Théorie de l'intégration

Jean JACOD

1999-2000

CHAPITRE 1

Introduction - La notion de mesure

1 Rappels sur les ensembles

Considérons un ensemble E, c'est-à-dire une collection d'objets appelés les "éléments", ou les "points", de E. L'appartenance d'un point x à l'ensemble E est notée $x \in E$, et $x \notin E$ signifie que le point x n'appartient pas à E.

Une partie de E est aussi un ensemble, appelé sous-ensemble de E: on écrit $F \subset E$ (on dit aussi que F est "inclus" dans E) lorsque F est un sous-ensemble de E.

Rappelons les opérations élémentaires sur les parties d'un ensemble:

Intersection: $A \cap B$ est l'intersection des ensembles A et B, i.e. l'ensemble des points appartenant à la fois à A et à B.

Réunion: $A \cup B$ est la réunion des ensembles A et B, i.e. l'ensemble des points appartenant à au moins l'un de ces deux ensembles.

Complémentaire: Si $A \subset E$, son complémentaire (dans E) est l'ensemble des points de E n'appartenant pas à A; on le note A^c, ou parfois $E \backslash A$.

Différence symétrique: $A \Delta B$ est l'ensemble des points appartenant à l'un des deux ensembles A ou B, mais pas aux deux; on a donc $A \Delta B = (A \backslash (A \cap B)) \cup (B \backslash (A \cap B))$.

Ensemble vide: C'est l'ensemble ne contenant aucun point; on le note $\emptyset$.

Ensembles disjoints: Les ensembles A et B sont dits *disjoints* si $A \cap B = \emptyset$.

La réunion et l'intersection sont des opérations commutatives et associatives: on a $A \cup B = B \cup A$ et $A \cap B = B \cap A$, et aussi $A \cup (B \cup C) = (A \cup B) \cup C$ et $A \cap (B \cap C) = (A \cap B) \cap C$, ensembles qu'on note naturellement $A \cup B \cup C$ et $A \cap B \cap C$. Plus généralement si on a une famille $(A_i)_{i \in I}$ d'ensembles, indexée par un ensemble quelconque I, on note $\cup_{i \in I} A_i$ (resp. $\cap_{i \in I} A_i$) la réunion (resp. l'intersection) de cette famille, i.e. l'ensemble des points appartenant à au moins l'un des A_i (resp. appartenant à tous les A_i): l'ordre d'indexation des A_i n'a pas d'importance.

Les ensembles suivants seront utilisés sans cesse:

$\mathbb{N}$ = ensemble des entiers naturels: $0, 1, 2, ...$

$\mathbb{N}^*$ = ensemble des entiers naturels non nuls: $1, 2, ...$

$\mathbb{Z}$ = ensemble des entiers relatifs: $..., -2, -1, 0, 1, 2, ...$

$\mathbb{Q}$ = ensemble des rationnels

$\mathbb{R}$ = ensemble des réels $=]-\infty, \infty[$

$\mathbb{R}^d$ = espace euclidien réel de dimension d (donc $\mathbb{R}^1 = \mathbb{R}$)

$\overline{\mathbb{R}} = [-\infty, \infty]$

$\mathbb{R}_+ = [0, \infty[$

$\overline{\mathbb{R}}_+ = [0, \infty]$

$\mathbb{C}$ = ensemble des nombres complexes.

L'ensemble des points a_i indexés par un ensemble I est noté $\{a_i : i \in I\}$. Si on a un nombre fini de points $a_1, ..., a_n$, on écrit aussi $\{a_1, a_2, ..., a_n\}$.

On sera amené très souvent à faire des opérations faisant intervenir $+\infty$ (qu'on écrit souvent, de manière plus simple, ∞) ou $-\infty$. Pour que ces opérations aient un sens précis, on fera *toujours* les conventions suivantes:

$$+\infty + \infty = +\infty, \quad -\infty - \infty = -\infty, \quad a + \infty = +\infty, \quad a - \infty = -\infty \text{ si } a \in \mathbb{R}, \tag{1}$$

$$0 \times \infty = 0, \quad a \in]0, \infty] \;\Rightarrow\; a \times \infty = +\infty, \quad a \in [-\infty, 0[\;\Rightarrow\; a \times \infty = -\infty. \tag{2}$$

Les ensembles dénombrables: on dit qu'un ensemble E est *dénombrable* s'il est en bijection avec $\mathbb{N}$, c'est-à-dire si on peut énumérer ses points en une suite $(x_n)_{n \in \mathbb{N}}$ (ce qui implique notamment que $x_n \neq x_m$ si $n \neq m$) : c'est le cas de $\mathbb{N}$ lui-même, ou de $\mathbb{N}^*$, de $\mathbb{Z}$, de $\mathbb{Q}$, ou encore des entiers pairs, ou de toute suite strictement croissante d'entiers. Ce n'est pas le cas ni de $\mathbb{R}$, ni des intervalles $[a, b]$ lorsque $a < b$.

Voici quelques propriétés des ensembles dénombrables: d'abord, toute partie d'un ensemble dénombrable est elle-même finie ou dénombrable. La réunion d'une famille finie ou dénombrable d'ensembles eux-mêmes finis ou dénombrables est un ensemble fini ou dénombrable. En revanche si A est n'est pas fini ou dénombrable, il en est de même de $A \backslash B$ pour tout $B \subset A$ qui est fini ou dénombrable.

Quelques résultats utiles sur les séries: Rappelons enfin quelques définitions et résultats sur les séries, notamment sur celles à termes positifs. Soit $(u_n)_{n \geq 1}$ une suite numérique, et $S_n = u_1 + ... + u_n$ la "somme partielle" à l'ordre n.

(S1) La série $\sum_n u_n$ est dite *convergente* si S_n converge vers une limite *finie* S, notée aussi $S = \sum_n u_n$ (c'est la "somme" de la série).

(S2) Si la série $\sum_n u_n$ converge, la *suite* $(u_n)_{n\geq 1}$ tend vers 0. La réciproque est **fausse**: on peut avoir $u_n \to 0$ sans que la série $\sum_n u_n$ converge.

(S3) La série $\sum_n u_n$ est dite *absolument convergente* si la série $\sum_n |u_n|$ converge.

(S4) Si on a $u_n \geq 0$ pour tout n, la suite S_n est croissante, donc elle tend toujours vers une limite $S \in \bar{\mathbb{R}}_+$. On écrit encore $S = \sum_n u_n$, bien que la série converge au sens de (S1) si et seulement si $S < \infty$. Avec les conventions (1) ceci s'applique même si les u_n sont à valeurs dans $\bar{\mathbb{R}}_+$.

En général l'ordre dans lequel on considère les termes d'une série est important. Il existe en effet de nombreux exemples de suites $(u_n)_{n\geq 1}$ et de bijections v de $\mathbb{N}^*$ dans lui-même pour lesquels $\sum_n u_n$ converge et $\sum_n u_{v(n)}$ diverge, ou converge vers une somme différente. Cela étant, il existe deux cas importants où l'ordre des termes n'a pas d'importance:

(S5) Lorsque les u_n sont des réels de signe quelconque et lorsque la série est absolument convergente, on peut modifier de manière arbitraire l'ordre des termes sans changer la propriété d'être absolument convergente, ni la somme de la série.

(S6) Si $u_n \in \bar{\mathbb{R}}_+$ pour tout n, la somme $\sum_n u_n$ (finie ou infinie: cf. (S4) ci-dessus) ne change pas si on change l'ordre de sommation. Rappelons rapidement la démonstration de cette propriété, qui est fondamentale pour les probabilités: soit v une bijection de $\mathbb{N}^*$ dans lui-même, $S_n = u_1 + \ldots + u_n$ et $S'_n = u_{v(1)} + \ldots + u_{v(n)}$; les suites (S_n) et (S'_n) sont croissantes, et on note S et S' leur limites respectives (dans $\bar{\mathbb{R}}_+$). Pour tout n il existe un entier $m(n)$ tel que $v(i) \leq m(n)$ dès que $i \leq n$; comme $u_i \geq 0$, on a donc clairement $S'_n \leq S_{m(n)} \leq S$, donc en passant à la limite on obtient $S' \leq S$. On montre de même que $S \leq S'$, donc $S = S'$.

2 Théorie de la mesure et théorie de l'intégration

La notion de *mesure* va étendre la notion usuelle de longueur pour les ensembles de $\mathbb{R}$, ou de volume pour ceux de $\mathbb{R}^d$, et ceci de deux manières: premièrement on veut pouvoir considérer des espaces de base plus généraux, ou plus “abstraits” (espaces de dimension infinie, espaces sur lesquels on définit les probabilités, etc...). Deuxièmement et surtout, on veut englober dans le même cadre mathématique d'une part les notions de longueurs, surface, volume, et d'autre part la notion de “masses” ou “charges ponctuelles” que l'on rencontre en mécanique ou en électricité, etc...

Prenons l'exemple de $\mathbb{R}^3$, supposé représenter un corps matériel ayant une densité $\rho(x)$ et une densité de charge électrique $\varepsilon(x)$ en chaque point x. Pour une partie raisonnable (on verra ce que veut dire “raisonnable” plus loin: pour le moment, on peut penser à une sphère, ou à un polyèdre) A de $\mathbb{R}^3$ on peut définir son volume $V(A)$, sa masse $M(A) = \int_A \rho(x)dx$ (intégrale de Riemann dans $\mathbb{R}^3$), sa charge électrique $E(A) = \int_A \varepsilon(x)dx$. Ces trois quantités ont a priori des propriétés “physiques” très différentes, mais elles partagent de manière évidente la propriété mathématique suivante (où $\mu(A)$ désigne $V(A)$, ou $M(A)$, ou $E(A)$):

(A) Additivité: On a $\mu(A \cup B) = \mu(A) + \mu(B)$ dès que A et B sont disjoints. $\square$

Ainsi, chaque partie raisonnable A de $\mathbb{R}^3$ a sa "mesure" (de volume, de masse, de charge) $\mu(A)$ et la propriété (A) ci-dessus est satisfaite: quitte à remplacer $\mathbb{R}^3$ par une ensemble E quelconque, on a là le contenu intuitif essentiel de la notion de *mesure*.

Malheureusement, la notion mathématique de mesure est un peu plus compliquée, pour deux raisons: d'abord, il faut définir ce qu'on entend par partie "raisonnable" de $\mathbb{R}^3$ (ou plus généralement de l'espace de base E sur lequel on se place); par exemple les polyèdres, et bien d'autres parties plus compliquées, ont des volumes, mais on peut construire des parties dont la "frontière" est si complexe que la notion de volume n'existe pas pour elles. Ensuite, la propriété (A) se révèle insuffisante pour avoir de bonnes propriétés pour les mesures.

Passons maintenant à l'intégration. Supposons que l'espace de base soit $E = [0,1]$.

Si f est une fonction réelle "convenable" sur E, on sait qu'on peut définir son *intégrale* $\int_0^1 f(x)dx$ au sens de Riemann. Rappelons en deux mots cette construction: pour chaque subdivision $\tau = \{0 = t_0 < t_1 < \ldots < t_k = 1\}$ de $[0,1]$ on pose

$$\begin{aligned} I_+(f,\tau) &= \sum_{i=1}^{k}(t_i - t_{i-1})\sup(f(x) : x \in [t_{i-1}, t_i]), \\ I_-(f,\tau) &= \sum_{i=1}^{k}(t_i - t_{i-1})\inf(f(x) : x \in [t_{i-1}, t_i]). \end{aligned}$$

On a bien sûr $I_-(f,\tau) \leq I_+(f,\tau)$, et la quantité $|\tau| = \sup(t_i - t_{i-1} : 1 \leq i \leq k)$ s'appelle le *pas* de la subdivision τ. On dit que f est *Riemann-intégrable* si, pour toute suite τ_n de subdivisions dont les pas $|\tau_n|$ tendent vers 0, la différence $I_+(f,\tau_n) - I_-(f,\tau_n)$ tend vers 0. Dans ce cas $I_+(f,\tau_n)$ et $I_-(f,\tau_n)$ convergent vers une limite commune et indépendante de la suite τ_n, et cette limite est l'intégrale de Riemann $\int_0^1 f(x)dx$ de f.

Cette notion d'intégrale semble à première vue assez naturelle, mais elle souffre de plusieurs inconvénients majeurs: d'abord, il est assez compliqué de décrire les fonctions Riemann-intégrables, et cette classe est plutôt petite comme on le verra ci-dessous; ensuite, elle s'étend assez facilement à $\mathbb{R}^d$, mais pas aux espaces de dimension infinie; mais surtout, elle est liée de manière intrinsèque à *une* mesure particulière sur $[0,1]$, à savoir la mesure de longueur, ou de *Lebesgue* comme elle sera appelée par la suite: en effet, si f est la fonction indicatrice du sous-intervalle $A = [a,b]$ de $[0,1]$ (i.e. $f(x) = 1$ quand $x \in A$ et $f(x) = 0$ quand $x \notin A$), alors $\int_0^1 f(x)dx = b - a$ est la longueur $\lambda(A) = b - a$ de A.

La théorie de l'intégration (au sens de Lebesgue) a pour but de pallier ces inconvénients: on pourra intégrer une classe de fonctions faciles à décrire, qu'on appellera les *fonctions mesurables*, sur un espace a-priori quelconque E, et par rapport à une mesure quelconque μ. Cette construction est en principe très simple: si f est l'indicatrice d'une partie A de E (donc $f(x) = 1$ si $x \in A$ et $f(x) = 0$ si $x \notin A$), l'intégrale de f "par rapport à μ" est $\int f d\mu = \mu(A)$. Puis, on "prolonge" cette intégrale à des fonctions plus générales par linéarité et continuité.

La construction de l'intégrale sera faite au chapitre 2, tandis que le reste de ce chapitre est consacré à la définition mathématique des mesures.

3 La classe des ensembles mesurables

Dans ce paragraphe, l'espace de base est un ensemble E quelconque. Comme on l'a mentionné ci-dessus dans le cas de la mesure "volume" sur $E = \mathbb{R}^3$, on ne peut pas en général, pour des raisons mathématiques, définir la mesure de n'importe quelle partie de E. Notre objectif ici est donc de définir la classe des parties de E dont on pourra définir la mesure.

1) Algèbres: Commençons par la notion la plus simple (mais mathématiquement insuffisante pour notre objectif):

Définition 1 Une classe $\mathcal{E}$ de parties de E est appelée *algèbre* (ou *algèbre de Boole*) si elle vérifie les trois axiomes suivants:

(T1) $E \in \mathcal{E}$,

(T2) $A \in \mathcal{E} \;\Rightarrow\; A^c \in \mathcal{E}$ ("stabilité par passage au complémentaire"),

(T3) $A, B \in \mathcal{E} \;\Rightarrow\; A \cup B \in \mathcal{E}$ ("stabilité par réunion").

Si $\mathcal{E}$ est une algèbre, les propriétés suivantes sont immédiates:

$$\emptyset \in \mathcal{E} \qquad \text{(par (T1) et (T2)).} \tag{3}$$

$$A_1, ..., A_n \in \mathcal{E} \quad \Rightarrow \quad A_1 \cup ... \cup A_n \in \mathcal{E} \qquad \text{("stabilité par réunion finie").} \tag{4}$$

$$A_1, ..., A_n \in \mathcal{E} \quad \Rightarrow \quad A_1 \cap ... \cap A_n \in \mathcal{E} \qquad \text{("stabilité par intersection finie").} \tag{5}$$

((4) s'obtient par récurrence à partir de (T3), et (5) s'obtient par (T2) et (4) puisque $A_1 \cap ... \cap A_n = (A_1^c \cup ... \cup A_n^c)^c$).

Il y a beaucoup d'algèbres sur E. La plus grosse est l'ensemble $\mathcal{P}(E)$ de toutes les parties de E. La plus petite est l'ensemble $\{\emptyset, E\}$ constituée des deux parties $\emptyset$ et E. Si $A \subset E$, la plus petite algèbre contenant A est $\{\emptyset, A, A^c, E\}$. L'intersection d'une famille quelconque d'algèbres est encore une algèbre.

2) Tribus: On a besoin en fait d'une notion (plus restrictive) de classe de parties de E:

Définition 2 Une classe $\mathcal{E}$ de parties de E est appelée *tribu* (ou *σ-algèbre de Boole*) si elle vérifie (T1), (T2) et l'axiome suivant:

(T4) $A_1, A_2, ... \in \mathcal{E} \;\Rightarrow\; \cup_{n \in \mathbb{N}^*} A_n \in \mathcal{E}$ ("stabilité par réunion dénombrable"). $\square$

Un élément de la tribu $\mathcal{E}$ s'appelle un *ensemble mesurable* (la terminologie se rapporte au fait que les "mesures" introduites au paragraphe suivant sont définies pour les éléments d'une tribu, qui sont donc "mesurables"); si on veut préciser la tribu, on dit que l'ensemble est "$\mathcal{E}$-mesurable", ou "mesurable par rapport à $\mathcal{E}$". Le couple $(E, \mathcal{E})$ constitué d'un ensemble E et d'une tribu s'appelle un *espace mesurable*.

On a (T4)$\Rightarrow$(T3) (prendre $A_1 = A$ et $A_2 = A_3 = ... = B$), donc *toute tribu est une algèbre*; en revanche il existe des algèbres qui ne sont pas des tribus (cf. ci-dessous).

Remarque: L'ensemble des propriétés (T1), (T2), (T3) (resp. (T1), (T2), (T4)) constitue ce qu'on appelle le système d'*axiomes* des algèbres (resp. des tribus). Il y a d'autres systèmes équivalents: si on pose

(T'1) $\emptyset \in \mathcal{E}$,

(T'3) $A, B \in \mathcal{E} \ \Rightarrow \ A \cap B \in \mathcal{E}$,

(T'4) $A_1, A_2, ... \in \mathcal{E} \ \Rightarrow \ \cap_{n \in I\!N^*} A_n \in \mathcal{E}$,

on a les équivalences

$$(T1)+(T2)+(T3) \ \Leftrightarrow \ (T1)+(T2)+(T'3) \ \Leftrightarrow \ (T'1)+(T2)+(T3) \ \Leftrightarrow \ (T'1)+(T2)+(T'3),$$

$$(T1)+(T2)+(T4) \ \Leftrightarrow \ (T1)+(T2)+(T'4) \ \Leftrightarrow \ (T'1)+(T2)+(T4) \ \Leftrightarrow \ (T'1)+(T2)+(T'4)$$

pour les algèbres et les tribus respectivement. $\square$

L'ensemble $\mathcal{P}(E)$ est une tribu (la plus grosse possible), tandis que $\{\emptyset, E\}$ est la plus petite. Si $A \subset E$, l'algèbre $\{\emptyset, A, A^c, E\}$ est une tribu. L'intersection d'une famille quelconque de tribus est encore une tribu, donc la définition suivante a un sens:

Définition 3 La *tribu engendrée* par une classe de parties $\mathcal{A}$ de E est la plus petite tribu contenant $\mathcal{A}$ (= l'intersection de toutes les tribus contenant $\mathcal{A}$; il y en a toujours au moins une, à savoir $\mathcal{P}(E)$). On la note $\sigma(\mathcal{A})$. $\square$

Exemples: 1) La tribu engendrée par $\mathcal{A} = \{A\}$ est $\{\emptyset, A, A^c, E\}$.

2) Soit $(E_i)_{i \in I}$ une *partition* de E (i.e. les ensembles E_i sont deux-à-deux disjoints, et $\cup_{i \in I} E_i = E$), indexée par un ensemble I fini ou dénombrable. La tribu engendrée par la classe $\{E_i : i \in I\}$ est l'ensemble des parties de la forme $A = \cup_{i \in J} E_i$, où J décrit l'ensemble des parties de I (avec la convention que $A = \emptyset$ si $J = \emptyset$). Si $I = \{1, 2\}$ et $E_1 = A$ et $E_2 = A^c$, on retrouve l'exemple 1. Si I est fini, cette tribu est aussi la plus petite algèbre contenant les A_i. Si I est dénombrable et si les E_i sont tous non vides, cette tribu contient strictement la plus petite algèbre contenant les A_i, qui peut être décrite ainsi: c'est l'ensemble des parties de la forme $A = \cup_{i \in J} E_i$, où J décrit l'ensemble des parties de I qui sont finies, ou de complémentaire fini: dans ce cas, cette algèbre n'est pas une tribu.

3) La tribu engendrée par la classe $\mathcal{A}$ des singletons de E, i.e. $\mathcal{A} = \{\{x\} : x \in E\}$, est l'ensemble des parties A de E qui sont finies ou dénombrables, ou qui sont de complémentaire A^c fini ou dénombrable. La plus petite algèbre contenant la classe $\mathcal{A}$ est l'ensemble des parties A de E qui sont finies ou de complémentaire fini. Cet exemple peut être vu comme une extension de l'exemple précédent. $\square$

Bien entendu, on peut avoir $\sigma(\mathcal{A}) = \sigma(\mathcal{B})$ pour deux classes différentes $\mathcal{A}$ et $\mathcal{B}$: dans l'exemple 1 ci-dessus, on a $\sigma(\{A\}) = \sigma(\{A^c\})$.

3) Quelques opérations sur les ensembles: On va introduire ci-dessous la notion de "limite" pour une suite $(A_n)_{\geq 1}$ de parties de E.

Définition 4 On dit qu'une suite $(A_n)_{n\geq 1}$ de parties de E *converge* (ou *tend*) vers la partie A, et on écrit $A_n \to A$, si pour tout $x \in A$ (resp. $x \notin A$) on a $x \in A_n$ (resp. $x \notin A_n$) pour tout n assez grand. En termes de quatificateurs, cela s'écrit:

$$\forall x \in A, \quad \exists n_0, \quad \forall n \geq n_0, \quad x \in A_n,$$

$$\forall x \notin A, \quad \exists n_0, \quad \forall n \geq n_0, \quad x \notin A_n, \quad \square$$

Il est facile de vérifier que cette définition revient à dire que la suite des fonctions indicatrices $(1_{A_n})_n$ converge simplement vers la fonction indicatrice 1_A (i.e., $1_{A_n}(x) \to 1_A(x)$ pour tout $x \in E$.

Si la suite $(A_n)_n$ est croissante (resp. décroissante), i.e. si $A_n \subset A_{n+1}$ (resp. $A_{n+1} \subset A_n$) pour tout n, alors elle converge vers $A = \cup_n A_n$ (resp. $A = \cap_n A_n$); on dit aussi dans ce cas que $(A_n)_n$ croit (resp. décroit) vers A, et on écrit $A_n \uparrow A$ ou $A = \lim_n \uparrow A_n$ (resp. $A_n \downarrow A$ ou $A = \lim_n \downarrow A_n$).

Il existe évidemment des suites $(A_n)_n$ de parties qui ne convergent pas. Mais dans tous les cas on peut poser:

Définition 5 On appelle *limite supérieure* et *limite inférieure* de la suite $(A_n)_n$ les ensembles suivants:

$$\left.\begin{array}{rcccl} \limsup_n A_n &=& \lim_n \downarrow \cup_{m\geq n} A_m &=& \cap_n \cup_{m\geq n} A_m \\ \liminf_n A_n &=& \lim_n \uparrow \cap_{m\geq n} A_m &=& \cup_n \cap_{m\geq n} A_m. \end{array}\right\} \tag{6}$$

On a une autre définition équivalente de ces ensembles:

$$x \in \limsup_n A_n \quad \Leftrightarrow \quad x \text{ appartient à } A_n \text{ pour une infinité d'indices } n, \tag{7}$$

$$x \in \liminf_n A_n \quad \Leftrightarrow \quad x \text{ appartient à } A_n \text{ pour tout } n \text{ sauf au plus un nombre fini.} \tag{8}$$

Dire que la suite $(A_n)_n$ converge revient à dire que $\limsup_n A_n = \liminf_n A_n$, et ce dernier ensemble est alors la limite des A_n. Le lecteur vérifiera aisément que

$$\limsup_n A_n = (\liminf_n A_n^c)^c, \qquad \liminf_n A_n = (\limsup_n A_n^c)^c. \tag{9}$$

Enfin, étant donnés (T4), (T'4) et (6), il est immédiat de vérifier que si $\mathcal{E}$ est une tribu,

$$A_n \in \mathcal{E} \quad \Rightarrow \quad \limsup_n A_n \in \mathcal{E}, \quad \liminf_n A_n \in \mathcal{E}. \tag{10}$$

En particulier on a:

$$A_n \in \mathcal{E} \quad \text{et} \quad A_n \to A \quad \Rightarrow \quad A \in \mathcal{E}. \tag{11}$$

4) La tribu borélienne de $\mathbb{R}$: La notion de tribu borélienne est liée à la structure "topologique" de l'ensemble de base. Comme la topologie n'est peut-être pas familière à tous les lecteurs nous allons essentiellement traiter le cas de $\mathbb{R}^d$, en commençant par le cas plus simple (au moins sur le plan des notations) de $\mathbb{R}$.

Etant donnée la structure relativement simple de cet ensemble, il existe plusieurs définitions équivalentes de la tribu borélienne de $\mathbb{R}$, et nous donnons la plus élémentaire:

Définition 6 La *tribu borélienne*, ou *tribu de Borel*, de $\mathbb{R}$ est la tribu engendrée par la classe des intervalles ouverts. On la note $\mathcal{R}$, ou $\mathcal{B}(\mathbb{R})$. Un élément de cette tribu est appelé une *partie borélienne*, ou un *borélien*. □

Voici quelques propriétés simples de cette tribu:

Proposition 7 *a) Tout intervalle ouvert, fermé, ou semi-ouvert, appartient à $\mathcal{R}$. Il en est de même de toute réunion finie ou dénombrable d'intervalles (ouverts, fermés, ou semi-ouverts).*

b) La tribu $\mathcal{R}$ est aussi la tribu engendrée par l'une quelconque des quatre classes suivantes d'ensembles:

(i) $\mathcal{J} = \{]-\infty, x] : x \in \mathbb{R}\}$,

(ii) $\mathcal{J}' = \{]-\infty, x] : x \in \mathbb{Q}\}$,

(iii) $\mathcal{K} = \{]-\infty, x[: x \in \mathbb{R}\}$,

(iv) $\mathcal{K}' = \{]-\infty, x[: x \in \mathbb{Q}\}$.

Preuve. a) On a $]a,b[\in \mathcal{R}$ par définition de $\mathcal{R}$. Comme $[a,b] = \cap_n]a - \frac{1}{n}, b + \frac{1}{n}[$ on a $[a,b] \in \mathcal{R}$ par (6). De même $[a,b[= \cap_n]a - \frac{1}{n}, b[$ et $]a,b] = \cap_n]a, b + \frac{1}{n}[$, on voit que ces deux intervalles semi-ouverts sont boréliens. La dernière assertion de (a) découle de (4) et (T4).

b) Nous ne montrons ici que les égalités $\sigma(\mathcal{J}) = \sigma(\mathcal{J}') = \mathcal{R}$, les deux autres se montrant de manière analogue. On a $\mathcal{J}' \subset \mathcal{J}$, et $\mathcal{J} \subset \mathcal{R}$ d'après (a). Il reste à montrer que $\mathcal{R} \subset \sigma(\mathcal{J}')$, et pour cela il suffit de vérifier que tout intervalle ouvert $]a,b[$ avec $a < b$ est dans $\sigma(\mathcal{J}')$. Il existe deux suites de rationnels $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ telles que $a < a_n < b_n < b$ et que $a_n \downarrow a$ et $b_n \uparrow b$. On a $]a_n, b_n] =]-\infty, b_n] \cap (]-\infty, a_n])^c$, donc $]a_n, b_n] \in \sigma(\mathcal{J}')$. On a aussi $]a,b[= \cup_n]a_n, b_n]$, donc $]a,b[\in \sigma(\mathcal{J}')$: le résultat est donc démontré. □

Remarques: 1) La proposition 7 montre que la tribu $\mathcal{R}$ est en fait engendrée par une classe dénombrable d'ensembles. Il est à noter que ce n'est pas le cas de toutes les tribus. Considérons par exemple la tribu $\mathcal{E}$ de $\mathbb{R}$ engendrée par la classe $\mathcal{A}$ des singletons (cf. Exemple 3 ci-dessus). Comme un singleton est un intervalle fermé, il appartient à $\mathcal{R}$, et par suite $\mathcal{E} \subset \mathcal{R}$. Cependant la classe $\mathcal{A}$ n'est pas dénombrable, et on peut d'ailleurs démontrer que $\mathcal{E}$ n'est engendrée (en tant que tribu) par aucune classe dénombrable, et ceci bien que $\mathcal{E}$ soit contenue dans $\mathcal{R}$.

2) Il n'est pas possible de donner une description plus concrète ou "constructive" de $\mathcal{R}$ que ci-dessus. Toutes les réunions finies ou dénombrables d'intervalles sont des boréliens, mais certains boréliens ne sont pas de cette forme. En fait, toutes les parties de $\mathbb{R}$ qu'on rencontre dans la pratique sont des boréliens, et il faut un peu se fatiguer pour construire une partie de $I\!R$ qui n'est pas borélienne: mais il en existe !

Examinons maintenant le cas de $\bar{I\!R}$, qui est tout-à-fait analogue à celui de $I\!R$, à ceci près qu'on doit distinguer les intervalles $]-\infty,x]$ (semi-ouvert) et $[-\infty,x]$ (fermé), et $]-\infty,x[$ (ouvert) et $[-\infty,x[$ (semi-ouvert), et de même en $+\infty$. Avec ces modifications triviales, la définition 6 reste valable, ainsi que la proposition 7 avec la même démonstration, à condition de remplacer $]-\infty,x]$ par $[-\infty,x]$. On notera $\bar{\mathcal{R}})$ la tribu borélienne de $\bar{I\!R}$.

La fin de ce paragraphe peut être omise. Elle a été rédigée en vue d'applications à des situations plus générales que celles de ce cours, mais qui se rencontrent parfois. En effet, la définition 6 de la tribu de Borel $\mathcal{R}$ n'est pas la définition "canonique". Celle-ci repose sur la notion d'ouvert: on dit qu'une partie A de $I\!R$ est un *ouvert* (ou une partie ouverte) si, pour tout $x\in A$, il existe un $\varepsilon>0$ tel qu'on ait l'inclusion $]x-\varepsilon,x+\varepsilon[\subset A$. Le complémentaire d'un ouvert est ce qu'on appelle un *fermé*, ou une partie fermée.

Les intervalles ouverts (resp. fermés) sont des ouverts (resp. des fermés); l'ensemble vide et $I\!R$ lui-même sont des ouverts, et donc aussi des fermés, mais il n'existe pas d'autre partie de $I\!R$ qui soit à la fois ouverte et fermée; les intervalles semi-ouverts $[a,b[$ et $]a,b]$ ne sont ni ouverts ni fermés lorsque $a,b\in I\!R$ et $a<b$ (toutefois $]-\infty,b]$ et $[a,\infty[$ sont fermés). Une réunion quelconque d'ouverts est un ouvert. Une intersection *finie* d'ouverts est un ouvert, mais une intersection infinie (dénombrable ou non) d'ouverts peut ne pas être un ouvert: par exemple l'intersection des intervalles ouverts $]-\frac{1}{n},\frac{1}{n}[$ est le fermé $\{0\}$.

La structure des ouverts de $I\!R$ est donc plutôt compliquée, et l'intérêt d'introduire une telle notion n'est peut-être pas évident a-priori. En fait elle offre la possibilité de définir de manière simple la convergence des suites: une suite de réels $(x_n)_{n\geq 1}$ converge vers une limite x si et seulement si pour tout ouvert A contenant x, les x_n sont dans A pour tout n assez grand (en termes "axiomatiques": si et seulement si pour tout ouvert A contenant x, il existe un entier N tel que $n>N\Rightarrow x_n\in A$); par ailleurs, elle s'étend à des espaces plus abstraits que $I\!R$. On a alors le résultat suivant:

Proposition 8 *a) Tout ouvert non vide A de $I\!R$ est réunion dénombrable d'intervalles ouverts, et aussi réunion dénombrable d'intervalles fermés.*

b) La tribu borélienne $\mathcal{R}$ est la tribu engendrée par la classe des ouverts, et aussi la tribu engendrée par la classe des fermés.

Preuve. a) Soit A un ouvert non vide. Soit $\mathcal{A}$ (resp. $\mathcal{B}$) la famille des intervalles $]a,b[$ (resp. $[a,b]$) qui sont contenus dans A et qui sont d'extrémités a et b dans l'ensemble des rationnels $\mathbb{Q}$. L'ensemble de ces intervalles est dénombrable. Si par ailleurs $x\in A$ il existe $\varepsilon>0$ avec $]x-\varepsilon,x+\varepsilon[\subset A$, donc il existe deux rationnels a,b avec $x-\varepsilon<a<x<b<x+\varepsilon$, donc $]a,b[\subset[a,b]\subset A$: donc x est dans l'un des éléments au moins de chacune des classes $\mathcal{A}$ et $\mathcal{B}$. Il s'ensuit que A est la réunion des intervalles appartenant à $\mathcal{A}$ (resp. à $\mathcal{B}$).

b) D'une part tout ouvert est réunion dénombrable d'intervalles ouverts, donc est dans $\mathcal{R}$ par (T4): donc la tribu engendrée par les ouverts est contenue dans $\mathcal{R}$. A l'inverse, les intervalles ouverts sont des ouverts, donc $\mathcal{R}$ est contenue dans la tribu engendrée par les ouverts: cela démontre la première partie de (b). Comme un ensemble est fermé si et seulement si c'est le complémentaire d'un ouvert, (T2) montre que la tribu engendrée par la classe des ouverts et celle engendrée par la classe des fermés sont identiques. □

C'est en fait la propriété (b) ci-dessus qui fournit la définition habituelle de la tribu borélienne. On dit qu'un ensemble E est un espace topologique s'il est muni d'une classe $\mathcal{A}$ d'ensembles (les ouverts) stable par intersection finie et par réunion quelconque, contenant $\emptyset$ et E. Les fermés sont par définition les complémentaires des ouverts, et on pose:

Définition 9 Si E est un espace topologique, la *tribu borélienne* de E, notée $\mathcal{B}(E)$, est la tribu engendrée par la classe des parties ouvertes de E (comme les fermés de E sont les complémentaires des ouverts, $\mathcal{B}(E)$ est aussi la tribu engendrée par la classe des fermés de E). Un élément de la tribu borélienne est aussi appelé une *partie borélienne*, ou un *borélien*, de E □

5) La tribu borélienne de $I\!R^d$: On va maintenant examiner le cas de $I\!R^d$. Rappelons que si les A_i sont des parties de $I\!R$, leur "produit" $\prod_{i=1}^d A_i$ est la partie de $I\!R^d$ constituée des points (ou "vecteurs") x dont les "coordonnées" x_i sont contenues dans les A_i. Donnons d'abord la définition "naïve" des boréliens de $I\!R^d$, analogue à la définition 6:

Définition 10 La tribu borélienne $\mathcal{R}^d$, ou $\mathcal{B}(I\!R^d)$, de $I\!R^d$ est la tribu engendrée par la classe des "rectangles ouverts" $\prod_{i=1}^d]a_i, b_i[$. Attention à la notation (usuelle) $\mathcal{R}^d$: la tribu borélienne de $I\!R^d$ n'est pas, comme on le varre plus tard, le $d^{\text{ème}}$ puissance cartésienne de la tribu $\mathcal{R}$ de $I\!R$.

Une démonstration analogue à celle de la proposition 7-b donne:

$$\left.\begin{array}{l}\text{La tribu } \mathcal{R}^d \text{ est la tribu engendrée par la classe des rectangles}\\ \text{de la forme } \prod_{i=1}^d]-\infty, x_i], \text{ avec les } x_i \text{ rationnels.}\end{array}\right\} \qquad (12)$$

Si on veut maintenant utiliser la définition 9, il convient d'abord de définir les ouverts de $I\!R^d$. Une partie A est dite ouverte si pour tout $x \in A$ il existe $\varepsilon > 0$ tel que tous les points y situés à une distance inférieure à ε de x sont dans A (la distance est ici la distance euclidienne usuelle). Là encore, une suite $(x_n)_{n\geq 1}$ converge vers une limite x dans $I\!R^d$ si et seulement si pour tout ouvert A contenant x, on a $x_n \in A$ pour tout n assez grand.

Proposition 11 *La tribu $\mathcal{R}^d$ est la tribu engendrée par les ouverts de $I\!R^d$, et aussi celle engendrée par les boules ouvertes de $I\!R^d$ (on appelle boule ouverte de centre x et de rayon $a > 0$ l'ensemble des $y \in I\!R^d$ qui sont à une distance strictement inférieure à a de x).*

Preuve. Soit $\mathcal{A}$ et $\mathcal{B}$ les tribus engendrées par les ouverts, et par les boules ouvertes, respectivement. Toute boule ouverte étant un ouvert, on a $\mathcal{B} \subset \mathcal{A}$.

Exactement comme dans la proposition 8, un ouvert A est la réunion (dénombrable) de toutes les boules ouvertes contenues dans A, dont le rayon a est rationnel et dont le centre x a des coordonnées qui sont rationnelles: cela implique que $\mathcal{A} \subset \mathcal{B}$, donc $\mathcal{B} = \mathcal{A}$.

Par ailleurs on voit qu'un rectangle ouvert est un ouvert (vérification immédiate), de sorte que $\mathcal{R}^d \subset \mathcal{B}$. Enfin, il est facile de vérifier qu'une boule ouverte B est la réunion (dénombrable) de tous les rectangles ouverts $\prod_{i=1}^{d}]a_i, b_i[$ qui sont contenus dans B et tels que les a_i et b_i sont des rationnels: cela implique que $\mathcal{B} \subset \mathcal{R}^d$, donc finalement $\mathcal{B} = \mathcal{R}^d$. □

4 Les mesures

Nous allons maintenant donner un sens mathématique précis à la notion de mesure. Dans tout ce paragraphe, l'espace de base E est fixé et muni d'une tribu $\mathcal{E}$ également fixée (on dit parfois que le couple $(E, \mathcal{E})$ est un *espace mesurable*, ce qui exprime bien qu'on a les ingrédients nécessaire à la construction des mesures).

Définition 12 Une *mesure* sur $(E, \mathcal{E})$ est une application μ de $\mathcal{E}$ dans $\overline{\mathbb{R}}_+ = [0, \infty]$, vérifiant "l'axiome de σ-additivité" suivant:

(SA) σ-additivité: $\mu(\cup_{n \in \mathbb{N}^*} A_n) = \sum_{n \in \mathbb{N}^*} \mu(A_n)$ pour toute suite $(A_n)_{n \geq 1}$ d'éléments de $\mathcal{E}$ qui sont deux-à-deux disjoints (i.e. $A_n \cap A_m = \emptyset$ si $n \neq m$),

ainsi que l'axiome suivant:

(O) $\mu(\emptyset) = 0$.

La mesure μ est dite *finie*, ou *de masse totale finie*, si $\mu(E) < \infty$. □

Une mesure est donc une *application* sur la tribu $\mathcal{E}$; mais par abus de langage la quantité $\mu(A)$ pour un $A \in \mathcal{E}$ s'appelle la "mesure de l'ensemble A" (ou parfois: la "valeur de μ sur A")

Dans l'axiome de σ-additivité (SA), la réunion $\cup_n A_n$ ne dépend pas de l'ordre par lequel on numérote les A_n; grâce à la propriété (S6), la somme $\sum_n \mu(A_n)$ ne dépend pas non plus de l'ordre de sommation !

On verra plus loin que les propriétés (SA) et (O) impliquent la propriété d'additivité (A), ce qui n'est pas complètement évident a-priori. Une application de $\mathcal{E}$ dans $\overline{\mathbb{R}}_+$ qui vérifie seulement (A) s'appelle une *mesure additive*, bien que ce ne soit pas nécessairement une mesure ! Intuitivement parlant, la notion de mesure additive est plus naturelle que celle de mesure, que ce soit pour les mesures "de volume", "de masse", etc... évoquées plus haut, ou dans le cadre de la théorie des probabilités. Mais elle a un défaut rédhibitoire: la classe des mesures additives a une structure mathématique extrêmement pauvre, ne permettant en particulier pas de définir une notion satisfaisante d'intégrale par rapport à ces mesures additives. On est donc conduit à utiliser les mesures au sens de la définition 12; et c'est la forme de l'axiome de σ-additivité (SA) qui nous oblige à considérer comme classe d'ensembles "mesurables" une tribu au lieu de la notion plus simple d'algèbre.

Le fait que $\mu(A) \geq 0$ pour tout A est une restriction propre à ce cours: il conviendrait d'appeler la notion définie ci-dessus une *mesure positive*, mais pour des raisons de simplicité nous ne le ferons pas en général.

Le fait que $\mu(A)$ puisse être infini pour certains A est indispensable pour les applications. Par exemple si $E = \mathbb{R}$ et si μ représente la mesure de longueur, $\mu(\mathbb{R})$ (qui est la "longueur totale" de $\mathbb{R}$) vaut $+\infty$.

Exemples:

1) La *mesure nulle* est celle qui vaut $\mu(A) = 0$ pour tout $A \in \mathcal{E}$: (0) et la σ-additivité (SA) sont évidemment vérifiés.

2) La *mesure infinie* est celle qui vaut $\mu(A) = +\infty$ pour tout $A \in \mathcal{E}$ qui n'est pas vide, et $\mu(\emptyset) = 0$: (SA) et (O) sont évidemment vérifiés.

3) La *mesure de Dirac en un point x*: c'est la mesure notée ε_x, qui vaut

$$\varepsilon_x(A) = \begin{cases} 1 & \text{si} \quad x \in A \\ 0 & \text{si} \quad x \notin A. \end{cases} \tag{13}$$

Là encore (SA) et (O) sont évidemment vérifiés. Si $E = I\!R^3$, la mesure ε_a peut être interprétée comme la "mesure de masse" associée à la masse "ponctuelle" au point a, au sens de la mécanique rationnelle.

4) La *mesure de comptage* est celle pour laquelle $\mu(A)$ est le nombre de points de l'ensemble A. $\square$

Tous ces exemples sont élémentaires, dans le sens où la vérification de (SA) est évidente. D'ailleurs, ces mesures sont définies sur une tribu quelconque, et en particulier sur la tribu $\mathcal{P}(E)$ de toutes les parties de E (et ceci, quel que soit l'espace E). Nous énoncerons plus bas des résultats d'existence de mesures plus complexes (et plus utiles), notamment pour la mesure de Lebesgue (mesure de longueur sur $I\!R$, ou de volume sur $I\!R^d$). Mais auparavant nous donnons quelques propriétés simples des mesures.

Proposition 13 *Toute mesure μ sur $(E, \mathcal{E})$ vérifie l'additivité (A), ainsi que les propriétés suivantes (ci-dessous on a $A, B, A_1, ..., A_n$ dans $\mathcal{E}$):*

$$\mu(A_1 \cup \ldots \cup A_n) = \mu(A_1) + \ldots + \mu(A_n) \quad \text{si les } A_1, .., A_n \text{ sont deux-à-deux disjoints,} \tag{14}$$

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B), \tag{15}$$

$$A \subset B \quad \Rightarrow \quad \mu(A) \leq \mu(B). \tag{16}$$

En particulier, (14) implique (A). Remarquer l'écriture de (15): on ne peut pas en général écrire $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$, puisque dans le second membre il se peut que tous les termes soient infinis, et que $\infty - \infty$ n'a pas de sens; en revanche $+\infty + \infty$ "vaut" naturellement $+\infty$, de sorte que (15) a bien un sens dans tous les cas.

Preuve. (14) se déduit immédiatement de (0) et de (SA) appliqué à la suite $B_1 = A_1$,..., $B_n = A_n$, $B_{n+1} = \emptyset$, $B_{n+2} = \emptyset$,...

Pour (15) on pose $C = A \cap B$, $A' = A \backslash C$ at $B' = B \backslash C$. On remarque que $A \cup B = A' \cup C \cup B'$, $A = A' \cup C$ et $B = B' \cup C$, tandis que les trois ensembles A', C, B' sont deux-à-deux disjoints. Par suite (14) implique

$$\mu(A \cup B) = \mu(A') + \mu(C) + \mu(B'),$$

$$\begin{aligned}\mu(A) &= \mu(A') + \mu(C),\\ \mu(B) &= \mu(B') + \mu(C).\end{aligned}$$

En additionnant ces trois égalités membre à membre, on obtient (15).

Enfin, si $A \subset B$, en posant $A' = B\backslash A$ on a $\mu(B) = \mu(A) + \mu(A')$ par (14), et comme $\mu(A') \geq 0$ on obtient (16). $\square$

Les mesures possèdent également des propriétés de "continuité" pour les suites d'ensembles, que nous énonçons ci-dessous:

Théorème 14 *Soit μ une mesure sur $(E, \mathcal{E})$.*

a) Pour toute suite croissante $(A_n)_{n\geq 1}$ d'éléments de $\mathcal{E}$, $\mu(\lim_n \uparrow A_n) = \lim_n \uparrow \mu(A_n)$.

b) Si $(A_n)_{n\geq 1}$ est une suite d'éléments de $\mathcal{E}$ convergeant vers une limite A (au sens de la définition 4), *et s'il existe un $B \in \mathcal{E}$ tel que $A_n \subset B$ pour tout n et $\mu(B) < \infty$, alors $\mu(A_n) \to \mu(A)$.*

L'assertion (b) ci-dessus est une version préliminaire d'un théorème plus général, fondamental dans la théorie de l'intégration, qu'on appelle le théorème de convergence dominée de Lebesgue. Ce résultat est en général faux sans l'hypothèse que les A_n sont contenus dans un ensemble de mesure finie, comme le montre le contre-exemple suivant: soit μ la mesure de comptage sur $E =]0, 1]$, et soit $A_n =]0, 1/n]$; on a $\mu(A_n) = \infty$ puisqu'il y a une infinité de points dans A_n; cependant, A_n décroît vers l'ensemble vide $A = \emptyset$, de sorte que $\mu(A_n)$ ne converge pas vers $\mu(A)$.

Preuve. a) Posons $A_0 = \emptyset$ et $B_n = A_n\backslash A_{n-1}$ pour $n \geq 1$. Les ensembles B_n sont deux-à-deux disjoints, et on a $A_n = B_1 \cup ... \cup B_n$, ainsi que $A = \cup_{n\geq 1} B_n$ si A désigne la limite croissante des A_n. (14) entraine $\mu(A_n) = \mu(B_1) + \ldots + \mu(B_n)$, tandis que (SA) entraine $\mu(A) = \sum_{n\geq 1} \mu(B_n)$. Par définition de la somme (éventuellement infinie) d'une série à termes positifs, on en déduit que $\mu(A)$ est la limite (évidemment croissante) des sommes partielles $\mu(A_n)$.

b) Supposons maintenant que $A_n \to A$ et que $A_n \subset B$ pour tout n, avec $\mu(B) < \infty$. Si la suite $(A_n)_n$ est croissante, le résultat a été obtenu dans (a). Supposons ensuite que (A_n) soit décroissante. Si $C_n = A_1\backslash A_n$, la suite (C_n) est clairement croissante, et sa limite est $C = A_1\backslash A$, donc $\mu(C_n) \uparrow \mu(A_1\backslash A)$; Mais $\mu(A_n) = \mu(A_1) - \mu(C_n)$ et $\mu(A) = \mu(A_1) - \mu(C)$ par (14): remarquer que les mesures de A_n, C_n, A, C sont toutes finies, puisque ces ensembles sont contenus dans B par hypothèses; on en déduit que $\mu(A_n) \downarrow \mu(A)$.

Passons au cas général. Soit $C_n = \cup_{m:m\geq n} A_m$ and $D_n = \cap_{m:m\geq n} A_m$. On a $D_n \subset A_n \subset C_n \subset B$, et les suites C_n et D_n sont respectivement décroissante et croissante, et convergent vers les limites $C = \limsup_n A_n$ et $D = \liminf_n A_n$ (cf. (6)); de plus comme $A_n \to A$, on a $C = D = A$. Les résultats précédents impliquent $\mu(C_n) \downarrow \mu(A)$ et $\mu(D_n) \uparrow \mu(A)$. Comme $\mu(D_n) \leq \mu(A_n) \leq \mu(C_n)$, il s'ensuit que $\mu(A_n) \to \mu(A)$. $\square$

Proposition 15 *Soit μ une mesure sur $(E, \mathcal{E})$ et $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{E}$. On a alors*

$$\mu(\cup_n A_n) \;\leq\; \sum_n \mu(A_n). \tag{17}$$

Preuve. Soit $B_n = A_1 \cup ... \cup A_n$, $C_1 = A_1$ et $C_n = B_n \backslash B_{n-1}$ si $n \geq 2$. Comme $C_i \subset A_i$ on a $\mu(C_i) \leq \mu(A_i)$. Par ailleurs les C_n sont deux-à-deux disjoints et $\cup_n C_n = \cup_n A_n$, donc $\mu(\cup_n A_n) = \mu(\cup_n C_n) = \sum_n \mu(C_n)$ par (SA), donc (17) est immédiat. $\square$

Il existe trois opérations simples sur les mesures:

La restriction d'une mesure: Si μ est une mesure sur $(E, \mathcal{E})$ et si $B \in \mathcal{E}$, la formule $\mu_B(A) = \mu(A \cap B)$ pour tout $A \in \mathcal{E}$ définit une nouvelle mesure μ_B (comme $B \cap (\cup_n A_n) = \cup_n (B \cap A_n)$, μ_B vérifie clairement (SA), et aussi (O)).

L'addition de deux mesures: si μ et ν sont deux mesures sur $(E, \mathcal{E})$, la formule $\eta(A) = \mu(A) + \nu(A)$ pour tout $A \in \mathcal{E}$ définit une nouvelle mesure η, notée $\eta = \mu + \nu$.

La multiplication par un réel positif: si μ est une mesure sur $(E, \mathcal{E})$ et si $a \in I\!\!R_+$, la formule $\nu(A) = a\mu(A)$ pour tout $A \in \mathcal{E}$ définit une nouvelle mesure, notée $\nu = a\mu$ (avec la convention $0 \times \infty = 0$, on a le même résultat si $a = +\infty$).

L'addition des mesures est évidemment commutative et associative. On a aussi $a(b\mu) = (ab)\mu$, et la distributivité: $a\mu + a\nu = a(\mu + \nu)$.

Proposition 16 *Soit $(\mu_n)_{n \geq 1}$ une suite de mesures sur $(E, \mathcal{E})$.*

a) Si la suite $(\mu_n)_n$ est croissante, ce qui signifie que $\mu_n(A) \leq \mu_{n+1}(A)$ pour tout n et tout $A \in \mathcal{E}$, la formule $\mu(A) = \lim_n \uparrow \mu_n(A)$ pour tout $A \in \mathcal{E}$ définit une nouvelle mesure appelée la limite croissante des μ_n.

b) La formule $\nu(A) = \sum_n \mu_n(A)$ pour tout $A \in \mathcal{E}$ définit une nouvelle mesure, notée $\nu = \sum_n \mu_n$.

Preuve. a) On a clairement $\mu(\emptyset) = 0$. Il reste donc à montrer que μ vérifie (SA). Pour cela, il suffit de prouver que si A_n est une suite d'éléments deux à deux disjoints de $\mathcal{E}$, si $A = \cup_n A_n$ et si $a = \sum_n \mu(A_n)$, alors $\mu(A) = a$.

On a $\mu_n(A) \geq \mu_n(A_1) + ... + \mu_n(A_p)$ pour tout p entier, et en passant à la limite en n on obtient $\mu(A) \geq \mu(A_1) + ... + \mu(A_p)$. Comme ceci est vrai pour tout p, on a aussi $\mu(A) \geq a$.

Si $a = +\infty$, on en déduit que $\mu(A) = a$. Si maintenant $a < \infty$, pour tout $\varepsilon > 0$ il existe p tel que $\sum_{i:i>p} \mu(A_i) \leq \varepsilon$. Comme $\mu_n(A_i) \leq \mu(A_i)$ on a aussi $\sum_{i:i>p} \mu_n(A_i) \leq \varepsilon$ pour tout n, ce qui entraîne $\mu_n(A) \leq \mu_n(A_1) + ... + \mu_n(A_p) + \varepsilon$ par (SA) appliqué à μ_n. En passant à la limite en n dans cette inégalité, on trouve $\mu(A) \leq \mu(A_1) + ... + \mu(A_p) + \varepsilon$; donc $\mu(A) \leq a + \varepsilon$, et comme cette inégalité est valide pour tout $\varepsilon > 0$ on a en fait $\mu(A) \leq a$. Par suite $\mu(A) = a$.

b) Si $\nu_n = \mu_1 + ... + \mu_n$ (se rappeler l'associativité de l'addition des mesures), on obtient une suite croissante $(\nu_n)_n$ de mesures, et $\nu(A) = \lim_n \uparrow \nu_n(A)$ pour tout $A \in \mathcal{E}$: il suffit alors d'appliquer (a) pour obtenir le résultat. $\square$

Parmi toutes les mesures, les seules qu'on sache vraiment étudier sont les mesures finies (i.e. telles que $\mu(E) < \infty$), et les suivantes:

Définition 17 Une mesure μ sur $(E, \mathcal{E})$ est dite *σ-finie* s'il existe une suite croissante $(E_n)_{n \geq 1}$ d'éléments de $\mathcal{E}$ dont la limite est E, et telle que $\mu(E_n) < \infty$ pour tout n. $\Box$

Ces mesures sont limites croissantes (au sens de la proposition 16-a) de mesures finies, à savoir des restrictions μ_{E_n} de μ à chaque E_n. On peut aussi les considérer comme des sommes infinies (au sens de la proposition 16-b) de mesures finies, à savoir les restrictions $\mu_{E'_n}$ de μ à chaque ensemble $E'_n = E_n \backslash E_{n-1}$ (avec la convention $E_0 = \emptyset$).

Noter qu'il existe des mesures qui ne sont pas σ-finies: la mesure infinie (exemple 2 ci-dessus), ou la mesure de comptage sur E lorsque E n'est pas fini ou dénombrable (cette dernière mesure est finie si E est fini, et σ-finie si E est dénombrable).

Enfin, on peut "normaliser" une mesure finie non nulle μ en la multipliant par la constante $a = 1/\mu(E)$. La nouvelle mesure $\nu = a\mu$ vérifie $\nu(E) = 1$. Ainsi, l'étude des mesures σ-finies se ramène, pour beaucoup de leurs propriétés, à celle des mesures de masse totale 1, qui portent un nom spécial:

Définition 18 Une *probabilité* (ou *mesure de probabilité*) sur $(E, \mathcal{E})$ est une mesure de masse totale $\mu(E) = 1$. $\Box$

5 La mesure de Lebesgue

Dans ce paragraphe nous définissons la mesure qui est de loin la plus importante en analyse (et en probabilités), qui est la mesure de Lebesgue (mesurant la "longueur" dans le cas de $\mathbb{R}$, la "surface" dans $\mathbb{R}^2$, le "volume" dans $\mathbb{R}^3$, etc...)

Nous commençons par le cas de $\mathbb{R}$, qu'on munit de la tribu borélienne $\mathcal{R}$. On connait bien-sûr la longueur des intervalles:

$$\lambda(A) \;=\; b-a \quad \text{si} \;\; A = [a,b], \;\text{ou}\;\; A = [a,b[, \;\text{ou}\;\; A =]a,b], \;\text{ou}\; A =]a,b[. \tag{18}$$

Cette propriété est compatible avec (SA), au sens ou $\lambda(A) = \sum \lambda(A_n)$ dès que les A_n sont des intervalles deux-à-deux disjoints dont la réunion A est encore un intervalle (cette propriété est assez facile à vérifier, mais pas complètement évidente sauf dans le cas où on peut numéroter les A_n de sorte que A_n soit à gauche de A_{n+1} pour tout n, ou bien à droite de A_{n+1} pour tout n; mais il y a des cas où aucune de ces deux propriétés n'est vérifiée).

La question qui se pose est donc la suivante: existe-t-il une (plusieurs) mesure(s) sur les boréliens de $\mathbb{R}$ qui vérifie(nt) (18) ? La réponse est donnée par le théorème suivant:

Théorème 19 *Il existe une mesure λ et une seule sur $(\mathbb{R}, \mathcal{R})$ qui vérifie (18), et qu'on appelle la mesure de Lebesgue.*

Ce résultat est difficile, et pour le moment nous l'admettrons. Il contient en fait deux résultats de nature différente. D'abord il y a l'existence de λ (qu'on appelle le théorème de prolongement): on connaît λ sur la classe $\mathcal{A}$ des intervalles; cette classe engendre la tribu borélienne (cf. proposition 7), et on peut "prolonger" λ à la tribu $\mathcal{R}$, de façon à obtenir une mesure (c'est la partie la plus difficile du théorème; la difficulté tient au fait qu'on ne sait pas décrire de manière "concrète" les boréliens). Ensuite, il y a un résultat d'unicité, qui sera démontré plus loin et qui est beaucoup plus facile.

En fait, la tribu $\mathcal{R}$ n'est pas tout à fait la plus grande possible sur laquelle on puisse définir la mesure de Lebesgue: ce qui veut dire que le prolongement dont il est question ci-dessus peut se faire sur une tribu $\mathcal{R}'$ plus grande que $\mathcal{R}$ (qu'on appellera plus loin la "complétée" de $\mathcal{R}$). Mais il est remarquable que la mesure de Lebesgue ne puisse pas se prolonger à la tribu $\mathcal{P}(\mathbb{R})$ de toutes les parties de $\mathbb{R}$: *il n'existe pas* de mesure sur $\mathcal{P}(\mathbb{R})$ vérifiant (18).

Voici quelques propriétés simples de la mesure de Lebesgue:

a) La mesure (ou "longueur") des singletons est $\lambda(\{a\}) = 0$ (appliquer (18) avec $A = [a, a]$).

b) Tout ensemble fini ou dénombrable A est borélien, de mesure $\lambda(A) = 0$: on peut écrire en effet $A = \cup_{n\geq 1}\{a_n\}$, où les a_n sont les points de A (qu'on peut toujours énumérer en une "suite" finie ou infinie). Il suffit alors d'appliquer (T4) et (SA) pour obtenir les résultats.

c) Un intervalle $A = [a, b]$ peut également s'écrire comme la réunion des singletons $\{x\}$ pour $x \in A$. Cependant on n'a pas $\lambda(A) = \sum_{x\in A} \lambda(\{x\})$ (en d'autres termes, la propriété (SA) ne s'étend pas à des familles non dénombrables d'ensembles): en effet $\lambda(A) > 0$, tandis que tous les termes de la somme de droite sont nuls, donc la seule valeur qu'on puisse raisonnablement donner à cette somme est 0 (une autre raison plus fondamentale est en fait que la somme d'une infinité non dénombrable de termes n'a a-priori pas de sens).

En particulier, la mesure de Lebesgue de l'ensemble $\mathbb{Q}$ de tous les rationnels est nulle: cette propriété manifeste le fait que la mesure de Lebesgue est une extension de la notion de longueur, mais ne se réduit pas à cette notion; en effet un ensemble de structure aussi compliquée que $\mathbb{Q}$ *n'a pas de longueur* au sens "physique" du terme, bien qu'il admette une mesure de Lebesgue. Le fait que que certaines parties de $\mathbb{R}$ n'admettent pas de mesure de Lebesgue montre qu'il y a des parties dont la structure est encore beaucoup plus compliquée que celle de $\mathbb{Q}$.

Passons maintenant au cas de $\mathbb{R}^d$, qu'on munit de la tribu borélienne $\mathcal{R}^d$. Le volume d'un rectangle de la forme $A = \prod_{i=1}^d]a_i, b_i[$ est

$$\lambda_d(A) = \prod_{i=1}^{d} (b_i - a_i), \tag{19}$$

et on a l'analogue du théorème 19:

Théorème 20 *Il existe une mesure λ_d et une seule sur $(\mathbb{R}^d, \mathcal{R}^d)$ qui vérifie (19), et qu'on appelle la mesure de Lebesgue.*

(Ce théorème se réduit au théorème 19 lorsque $d = 1$.) Une autre manière de voir les choses consiste à remarquer que (19) peut s'écrire

$$\lambda_d(\prod_{i=1}^{d} A_i) \;=\; \prod_{i=1}^{d} \lambda(A_i) \tag{20}$$

lorsque les A_i sont des intervalles. Cette propriété, qui d'une certaine manière traduit le fait que la mesure de Lebesgue λ_d sur $I\!R^d$ est la puissance $d^{\text{ème}}$ de la mesure de Lebesgue $\lambda = \lambda_1$ sur $I\!R$, se généralise ainsi:

Théorème 21 *Si les A_i sont des boréliens de $I\!R$, le produit $A = \prod_{i=1}^{d} A_i$ est un borélien de $I\!R^d$, et on a la propriété (20).*

Ce résultat sera démontré dans le chapitre consacré aux produits de mesures, et il préfigure les résultats de ce chapitre.

CHAPITRE 2

L'intégration par rapport à une mesure

Ce chapitre est consacré à la construction de l'intégrale des fonctions par rapport à une mesure. On fixe donc dans tout le chapitre un espace E, muni d'une tribu $\mathcal{E}$ et d'une mesure μ. Le lecteur pourra avoir à l'esprit les trois exemples fondamentaux suivants: celui de $E = \mathbb{R}$ avec $\mathcal{E} = \mathcal{R}$ (tribu borélienne) et $\mu = \lambda$ (mesure de Lebesgue); celui de $E = \mathbb{N}^*$ avec $\mathcal{E} = \mathcal{P}(E)$ (tribu de toutes les parties de E) et μ la mesure de comptage ($\mu(A) =$ le nombre de points de A); enfin celui d'un ensemble E arbitraire, avec $\mathcal{E} = \mathcal{P}(E)$ et $\mu = \varepsilon_x$ la masse de Dirac en un point x: voir (1-13). Dans le premier cas, la théorie de l'intégration permet d'étendre l'intégrale de Riemann; dans le second cas elle est une autre manière de considérer la sommation des séries; le troisième cas est essentiellement trivial, mais permet de vérifier la compréhension des notions et résultats présentés.

Il est important de remarquer que l'intégration est une construction abstraite, n'utilisant pas la structure particulière de tel ou tel ensemble E: la construction de l'intégrale par rapport à la mesure de Lebesgue sur $\mathbb{R}$ n'est absolument pas plus simple que la théorie générale.

1 Les fonctions mesurables

1) Les définitions: Lors de l'intégration d'une fonction, deux obstacles peuvent se présenter: d'une part la fonction peut être "trop grande"; d'autre part elle peut ne pas être assez "régulière". Ce paragraphe est consacré à la notion de "régularité" nécessaire à la définition de l'intégrale.

Rappelons d'abord que si f est une application d'un espace E dans un espace F, l'*image réciproque* d'une partie A de F par f est la partie de E notée $f^{-1}(A)$ (ou parfois $\{f \in A\}$, ce qui est une notation moins "canonique" mais plus parlante) et définie par

$$f^{-1}(A) \;=\; \{x \in E : f(x) \in A\} \tag{1}$$

(ne pas confondre cette notation avec celle désignant la "fonction réciproque" ou "fonction inverse" de f, lorsque celle-ci est bijective). Les propriétés suivantes, où A et les A_i sont des parties quelconques de F et I est une ensemble fini, dńombrable, ou infini non dénombrable, se vérifient immédiatement:

$$\left.\begin{array}{l} f^{-1}(F) = E, \qquad f^{-1}(\emptyset) = \emptyset, \qquad f^{-1}(A^c) = (f^{-1}(A))^c, \\ f^{-1}(\cup_{i\in I} A_i) = \cup_{i\in I} f^{-1}(A_i), \qquad f^{-1}(\cap_{i\in I} A_i) = \cap_{i\in I} f^{-1}(A_i). \end{array}\right\} \tag{2}$$

On énonce les trois dernières propriétés ci-dessus en disant que l'image réciproque commute avec le passage au complémentaire, la réunion et l'intersection. Si $\mathcal{A}$ est une classe quelconque de parties de F, on note $f^{-1}(\mathcal{A})$ la classe de parties de E définie ainsi: $f^{-1}(\mathcal{A}) = \{f^{-1}(A) : A \in \mathcal{A}\}$. Il découle immédiatement de (2) que:

$$\text{Si } \mathcal{F} \text{ est une tribu de } F\text{, la classe } f^{-1}(\mathcal{F}) \text{ est une tribu de } E. \tag{3}$$

Définition 1 Soit $(E,\mathcal{E})$ et $(F,\mathcal{F})$ deux espaces mesurables, et f une application de E dans F.

a) On dit que f est une application *mesurable* de $(E,\mathcal{E})$ dans $(F,\mathcal{F})$ si la tribu $f^{-1}(\mathcal{F})$ est contenue dans $\mathcal{E}$, ce qui revient à dire que $f^{-1}(A) \in \mathcal{E}$ pour tout $A \in \mathcal{F}$. On écrit aussi parfois: $f : (E,\mathcal{E}) \mapsto (F,\mathcal{F})$.

b) Une fonction sur E (i.e. une application de E dans $\mathbb{R}$ ou dans $\overline{\mathbb{R}}$) est dite *mesurable par rapport à la tribu* $\mathcal{E}$, ou "$\mathcal{E}$-mesurable", ou simplement "mesurable" s'il n'y a pas d'ambiguïté quant à la tribu $\mathcal{E}$, si elle est mesurable de $(E,\mathcal{E})$ dans $\mathbb{R}$ ou $\overline{\mathbb{R}}$ muni de sa tribu borélienne.

c) Lorsque $E = \mathbb{R}^d$ et $F = \mathbb{R}^q$ (ou plus généralement si E et F sont des espaces topologiques), avec leurs tribus boréliennes respectives $\mathcal{E}$ et $\mathcal{F}$, une fonction mesurable de $(E,\mathcal{E})$ dans $(F,\mathcal{F})$ est dite *borélienne.*

d) Si $(f_i)_{i\in I}$ est une famille quelconque de fonctions sur E, on appelle *tribu engendrée* par cette famille, et on note $\sigma(f_i : i \in I)$, la plus petite tribu de E rendant mesurables les fonctions f_i (i.e. la plus petite tribu contenant les tribus $f_i^{-1}(\mathcal{F})$ pour tout $i \in I$). $\square$

Le résultat suivant, que le lecteur vérifiera par lui-même, montre la cohérence entre la mesurabilité d'une fonction et celle d'un ensemble. On rappelle que si $A \subset E$, la fonction indicatrice 1_A de A est la fonction de E dans $\mathbb{R}$ qui vaut 1 sur A et 0 sur le complémentaire A^c:

$$\text{si } A \subset E\text{, on a } A \in \mathcal{E} \text{ si et seulement si } 1_A \text{ est } \mathcal{E}\text{-mesurable.} \tag{4}$$

Exemples:

1) Si E est muni de la tribu $\mathcal{E} = \mathcal{P}(E)$ de toutes ses parties, toute application de E dans un ensemble mesurable $(F,\mathcal{F})$ est mesurable.

2) Si $(E,\mathcal{E})$ est un espace mesurable quelconque, toute fonction constante (i.e. $f(x) = a$ pour tout x, où a est un réel fixé) est mesurable. En effet $f^{-1}(A) = E$ si $a \in A$ et $f^{-1}(A) = \emptyset$ sinon.

2) Critères de mesurabilité: Pour vérifier la mesurabilité d'une fonction, on dispose des trois outils suivants:

Proposition 2 *Soit f une application de E dans F, et soit $\mathcal{A}$ une classe de parties de F telle que $\mathcal{F} = \sigma(\mathcal{A})$* (rappelons que cela signifie que la tribu engendrée par $\mathcal{A}$ est $\mathcal{F}$). *Pour que f soit mesurable de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$ il faut et il suffit que $f^{-1}(A) \in \mathcal{E}$ pour tout $A \in \mathcal{A}$ ($\Leftrightarrow$ $f^{-1}(\mathcal{A}) \subset \mathcal{E}$).*

Preuve. La nécessité est évidente. Inversement, supposons que $f^{-1}(\mathcal{A}) \subset \mathcal{E}$. Soit aussi $\mathcal{A}'$ l'ensemble des parties de F telles que $f^{-1}(A) \in \mathcal{E}$. D'après (2) il est très facile de vérifier que $\mathcal{A}'$ est une tribu de F. Par hypothèse on a $\mathcal{A} \subset \mathcal{A}'$. Comme $\mathcal{A}'$ est une tribu et comme $\mathcal{F}$ est la tribu engendrée par $\mathcal{A}$, on a donc $\mathcal{F} \subset \mathcal{A}'$. Par suite $f^{-1}(\mathcal{F}) \subset \mathcal{E}$ et f est mesurable. □

Proposition 3 *Soit $(E, \mathcal{E})$, $(F, \mathcal{F})$ et $(G, \mathcal{G})$ trois espaces mesurables. Si f est une application mesurable de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$ et si g est une application mesurable de $(F, \mathcal{F})$ dans $(G, \mathcal{G})$, l'application composée $h = g \circ f$ est une application mesurable de $(E, \mathcal{E})$ dans $(G, \mathcal{G})$.*

Preuve. Si $A \in \mathcal{G}$ l'image réciproque $B = g^{-1}(A)$ est dans $\mathcal{F}$ et donc $f^{-1}(B) \in \mathcal{E}$. Comme $h^{-1}(A) = f^{-1}(g^{-1}(A))$, on en déduit $h^{-1}(A) \in \mathcal{E}$, d'où le résultat. □

Proposition 4 *Toute application continue de $E = \mathbb{R}^d$ dans $F = \mathbb{R}^q$ est borélienne. Plus généralement si E et F sont des espaces topologiques, toute application continue de E dans F est borélienne.*

Preuve. a) On va d'abord montrer que si $E = \mathbb{R}^d$ et $F = \mathbb{R}^q$ et si f est une application de E dans F, alors

$$f \text{ est continue} \quad \Leftrightarrow \quad \text{l'image réciproque d'un ouvert de } F \text{ est un ouvert de } E. \tag{5}$$

Supposons d'abord f continue. Rappelons que cela signifie la chose suivante, en notant $|x - x'|_d$ (resp. $|y - y'|_q$) la distance euclidienne de x à x' dans E (resp. de y à y' dans F):

$$\forall x \in E, \ \ \forall \varepsilon > 0, \ \ \exists \eta > 0, \ \ \forall x' \text{ avec } \ |x - x'|_d < \eta, \quad \text{on a} \quad |f(x) - f(x')|_q < \varepsilon. \tag{6}$$

Soit B un ouvert de F et $A = f^{-1}(B)$. Soit $x \in A$ et $y = f(x)$. Comme $y \in B$, il existe un $\varepsilon > 0$ tel que la boule de F centrée en y et de rayon ε soit contenue dans B. Si η est associé à x et ε comme dans (6), cette propriété implique que la boule de E centrée en x et de rayon η est contenue dans A: cela veut exactement dire que A est un ouvert.

Supposons inversement que l'image réciproque de tout ouvert de F par f soit un ouvert de E. Soit $x \in E$ et $\varepsilon > 0$. L'image réciproque de la boule ouverte B de F centrée en $f(x)$ et de rayon ε est un ouvert contenant x, donc il existe $\eta > 0$ tel que $f^{-1}(B)$ contienne la boule de E centrée en x et de rayon η: en d'autres termes, on a (6). Par suite f est continue.

b) Passons à la preuve proprement dite. On a vérifié (5) ci-dessus lorsque $E = \mathbb{R}^d$ et $F = \mathbb{R}^q$. Lorsque E et F sont des espaces topologiques quelconques, (5) est en fait la *définition* des fonctions continues. Si $\mathcal{A}$ (resp. $\mathcal{B}$) désigne la classe des ouverts de E (resp. de F), (5) implique que pour toute fonction continue on a $f^{-1}(\mathcal{B}) \subset \mathcal{A}$. Comme les tribus boréliennes sont les tribus engendrées par les ouverts, le résultat découle immédiatement de la proposition 2. □

On va maintenant donner quelques applications utiles de ces trois résultats.

Proposition 5 *Soit $(E,\mathcal{E})$ un espace mesurable. Pour qu'une fonction f sur E soit mesurable, il faut et il suffit qu'elle vérifie l'une des conditions suivantes:*

(i) $\{f \leq x\} \in \mathcal{E}$ pour tout $x \in \mathbb{R}$ (rappelons que $\{f \leq x\} = f^{-1}([-\infty,x]) = \{y \in E : f(y) \leq x\}$).

(ii) $\{f \leq x\} \in \mathcal{E}$ pour tout $x \in \mathbb{Q}$.

(iii) $\{f < x\} \in \mathcal{E}$ pour tout $x \in \mathbb{R}$.

(iv) $\{f < x\} \in \mathcal{E}$ pour tout $x \in \mathbb{Q}$.

Preuve. Il suffit de combiner les propositions 1-7 et 2. □

Proposition 6 *Soit $f_1,...,f_d$ des fonctions réelles mesurables sur $(E,\mathcal{E})$. Soit g une fonction borélienne sur $\mathbb{R}^d$. La fonction h sur E définie par $h(x) = g(f_1(x), f_2(x), ..., f_d(x))$ est alors mesurable sur $(E,\mathcal{E})$.*

Preuve. On peut considérer le d-uplet $(f_1,...,f_d)$ comme une application de E dans $\mathbb{R}^d$, qu'on notera f: si $x \in E$, $f(x)$ est le vecteur de $\mathbb{R}^d$ dont les composantes sont $f_1(x),...,f_d(x)$. Comme $h = g \circ f$, en vertu de la proposition 3 il suffit de démontrer que f est mesurable de $(E,\mathcal{E})$ dans $(\mathbb{R}^d,\mathcal{R}^d)$.

Pour cela, en utilisant 1-(12) et la proposition 2, on voit qu'il suffit de montrer que pour tout rectangle $A = \prod_{i=1}^d]-\infty, a_i]$, où les a_i sont des réels, on a $f^{-1}(A) \in \mathcal{E}$. Mais comme $f^{-1}(A) = \cap_{1\leq i\leq d}\{f_i \leq a_i\}$ cette propriété découle de la mesurabilité des f_i et de la propriété (T'4) des tribus. □

Ce résultat s'applique en particulier lorsque la fonction g ci-dessus est continue. Cela donne une série de propriétés d'usage constant. Par exemple si les fonctions réelles f_i sont mesurables sur $(E,\mathcal{E})$, il en est de même des fonctions suivantes:

$$\sum_{i=1}^d a_i f_i, \quad \text{où les } a_i \text{ sont réels.} \tag{7}$$

$$\prod_{i=1}^d (f_i)^{a_i}, \quad \text{où } \ a_i \in \mathbb{Z}, \text{ et } \ a_i > 0 \ \text{ si } \ f_i \text{ peut s'annuler.} \tag{8}$$

$$f_1 \wedge f_2 = \min(f_1, f_2), \qquad f_1 \vee f_2 = \max(f_1, f_2). \tag{9}$$

(Pour (7) par exemple, il suffit d'appliquer la proposition précédente avec $g(x_1,...,x_d) = \sum_{i=1}^d a_i x_i$, qui est continue). On déduit de ces propriétés que l'ensemble de toutes les fonctions réelles mesurables sur $(E,\mathcal{E})$ est une *algèbre* (i.e. un espace vectoriel stable par produit des fonctions), et un espace *réticulé* (i.e. stable par les opérations "sup" et "inf"); on verra mieux dans la proposition 8 ci-dessous.

En particulier $g = f_1 - f_2$ est une fonction mesurable, et donc les ensembles suivants

$$\{f_1 = f_2\} = \{g = 0\}, \qquad \{f_1 < f_2\} = \{g < 0\}, \qquad \{f_1 \leq f_2\} = \{g \leq 0\} \tag{10}$$

sont mesurables.

3) Les limites de fonctions mesurables: Chacun sait qu'une suite $(f_n)_{n\geq 1}$ de fonctions sur E et à valeurs dans $\mathbb{R}$ ou dans $\overline{\mathbb{R}}$ converge *simplement* vers une limite f si $f_n(x) \to f(x)$ pour tout x. Lorsque la suite de fonctions est quelconque, on peut toujours introduire les notions suivantes:

Définition 7 On appelle *limite supérieure* et *limite inférieure* d'une suite $(f_n)_{n\geq 1}$ de fonctions sur E et à valeurs dans $\overline{\mathbb{R}}$ les fonctions suivantes:

$$\left.\begin{array}{rcccl} \limsup_n f_n(x) & = & \lim_n \downarrow \sup_{m\geq n} f_m(x) & = & \inf_n \sup_{m\geq n} f_m(x), \\ \liminf_n f_n(x) & = & \lim_n \uparrow \inf_{m\geq n} f_m(x) & = & \sup_n \inf_{m\geq n} f_m(x). \end{array}\right\} \tag{11}$$

Noter que les fonctions $\limsup_n f_n$ et $\liminf_n f_n$ définies ci-dessus sont a-priori à valeurs dans $\overline{\mathbb{R}}$, même si les f_n sont à valeurs dans $\mathbb{R}$.

Rappelons qu'une suite de fonction $(f_n)_n$ *converge simplement vers la limite* f si on a $f_n(x) \to f(x)$ pour tout x. Si la suite $(f_n)_n$ est croissante (resp. décroissante), c'est-à-dire si $f_n \leq f_{n+1}$ (resp. $f_n \geq f_{n+1}$) pour tout n, elle converge simplement vers une limite f vérifiant $f = \limsup_n f_n = \liminf_n f_n$ et aussi $f = \sup_n f_n$ (resp. $f = \inf_n f_n$). Dans le cas général, dire que la suite (f_n) converge simplement revient à dire que $\limsup_n f_n = \liminf_n f_n$, et dans ce cas la valeur commune de ces deux fonctions est la limite de la suite (f_n). La propriété suivante est immédiate:

$$\limsup_n f_n \;=\; -\liminf_n (-f_n), \tag{12}$$

et si les $(A_n)_{n\geq 1}$ sont des parties de E, en se rappelant la définition 1-5 on a:

$$\limsup_n 1_{A_n} \;=\; 1_{\limsup_n A_n}, \qquad \liminf_n 1_{A_n} \;=\; 1_{\liminf_n A_n}. \tag{13}$$

Proposition 8 *Soit $(f_n)_{n\geq 1}$ une suite de fonctions mesurables sur $(E, \mathcal{E})$, à valeurs dans $\mathbb{R}$ ou dans $\overline{\mathbb{R}}$.*

a) Les fonctions $\sup_n f_n$ et $\inf_n f_n$ sont mesurables.

b) Les fonctions $\limsup_n f_n$ et $\liminf_n f_n$ sont mesurables.

c) L'ensemble des $x \in E$ où la suite numérique $(f_n(x))$ converge (dit "ensemble de convergence" de la suite (f_n)) est dans $\mathcal{E}$.

d) Si la suite (f_n) converge simplement, sa limite est une fonction mesurable.

Preuve. Pour (a) on utilise le fait que $\{\sup_n f_n \leq x\} = \cap_n \{f_n \leq x\}$ et $\{\inf_n f_n < x\} = \cup_n \{f_n < x\}$ et la proposition 5. (b) s'obtient par application répétée de (a). Si

$g = \limsup_n f_n$ et $h = \liminf_n f_n$, l'ensemble de convergence de la suite (f_n) est l'ensemble $\{g = h\}$, qui est mesurable d'après (10). Enfin si (f_n) converge simplement sa limite est égale à $g = h$, donc (d) découle de (b). $\square$

4) Image d'une mesure par une application: Ci-dessous on considère d'une part une application mesurable de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$, et d'autre part une mesure μ sur $(E, \mathcal{E})$. On peut "transporter" la mesure μ sur F par f, selon le schéma suivant:

Théorème 9 *Si pour tout $B \in \mathcal{F}$ on pose*

$$\nu(B) \;=\; \mu(f^{-1}(B)), \tag{14}$$

on définit une mesure ν sur $(F, \mathcal{F})$, appelée la mesure image *de μ par f.*

Preuve. On utilise (2): d'une part, $\nu(\emptyset) = \mu(\emptyset) = 0$. D'autre part si on a une suite $(B_n)_{n \geq 1}$ de parties deux-à-deux disjointes et appartenant à $\mathcal{F}$, les $A_n = f^{-1}(B_n)$ sont aussi deux-à-deux disjointes, tandis que $\cup_n A_n = f^{-1}(\cup_n B_n)$. Par suite

$$\nu(\cup_n B_n) \;=\; \mu(f^{-1}(\cup_n B_n)) \;=\; \mu(\cup_n A_n) \;=\; \sum_n \mu(A_n) \;=\; \sum_n \nu(B_n). \quad \square$$

2 L'intégrale des fonctions mesurables

Nous fixons ci-dessous un espace E muni d'une tribu $\mathcal{E}$ et d'une mesure μ. On appelle $\mathcal{F}$ l'ensemble de toutes les fonctions réelles mesurables sur $(E, \mathcal{E})$: c'est un espace vectoriel d'après (7).

Nous nous proposons de définir l'intégrale d'une fonction f par rapport à μ, notée $\int f d\mu$, pour une classe aussi grande que possible de fonctions de $\mathcal{F}$. Cette intégrale devra avoir les propriétés suivantes:

$$\int 1_A d\mu = \mu(A) \qquad \text{si} \;\; A \in \mathcal{E}, \tag{15}$$

$$\left.\begin{array}{l} \text{L'application } f \mapsto \int f d\mu \text{ est "linéaire", i.e. } \int (af) d\mu = a \int f d\mu \\ \text{si } a \in I\!\!R, \text{ et } \int (f+g) d\mu = \int f d\mu + \int g d\mu, \end{array}\right\} \tag{16}$$

ainsi que des propriétés de "continuité" qui seront précisées plus loin.

Le principe de la construction, qui se fait en plusieurs étapes, est assez simple:

1) En combinant (15) et (16), on construit $\int f d\mu$ pour les fonctions f positives mesurables ne prenant qu'un nombre fini de valeurs.

2) Toute fonction positive mesurable étant limite croissante d'une suite de fonctions du type précédent, on obtient son intégrale par passage à la limite.

3) Toute fonction mesurable étant différence de deux fonctions mesurables positives, on construit son intégrale par différence.

1) Les fonctions étagées: On dit qu'une fonction est *étagée* si elle ne prend qu'un nombre fini de valeurs dans $\overline{I\!R}$. On note $\mathcal{F}_+^0$ l'ensemble de toutes les fonctions étagées positives mesurables. Cet ensemble n'est pas un espace vectoriel (c'est seulement ce qu'on appelle un "cône"), mais il est stable par addition, et par multiplication par les réels positifs (et par $+\infty$: rappelons les conventions 1-(1) et 1-(2)).

Etant donnés les nombres $a_1, \ldots, a_n$ de $\overline{I\!R}_+$ et les ensembles mesurables $A_1, \ldots, A_n$, on obtient une fonction $f \in \mathcal{F}_+^0$ en posant

$$f = \sum_{i=1}^{n} a_i 1_{A_i}. \tag{17}$$

(il est clair que cette fonction ne peut prendre que les valeurs qui sont des sommes d'un nombre quelconque de a_i, donc ne prend qu'un nombre fini de valeurs; d'autre part f est mesurable par (4) et (7)). Il y a évidemment plusieurs manières d'écrire la même fonction f sous la forme (17).

Inversement, toute $f \in \mathcal{F}_+^0$ s'écrit sous cette forme, et même admet une écriture (17) "canonique" qui est unique et qui a la forme suivante: Si U est l'ensemble des valeurs prises par f, la famille $A_a = \{f = a\}$ indicée par l'ensemble fini U (i.e. a parcourt U) constitue une partition mesurable de E, et on a

$$f = \sum_{a \in U} a 1_{A_a}. \tag{18}$$

Cette écriture est un cas particulier de (17).

Définition 10 *Par définition, on appelle intégrale par rapport à μ de la fonction $f \in \mathcal{F}_+^0$ admettant la décomposition canonique (18), et on note $\int f d\mu$ ou $\int f(x)\mu(dx)$, le nombre suivant de $[0, \infty]$:*

$$\int f d\mu = \sum_{a \in U} a\mu(A_a) = \sum_{a \in U} a\mu(\{f = a\}). \quad \square \tag{19}$$

Exemples:

1) L'intégrale de la fonction nulle (qui appartient à $\mathcal{F}_+^0$) est 0.

2) L'intégrale de la fonction constante égale à $a \geq 0$ (qui appartient aussi à $\mathcal{F}_+^0$) vaut $a\mu(E)$ (donc vaut $+\infty$ si la mesure μ est de masse totale infinie, ou si $a = +\infty$ et μ n'est pas la mesure nulle).

3) Rappelons que $f = 1_A$ est dans $\mathcal{F}_+^0$ si et seulement si $A \in \mathcal{E}$. Dans ce cas son intégrale est $\mu(A)$: on a donc (14). $\square$

Proposition 11 *(i) Si $f \in \mathcal{F}_+^0$ est donnée par (17), on a*

$$\int f d\mu = \sum_{i=1}^{n} a_i \mu(A_i) \tag{20}$$

(ii) *Si* $a \geq 0$ *et* $f \in \mathcal{F}_+^0$, *on a* $\int (af) d\mu = a \int f d\mu$.

(iii) *Si* $f, g \in \mathcal{F}_+^0$, *on a* $\int (f+g) d\mu = \int f d\mu + \int g d\mu$.

(iv) *Si* $f, g \in \mathcal{F}_+^0$ *et* $f \leq g$, *on a* $\int f d\mu \leq \int g d\mu$.

Preuve. (ii) est évident. Pour montrer (iii), notons U et V les ensembles (finis) de valeurs prises par f et g respectivement, ainsi que $A_a = \{f = a\}$ pour $a \in U$ et $B_b = \{g = b\}$ pour $b \in V$. Remarquons que si $a \in U$ l'ensemble A_a est la réunion des ensembles mesurables deux-à-deux disjoints $(A_a \cap B_b)_{b \in V}$ (certains de ces ensembles peuvent être vides). De même B_b est la réunion des ensembles mesurables deux-à-deux disjoints $(A_a \cap B_b)_{a \in U}$. D'après (19) et l'additivité (A) de μ on a donc

$$\int f d\mu = \sum_{a \in U} a\mu(A_a) = \sum_{a \in U, b \in V} a\mu(A_a \cap B_b),$$

$$\int g d\mu = \sum_{b \in V} b\mu(B_b) = \sum_{a \in U, b \in V} b\mu(A_a \cap B_b).$$

En additionnant, il vient

$$\int f d\mu + \int g d\mu = \sum_{a \in U, b \in V} (a+b)\mu(A_a \cap B_b). \tag{21}$$

Par ailleurs notons W l'ensemble des valeurs prises par $h = f + g$. Tout point c de W s'écrit $c = a + b$ pour une certaines famille (finie) I_c de couples (a, b) dans le produit $U \times V$ (noter que I_c peut contenir un ou plusieurs couples). L'ensemble $C_c = \{h = c\}$ est alors la réunion des ensembles deux-à-deux disjoints $(A_a \cap B_b)_{(a,b) \in I_c}$, de sorte que

$$\int h d\mu = \sum_{c \in W} c\mu(C_c) = \sum_{c \in W} \sum_{(a,b) \in I_c} c\mu(A_a \cap B_b). \tag{22}$$

Si le couple $(a, b) \in U \times V$ n'appartient à aucun I_c on a $A_a \cap B_b = \emptyset$, de sorte que $\mu(A_a \cap B_b) = 0$. Comme $c = a + b$ lorsque $(a, b) \in I_c$, il est alors facile de vérifier que les expressions (21) et (22) sont égales: on a donc (iii).

Pour obtenir (i), il suffit alors d'appliquer (ii), (iii) et (14). Enfin si $f, g \in \mathcal{F}_+^0$ et si $f \leq g$, la fonction $h = g - f$ est aussi dans $\mathcal{F}_+^0$. Par (iii) on a $\int g d\mu = \int f d\mu + \int h d\mu$. Comme $\int h d\mu \geq 0$ par constrution (cf. (19)), on obtient (iv). $\square$

Proposition 12 *Soit* $(f_n)_{n \geq 1}$ *une suite croissante (i.e.* $f_n \leq f_{n+1}$ *pour tout* n*) de fonctions de* $\mathcal{F}_+^0$ *et* $f(x) = \lim_n \uparrow f_n(x)$ noter que f n'est pas nécessairement étagée).

(i) Si $g \in \mathcal{F}_+^0$ *vérifie* $g \leq f$, *on a* $\int g d\mu \leq \lim_n \uparrow \int f_n d\mu$.

(ii) Si de plus $f \in \mathcal{F}_+^0$, *on a* $\int f d\mu = \lim_n \uparrow \int f_n d\mu$.

Preuve. D'après (iv) de la proposition précédente la suite $\alpha_n = \int f_n d\mu$ est croissante, et on note α sa limite.

(i) Soit $g \in \mathcal{F}_+^0$ avec $g \leq f$. Soit $\varepsilon \in]0, 1[$ fixé. La fonction $g' = (1 - \varepsilon)g$ vérifie $g' \in \mathcal{F}_+^0$, $g' \leq f$, et $g'(x) < f(x)$ si $f(x) > 0$.

Soit U l'ensemble des valeurs prises par g'. Pour tout $a \in U$ on a $a1_{\{g'=a\leq f_n\}} \leq f_n 1_{\{g'=a\}}$; donc en appliquant les assertions (i) et (iv) de la proposition précédente, on obtient

$$a\mu(\{g' = a \leq f_n\}) \;=\; \int (a1_{\{g'=a\leq f_n\}})d\mu \;\leq\; \int (f_n 1_{\{g'=a\}})d\mu.$$

Comme $\sum_{a\in U} f_n 1_{\{g'=a\}} = f_n$, en sommant les inégalités ci-dessus pour tous les $a \in U$ et en utilisant (iii) de la proposition 11, il vient

$$\sum_{a\in U} a\mu(\{g' = a \leq f_n\}) \;\leq\; \int \sum_{a\in U} (f_n 1_{\{g'=a\}})d\mu \;=\; \alpha_n.$$

Rappelons que si $f(x) = 0$ on a $g'(x) = f_n(x) = 0$ pour tout n, tandis que si $f(x) > 0$ on a $g'(x) < f(x)$ et donc $g'(x) < f_n(x)$ pour n assez grand (dépendant de x). Par suite $\{g' = a \leq f_n\} \uparrow \{g' = a\}$ quand n croît vers l'infini. Donc en utilisant le théorème 14, on obtient en passant à la limite dans l'inégalité précédente:

$$\int g' d\mu \;=\; \sum_{a\in U} a\mu(\{g' = a\}) \;\leq\; \alpha.$$

Enfin comme $g = \frac{g'}{1-\varepsilon}$ on a $\int g d\mu = \frac{1}{1-\varepsilon}\int g' d\mu \leq \frac{\alpha}{1-\varepsilon}$. Comme ε est arbitrairement proche de 0 et comme $\lim_{\varepsilon\downarrow 0}\frac{\alpha}{1-\varepsilon} = \alpha$, on en déduit finalement que $\int g d\mu \leq \alpha$.

(ii) Si maintenant $f \in \mathcal{F}^0_+$, (i) appliqué à $g = f$ montre que $\int f d\mu \leq \alpha$. Par ailleurs $f_n \leq f$, donc $\alpha_n \leq \int f d\mu$ pour tout n, et en passant à la limite on obtient $\alpha \leq \int f d\mu$. Par suite $\int f d\mu = \alpha$. $\square$

2) Les fonctions positives: Dans la suite on note $\mathcal{F}_+$ l'ensemble des fonctions mesurables à valeurs dans $\overline{\mathbb{R}}_+$

Lemme 13 *Toute fonction f de $\mathcal{F}_+$ est limite simple d'une suite croissante $(f_n)_{n\geq 1}$ de fonctions mesurables positives étagées (i.e. $f(x) = \lim_n \uparrow f_n(x)$ pour tout $x \in E$).*

Preuve. Il suffit de poser:

$$f_n(x) \;=\; \begin{cases} \frac{k}{2^n} & \text{si } \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \text{ et } k = 0, 1, \ldots, n2^n - 1, \\ n & \text{si } f(x) \geq n. \end{cases} \qquad \square$$

Définition 14 On appelle *intégrale* par rapport à μ de la fonction $f \in \mathcal{F}_+$ le nombre suivant de $[0, \infty]$:

$$\int f d\mu \;=\; \int f(x)\mu(dx) \;=\; \sup(\int g d\mu : g \in \mathcal{F}^0_+, g \leq f). \;\; \square \tag{23}$$

Lemme 15 *Si $f \in \mathcal{F}_+$, toute suite croissante $(f_n)_{n\geq 1}$ de fonctions de $\mathcal{F}^0_+$ admettant f pour limite* (il existe de telles suites d'après le lemme 13) *vérifie $\int f d\mu = \lim_n \uparrow \int f_n d\mu$.*

Preuve. La suite de nombres $\alpha_n = \int f_n d\mu$ croît vers une limite $\alpha \in [0, \infty]$. D'après (23) on a $\alpha_n \leq \int f d\mu$, donc aussi $\alpha \leq \int f d\mu$. A l'inverse, toute fonction $g \in \mathcal{F}_+^0$ telle que $g \leq f$ vérifie $\int g d\mu \leq \alpha$ par la proposition 12, de sorte que $\int f d\mu \leq \alpha$ en vertu de (23): on en déduit que $\alpha = \int f d\mu$. $\square$

Nous pouvons maintenant énoncer l'un des résultats essentiels de la théorie:

Théorème 16 *(i) Si $a \in I\!R_+$ et $f \in \mathcal{F}_+$, on a $\int (af) d\mu = a \int f d\mu$.*

(ii) Si $f, g \in \mathcal{F}_+$ on a $\int (f+g) d\mu = \int f d\mu + \int g d\mu$.

(iii) Si $f, g \in \mathcal{F}_+$ et si $f \leq g$, on a $\int f d\mu \leq \int g d\mu$.

(iv) (THEOREME DE CONVERGENCE MONOTONE) *Si la suite $(f_n)_{n\geq 1}$ de fonctions de $\mathcal{F}_+$ croît vers une limite f* (nécessairement dans $\mathcal{F}_+$)*, alors la suite $(\int f_n d\mu)_{n\geq 1}$ croît vers $\int f d\mu$.*

(v) Pour toute suite $(f_n)_{n\geq 1}$ de fonctions de $\mathcal{F}_+$ on a

$$\int (\inf_n f_n) d\mu \ \leq \ \inf_n \int f_n d\mu, \qquad \int (\sup_n f_n) d\mu \ \geq \ \sup_n \int f_n d\mu. \tag{24}$$

(vi) Pour toute suite $(f_n)_{n\geq 1}$ de fonctions de $\mathcal{F}_+$ on a

$$\int (\liminf_n f_n) d\mu \ \leq \ \liminf_n \int f_n d\mu. \tag{25}$$

Attention: (vi) est une version de ce qu'on appelle le *lemme de Fatou* (on en verra une forme plus générale plus loin). Contrairement à ce que pourrait faire penser (24), dans lequel "sup" et "inf" jouent des rôles analogues, on n'a pas dans (vi) l'inégalité en sens opposé en remplaçant "liminf" par "limsup": si par exemple μ est une mesure de masse totale infinie et si $f_n(x) = 1/n$, on a $\limsup_n f_n = \liminf_n f_n = f$, avec $f(x) = 0$ pour tout x; donc $\int \limsup_n f_n d\mu = \int \liminf_n f_n d\mu = 0$; cependant $\int f_n d\mu = \infty$ pour tout n, donc $\limsup_n \int f_n d\mu = \liminf_n \int f_n d\mu = \infty$.

Preuve. Pour (i), (ii) et (iii) On considère des suites (f_n) et (g_n) de fonctions de $\mathcal{F}_+^0$ croissant respectivement vers f et g. On a $f_n + g_n \in \mathcal{F}_+^0$ et $f_n + g_n \uparrow f + g$, donc le lemme 15 et les assertions (ii), (iii) et (iv) de la proposition 11 impliquent (i), (ii) et (iii).

(iv) D'après (iii), la suite $\alpha_n = \int f_n d\mu$ croît vers une limite α et vérifie $\alpha_n \leq \int f d\mu$, de sorte que $\alpha \leq \int f d\mu$. Pour chaque n il existe une suite croissante $(g_{n,i})_{i\geq 1}$ de fonctions de $\mathcal{F}_+^0$ telle que $\lim_i \uparrow g_{n,i} = f_n$. On pose $h_i = \sup_{n:1\leq n\leq i} g_{n,i}$. Chaque h_i est dans $\mathcal{F}_+^0$; on a $g_{n,i} \leq g_{n,i+1}$, donc $h_i \leq h_{i+1}$ et la suite (h_i) croît vers une limite h quand i tend vers l'infini; comme $g_{n,i} \leq f$ on a $h_i \leq f$ et donc $h \leq f$; enfin $h_i \geq g_{n,i}$ pour tout $i \geq n$, donc $h \geq f_n$ pour tout n, donc $h \geq f$: on en déduit finalement que (h_i) est une suite croissante de fonctions de $\mathcal{F}_+^0$ admettant la limite $h = f$.

On a donc $\int h_i d\mu \uparrow \int f d\mu$ quand i tend vers l'infini, d'après le lemme 15. Mais $h_i \leq \sup_{n:1\leq n\leq i} f_n = f_i$, de sorte que $\int h_i d\mu \leq \alpha_i$. Par suite en passant à la limite en i on obtient $\int f d\mu \leq \alpha$: donc $\alpha = \int f d\mu$ et le résultat est démontré.

(v) Soit $g = \inf_n f_n$ et $h = \sup_n f_n$, qui sont des fonctions de $\mathcal{F}_+$ Pour tout n on a $g \leq f_n \leq h$, donc $\int g d\mu \leq \int f_n d\mu \leq \int h d\mu$ par (iii), et (24) est immédiat.

(vi) Si $g_n = \inf_{i \geq n} f_i$, on a $\int g_n d\mu \leq \inf_{i \geq n} \int f_n d\mu$ d'après (v). Lorsque n tend vers l'infini, les nombres $\inf_{i \geq n} \int f_n d\mu$ croissent vers le nombre $\liminf_n \int f_n d\mu$. Par ailleurs la suite (g_n) croît vers la fonction $\liminf_n f_n$, donc (iv) implique que $\int g_n d\mu$ croît vers $\int \liminf_n f_n d\mu$. L'inégalité (25) est alors immédiate. $\square$

Lorsque les f_n sont des fonctions mesurables positives, en appliquant (iv) ci-dessus aux fonctions $g_n = f_1 + \ldots + f_n$ on obtient le

Corollaire 17 *Si les $(f_n)_{n \geq 1}$ sont des fonctions mesurables positives, on a $\int (\sum_n f_n) d\mu = \sum_n \int f_n d\mu$ (on peut "intervertir" somme d'une série et intégrale, lorsque les termes sont positifs).*

Exemple: Si $(u_{n,i})_{n,i \geq 1}$ est une double suite de nombres positifs, un résultat bien connu de la théorie des séries affirme que

$$\sum_{n \geq 1} \sum_{i \geq 1} u_{n,i} \;=\; \sum_{i \geq 1} \sum_{n \geq 1} u_{n,i} \tag{26}$$

(appelé "interversion des sommations", ou encore "sommation par paquets"). Ce résultat est aussi une conséquence du corollaire précédent: en effet, soit $E = \mathbb{N}^*$, muni de la tribu $\mathcal{E}$ de toutes les parties et de la mesure de comptage μ (i.e. $\mu(A)$ est le nombre de point de A). Noter que toute fonction sur E est $\mathcal{E}$-mesurable. La formule ci-dessus provient alors du corollaire, si on pose $f_n(i) = u_{n,i}$. $\square$

3) Les fonctions de signe quelconque: Il nous reste à définir l'intégrale des fonctions de signe quelconque. Pour cela, on utilise le fait qu'une fonction f est toujours la différence $f = g - h$ de deux fonctions positives, cette décomposition n'étant bien-sûr pas unique. On verra ci-dessous que si f est mesurable, on peut choisir g et h mesurables également. L'idée consiste à définir $\int f d\mu$ comme la différence $\int g d\mu - \int h d\mu$: mais pour que cela ait un sens, il ne faut pas que la différence ci-dessus soit $\infty - \infty$.

On a donc intérêt à choisir g et h ci-dessus aussi petites que possibles (car si on augmente g, on augmente h de la même quantité pour préserver l'égalité $g - h = f$, et donc on augmente les intégrales de g et h). Le choix "minimal" est le suivant:

$$f^+(x) \;=\; \sup(0, f(x)), \qquad f^-(x) \;=\; \sup(0, -f(x)), \tag{27}$$

de sorte qu'on a

$$f = f^+ - f^-, \qquad |f| = f^+ + f^-. \tag{28}$$

f^+ et f^- sont ce qu'on appelle les *parties positive et négative* de f, et toute autre décomposition $f = g - h$ avec g et h positives vérifie $g \geq f^+$ et $h \geq f^-$. Remarquer aussi que si f est mesurable, alors f^+ et f^- sont mesurables par (9). Avec ces notations, on peut enfin donner la définition de l'intégrale dans le cas général:

Définition 18 a) On dit que la fonction mesurable f à valeurs dans $\overline{I\!\!R}$ *admet une intégrale* par rapport à μ, ou que "son intégrale existe", si on n'a pas à la fois $\int f^+ d\mu = \infty$ et $\int f^- d\mu = \infty$; dans ce cas *l'intégrale* de f est le nombre

$$\int f d\mu \;=\; \int f(x)\mu(dx) \;=\; \int f^+ d\mu - \int f^- d\mu. \qquad (29)$$

b) On dit que la fonction mesurable f est *intégrable* par rapport à μ (ou: μ-intégrable) si l'intégrale $\int |f| d\mu$ est finie. Ceci équivaut à dire que les intégrales de f^+ et f^- sont finies (utiliser (28) et le théorème 16-(ii)), de sorte que l'intégrale $\int f d\mu$ existe et est finie.

c) Finalement on note $\mathcal{L}^1(E, \mathcal{E}, \mu)$ (ou plus simplement $\mathcal{L}^1$) l'ensemble des fonctions à valeurs dans $I\!\!R$, mesurables et intégrables. $\square$

Cette terminologie est un peu malheureuse, puisqu'une fonction peut ne pas être intégrable, et cependant avoir une intégrale (qui vaut alors nécessairement $-\infty$ ou $+\infty$). Si f admet une intégrale, elle est intégrable si et seulement si son intégrale est finie. Avant de donner les principales propriétés de l'intégrale, voici quelques exemples.

Exemples:

1) Soit $(E, \mathcal{E})$ un espace mesurable quelconque, et $\mu = \varepsilon_a$ la mesure de Dirac au point a (rappelons que $\mu(A)$ vaut 1 ou 0 selon que a est dans A ou non). Il est facile de vérifier que toute fonction mesurable f admet une intégrale, qui vaut $\int f d\mu = f(a)$. Les fonctions intégrables sont celles qui vérifient $f(a) \in I\!\!R$ (elles peuvent prendre les valeurs $+\infty$ et $-\infty$ en dehors de a).

2) Soit $E = \{1, \ldots, k\}$, muni de la tribu de toutes les parties et de la mesure de comptage μ. On a déjà dit que toute fonction sur E est mesurable, et évidemment toute fonction ne prend qu'un nombre fini de valeurs. Ainsi $\mathcal{F}_+^0 = \mathcal{F}_+$ est l'ensemble des fonctions à valeurs dans $\overline{I\!\!R}_+$.

   Dans cet exemple, une fonction est intégrable si et seulement si elle est à valeurs dans $I\!\!R$. Une fonction admet une intégrale si et seulement si elle est à valeurs dans $]-\infty, \infty]$ ou dans $[-\infty, \infty[$. Dans tous ces cas, on a $\int f d\mu = \sum_{i=1}^k f(i)$

3) Soit $E = I\!\!N^*$, muni de la tribu de toutes les parties et de la mesure de comptage μ. Une fonction f sur E peut être identifiée à la suite $(u_n = f(n))_{n \geq 1}$ des valeurs qu'elle prend, et là encore toute fonction sur E est mesurable. Si f est une fonction positive, on peut construire une suite particulière $(f_n)_{n \geq 1}$ de fonctions étagées croissant vers f en posant:

$$f_n(i) \;=\; \begin{cases} f(i) & \text{si} \quad i \leq n, \\ 0 & \text{si} \quad i > n. \end{cases}$$

   D'après (20) on a $\int f_n d\mu = \sum_{i=1}^n f(i)$, et le lemme 15 implique que $\int f d\mu = \sum_{i \geq 1} f(i)$: l'intégrale de f est ainsi la somme de la série de terme général $f(i)$.

   La définition 18 entraine alors qu'une fonction f (de signe quelconque) est *intégrable* si et seulement si la série de terme général $f(i)$ est *absolument convergente*, et dans ce cas $\int f d\mu = \sum_{i \geq 1} f(i)$. Notons qu'on retrouve ici, en particulier, la propriété (S5) du chapitre 1.

La fonction f n'est pas intégrable, mais admet une intégrale, si et seulement si on est dans l'un des cas suivants:

(a) $\sum_{i:f(i)<0}|f(i)| < \infty$ et $\sum_{i:f(i)>0} f(i) = \infty$, auquel cas $\int f d\mu = +\infty$,

(b) $\sum_{i:f(i)>0} f(i) < \infty$ et $\sum_{i:f(i)<0}|f(i)| = \infty$, auquel cas $\int f d\mu = -\infty$. $\square$

Théorème 19 (i) *L'ensemble $\mathcal{L}^1(E,\mathcal{E},\mu)$ de toutes les fonctions qui sont à valeurs réelles et qui sont mesurables et intégrables, est un espace vectoriel.*

(ii) *L'application $f \mapsto \int f d\mu$ de $\mathcal{L}^1(E,\mathcal{E},\mu)$ dans $I\!R$ est une forme linéaire positive:* on rappelle que cela veut dire que c'est une application linéaire de $\mathcal{L}^1(E,\mathcal{E},\mu)$ dans $I\!R$, i.e. $\int(f+g)d\mu = \int f d\mu + \int g d\mu$ et $\int(af)d\mu = a\int f d\mu$ si $a \in I\!R$, et qu'elle est en outre “positive” au sens où $\int f d\mu \geq 0$ si $f \geq 0$

(iii) *Pour toute fonction f de $\mathcal{L}^1(E,\mathcal{E},\mu)$ on a*

$$\left|\int f d\mu\right| \;\leq\; \int |f| d\mu. \tag{30}$$

(iv) *Enfin si $f \in \mathcal{L}^1(E,\mathcal{E},\mu)$ et si g est mesurable et vérifie $|g| \leq |f|$, alors $g \in \mathcal{L}^1(E,\mathcal{E},d\mu)$.*

Avant de prouver ce théorème on va énoncer un lemme de “linéarité” qui généralise la propriété (29) et qui concerne les fonctions admettant une intégrale sans être nécessairement intégrables.

Lemme 20 *Soit $f = g - h$ la différence de deux fonctions g et h de $\mathcal{F}_+$. Si l'une des deux intégrales $\int g d\mu$ ou $\int h d\mu$ au moins est finie, alors f admet une intégrale, qui vaut $\int f d\mu = \int g d\mu - \int h d\mu$.*

Preuve. Supposons par exemple que $\int g d\mu < \infty$. D'une part $f^+ \leq g$, d'autre part $f^+ + h = f^- + g$. Donc le théorème 16 implique d'une part $\int f^+ d\mu \leq \int g d\mu < \infty$, et d'autre part

$$\int f^+ d\mu + \int h d\mu \;=\; \int f^- d\mu + \int g d\mu.$$

On en déduit que $\int f d\mu$ est bien défini par la formule (29), à valeurs dans $[-\infty, \infty[$, et que

$$\int h d\mu \;=\; -\int f^+ d\mu + \int f^- d\mu + \int g d\mu \;=\; -\int f d\mu + \int g d\mu,$$

d'où le résultat. $\square$

Preuve du théorème 19. Si $f \geq 0$ on a $f = f^+$ et $f^- = 0$, donc $\int f d\mu = \int f^+ d\mu \geq 0$.

Si $a \in I\!R_+$ on a $(af)^+ = af^+$ et $(af)^- = af^-$. Donc le théorème 16-(i) et la définition 18 impliquent $af \in \mathcal{L}^1$ et $\int(af)d\mu = a\int f d\mu$. Si maintenant $a \in]-\infty, 0[$, on a $(af)^+ = -af^- = |a|f^-$ et $(af)^- = -af^+ = |a|f^+$: on en déduit par les mêmes arguments que $af \in \mathcal{L}^1$ et que $\int(af)d\mu = a\int f d\mu$.

Soit maintenant $f, g \in \mathcal{L}^1$. D'abord $|f+g| \leq |f| + |g|$, donc le théorème 16-(ii,iii) implique $f + g \in \mathcal{L}^1$: cela termine la preuve du fait que $\mathcal{L}^1$ est un espace vectoriel. Ensuite

$f+g = f^+ + g^+ - f^- - g^-$ et les fonctions du second membre ci-dessus sont toutes d'intégrale finie. Le lemme précédent entraine alors

$$\int (f+g)d\mu \;=\; \int f^+ d\mu + \int g^+ d\mu - \int f^- d\mu - \int g^- d\mu \;=\; \int f d\mu + \int g d\mu.$$

On a donc achevé la preuve de la linéarité et de la positivité de $f \mapsto \int f d\mu$.

Pour tous $a, b \in I\!\!R_+$ on a $|a-b| \le a+b$, donc en utilisant (28) on obtient

$$|\int f d\mu| \;=\; |\int f^+ d\mu - \int f^- d\mu| \;\le\; \int f^+ d\mu + \int f^- d\mu \;=\; \int |f| d\mu,$$

donc on a (30). Enfin la dernière assertion découle du théorème 16-(iii). $\square$

Nous terminons par des résultats de "continuité" concernant l'intégrale. Il s'agit des résultats *essentiels* de la théorie, qui doivent absolument être assimilés. Ils seront encore améliorés plus loin, mais vu leur importance il ne faut pas lésiner sur les répétitions...)

Théorème 21 *Soit $(f_n)_{n \ge 1}$ une suite de fonctions mesurables.*
a) (LEMME DE FATOU) *Si g est une fonction à valeurs dans $I\!\!R$ et intégrable, on a les implications:*

$$f_n \ge g \quad \forall n \quad \Rightarrow \quad \int (\liminf_n f_n) d\mu \;\le\; \liminf_n \int f_n d\mu, \tag{31}$$

$$f_n \le g \quad \forall n \quad \Rightarrow \quad \int (\limsup_n f_n) d\mu \;\ge\; \limsup_n \int f_n d\mu, \tag{32}$$

b) (THEOREME DE CONVERGENCE DOMINEE DE LEBESGUE) *S'il existe une fonction intégrable g telle que $|f_n| \le g$ pour tout n et si la suite (f_n) converge simplement vers une limite f on a*

$$f \in \mathcal{L}^1(E, \mathcal{E}, \mu) \quad \textit{et} \quad \int f_n d\mu \;\to\; \int f d\mu. \tag{33}$$

Preuve. a) Remarquons d'abord que (31) implique (32): en effet si $f'_n = -f_n$, on a $\limsup_n f_n = -\liminf_n f'_n$; si de plus $f_n \le g$ on a $f'_n \ge -g$, tandis que si g et intégrable il en est de même de $-g$: pour obtenir (32) pour la suite (f_n) il suffit alors d'appliquer (31) à la suite (f'_n).

Pour montrer (31), on pose $f'_n = f_n - g$, qui par hypothèse est positive. On a $f_n = f'_n + g^+ - g^-$ et g^- est intégrable, donc le lemme 20 entraine que $\int f_n d\mu$ est bien définie et vaut $\int f'_n d\mu + \int g d\mu$. De même si $f = \liminf_n f_n$ et $f' = f - g$ on a $f' \ge 0$, donc $\int f d\mu$ est bien définie et vaut $\int f' d\mu + \int g d\mu$. Comme enfin $f' = \liminf_n f'_n$, il suffit d'appliquer (25) pour obtenir (31).

b) On a clairement $|f| \le g$, donc f est intégrable. On a aussi $f = \limsup_n f_n = \liminf_n f_n$ et $-g \le f_n \le g$. Par suite (31) et (32) entrainent

$$\int f d\mu \;\le\; \liminf_n \int f_n d\mu \;\le\; \limsup_n \int f_n d\mu \;\le\; \int f d\mu.$$

La propriété $\int f_n d\mu \to \int f d\mu$ en découle immédiatement. $\square$

Le lecteur sera particulièrement attentif à l'énoncé du théorème de Lebesgue, dans lequel il y a deux hypothèses: 1) la suite (f_n) converge simplement, ce qui signifie $f_n(x) \to f(x)$ pour tout x, et 2) la suite (f_n) est "dominée" par la fonction g, ce qui signifie $|f_n(x)| \leq g(x)$ pour tout x et tout n, et en plus g est intégrable. Sans la première hypothèse l'énoncé n'a pas de sens car la fonction f n'est pas définie. Sans la seconde le théorème est faux, comme le montre l'exemple cité après le théorème 16: on prend $f_n(x) = 1/n$ pour tout $x \in E$, qui converge simplement (et même uniformément !) vers la fonction nulle $f = 0$, alors que si μ est une mesure infinie les intégrales $\int f_n d\mu$ (qui sont infinies) ne convergent pas vers $\int f d\mu = 0$: dans cet exemple la plus petite fonction g dominant la suite (f_n) est $g(x) = 1$, et elle n'est pas intégrable.

Signalons que le théorème de Lebesgue généralise le théorème 1-14-(b): avec les notations de ce dernier théorème, et si $f_n = 1_{A_n}$, on a convergence simple de (f_n) vers $f = 1_A$, et domination par la fonction $g = 1_B$.

3 L'intégrale des fonctions à valeurs complexes

Il est utile (en particulier en analyse de Fourier, comme on le verra plus loin) d'intégrer des fonctions complexes. Nous allons voir que cette opération est très simple, à condition de considérer une fonction complexe comme un couple de deux fonctions réelles.

Comme dans la section précédente, on fixe un ensemble E muni d'une tribu $\mathcal{E}$ et d'une mesure μ. Une fonction complexe sur E est une application de E dans $\mathbb{C}$. Rappelons que tout nombre complexe y peut s'écrire de manière unique comme $y = a + ib$ où a et b sont des réels appelés respectivement partie réelle et partie imaginaire de y. On écrit aussi $a = \mathcal{R}(y)$ et $b = \mathcal{I}(y)$. Inversement si a, b sont des réels on leur associe le complexe $y = a + ib$. On peut ainsi identifier les ensembles $\mathbb{C}$ et $\mathbb{R}^2$, et cette identification est encore valable pour les notions de convergence (et donc pour la topologie): les complexes $y_n = a_n + ib_n$ convergent vers le complexe $y = a + ib$ si et seulement si les deux suites réelles (a_n) et (b_n) convergent respectivement vers a et b. Par suite la tribu borélienne $\mathcal{C}$ de $\mathbb{C}$ peut être identifiée à la tribu borélienne $\mathcal{R}^2$ de $\mathbb{R}^2$.

Toute fonction complexe f sur E s'écrit $f = \mathcal{R}(f) + i\mathcal{I}(f)$ où $\mathcal{R}(f)$ et $\mathcal{I}(f)$ sont les fonctions réelles sur E définies par $\mathcal{R}(f)(x) = \mathcal{R}(f(x))$ et $\mathcal{I}(f)(x) = \mathcal{I}(f(x))$. La fonction f est mesurable de $(E, \mathcal{E})$ dans $(\mathbb{C}, \mathcal{C})$ si et seulement si les deux fonctions $\mathcal{R}(f)$ et $\mathcal{I}(f)$ sont mesurables de $(E, \mathcal{E})$ dans $(\mathbb{R}, \mathcal{R})$.

Rappelons encore que le *module* du complexe $y = a + ib$ est $|y| = \sqrt{a^2 + b^2}$. Si f est une fonction complexe, on a

$$|f| \;\leq\; |\mathcal{R}(f)| + |\mathcal{I}(f)|, \qquad |\mathcal{R}(f)| \;\leq\; |f|, \qquad |\mathcal{I}(f)| \;\leq\; |f|. \tag{34}$$

Si de plus f est mesurable, la fonction $|f|$ est aussi mesurable par les propositions 6 et 8.

Définition 22 La fonction complexe f sur $(E, \mathcal{E})$ est dite *intégrable par rapport à la mesure* μ si d'une part elle est mesurablen et si d'autre part la fonction réelle $|f|$ est intégrable. Cela entraine d'après (34) que les fonctions réelles $\mathcal{R}(f)$ et $\mathcal{I}(f)$ sont intégrables, et l'intégrale de f est le nombre complexe suivant:

$$\int f d\mu \;=\; \int f(x)\mu(dx) \;=\; \int \mathcal{R}(f)d\mu + i\int \mathcal{I}(f)d\mu. \qquad \square \tag{35}$$

Théorème 23 (i) *L'ensemble des fonctions complexes intégrables est un espace vectoriel sur $\mathbb{C}$.*

(ii) *L'application $f \mapsto \int f d\mu$ de cet espace dans $\mathbb{C}$ est une forme linéaire.*

(iii) *On a pour toute fonction complexe intégrable:*

$$|\int f d\mu| \ \leq \ \int |f| d\mu. \tag{36}$$

Preuve. Compte tenu du théorème 19 les deux premières assertions sont évidentes. Soit f une fonction complexe intégrable. Il existe un $z \in \mathbb{C}$ avec $|z| = 1$ et tel que le produit $z \int f d\mu$ soit réel, et bien entendu $|z \int f d\mu| = |\int f d\mu|$. Par ailleurs la linéarité montre que $z \int f d\mu = \int (zf) d\mu$. Comme cette expression est réelle, en comparant à (35) on voit qu'en fait $z \int f d\mu = \int \mathcal{R}(zf) d\mu$. Mais $|\mathcal{R}(zf)| \leq |zf| = |f|$ par (34), donc (30) et le théorème 16-(iii) entraînent que $|z \int f d\mu| \leq \int |f| d\mu$ et on obtient ainsi (36). $\square$

4 L'intégrale par rapport à la mesure de Lebesgue

Dans cette dernière section nous allons considérer le cas particulier où $E = \mathbb{R}$ est muni de sa tribu borélienne et de la mesure de Lebesgue λ. La théorie de l'intégration dans ce cas n'est nullement plus simple que dans le cas général vu plus haut, mais il est évidemment important de vérifier que l'intégrale obtenue dans ce chapitre (qu'on appelle "intégrale de Lebesgue") coïncide avec l'intégrale de Riemann lorsque celle-ci existe.

Pour montrer en toute généralité qu'une fonction Riemann-intégrable est aussi Lebesgue-intégrable il nous manque encore un outil qui sera développé dans le chapitre suivant. Mais nous pouvons dès à présent montrer que pour une fonction f qui est continue par morceaux (cela veut dire qu'il existe un nombre *fini* de réels $a_1 < \ldots < a_k$ tels que la fonction f soit continue en tout point x différent de tous les a_i, et telle qu'en plus elle admette une limite à droite et une limite à gauche finies en chacun des points a_i), les deux intégrales coïncident (dans la pratique, on n'intègre jamais au sens de Riemann des fonctions qui ne sont pas continues par morceaux).

Considérons donc une fonction f sur $\mathbb{R}$, continue par morceaux, qu'on va intégrer sur un intervalle borné $[a, b]$. On note D l'ensemble fini constitué des points a et b et des points de $]a, b[$ où f n'est pas continue, et $C = [a, b] \backslash D$. On va considérer pour chaque n une subdivision $\alpha(n, 0) < \ldots < \alpha(n, k_n)$ de $[a, b]$ en k_n sous-intervalles (donc $\alpha(n, 0) = a$ et $\alpha(n, k_n) = b$), de sorte que tous les points de D soient des points de subdivision, et que le pas de cette subdivision (i.e. $\sup_i(\alpha(n, i) - \alpha(n, i - 1))$) tende vers 0 quand $n \to \infty$. Soit aussi $\beta(n, i)$ un point quelconque de $]\alpha(n, i - 1), \alpha(n, i)[$. Avec ces notations, on sait que l'intégrale de Riemann $\int_a^b f(x) dx$ est la limite des suites

$$I_n \ = \ \sum_{i=1}^{k_n} f(\beta(n, i))(\alpha(n, i) - \alpha(n, i - 1)).$$

Soit alors pour chaque n la fonction

$$f_n(x) \ = \ \begin{cases} f(\beta(n, i)) & \text{si } \ x \in [\alpha(n, i - 1), \alpha(n, i)[\cap C \\ f(x) & \text{si } \ x \in D \\ 0 & \text{si } \ x \notin [a, b]. \end{cases}$$

Une autre manière d'écrire f_n est la suivante:

$$f_n \;=\; \sum_{i=1}^{k_n} f(\beta(n,i)) 1_{[\alpha(n,i-1),\alpha(n,i)[\cap C} + \sum_{u\in D} f(u)1_{\{u\}},$$

et sur cette expression on voit immédiatement que f_n est borélienne et que son intégrale par rapport à la mesure de Lebesgue est

$$\int f_n d\lambda \;=\; \sum_{i=1}^{k_n} f(\beta(n,i))\lambda([\alpha(n,i-1),\alpha(n,i)]\cap C) + \sum_{u\in D} f(u)\lambda(\{u\}).$$

La mesure de Lebesgue d'un singleton est nulle, et $\lambda([\alpha(n,i-1),\alpha(n,i)[\cap C) = \alpha(n,i) - \alpha(n,i-1)$: donc $\int f_n d\lambda = I_n$.

Par ailleurs, étant données les propriétés de f il est très facile de voir que la suite $(f_n)_n$ converge simplement (et même uniformément) vers la fonction $f' = f1_{[a,b]}$, de sorte que f est borélienne. De plus $|f_n| \leq g$ pour tout n, si g désigne la fonction égale à 0 sur le complémentaire de $[a,b]$ et à $\sup_{x\in[a,b]}(|f(x)|)$ sur $[a,b]$. La fonction g étant intégrable, on peut appliquer le théorème de Lebesgue, qui implique que $\int f_n d\lambda = I_n$ converge vers $\int f' d\lambda$. Par suite on a

$$\int_a^b f(x)dx \;=\; \int (f1_{[a,b]}d\lambda. \tag{37}$$

Remarquons au passage que la notation $\int_a^b f(x)dx$ est très commode. On va donc l'utiliser aussi pour l'intégrale de Lebesgue. Plus précisément, si μ est une mesure quelconque sur un espace mesurable $(E,\mathcal{E})$ et si une fonction f admet une intégrale $\int f d\mu$, pour tout $A\in\mathcal{E}$ la fonction $f1_A$ admet également une intégrale (exercice: pourquoi ?), et on utilise les notations $\int_A f d\mu$ ou $\int_A f(x)\mu(dx)$ au lieu de $\int (f1_A)d\mu$. Lorsque de plus μ est la mesure de Lebesgue sur $I\!\!R$ on écrit aussi $\int_A f(x)dx$ au lieu de $\int_A f(x)\lambda(dx)$. Si enfin $A=[a,b]$ on écrira $\int_a^b f(x)dx$, même si f n'est pas intégrable au sens de Riemann.

Noter qu'il existe beaucoup de fonctions qui sont intégrables au sens de Lebesgue, mais pas de Riemann; par exemple l'indicatrice $f = 1_{\boldsymbol{Q}\cap[0,1]}$ de l'ensemble des rationnels de $[0,1]$ est mesurable (et en fait étagée), intégrable et d'intégrale nulle, mais elle n'est pas Riemann-intégrable.

Passons maintenant aux intégrales "sur $I\!\!R$ tout entier": on peut définir sous certaines conditions l'intégrale impropre $\int_{-\infty}^{\infty} f(x)dx$ au sens de Riemann, comme la limite des intégrales de Riemann $\int_a^b f(x)dx$ lorsque $a\to-\infty$ et $b\to+\infty$. La situation est en fait analogue à celle des séries (ce n'est pas un hasard: on a vu que la somme d'une série est en fait l'intégrale d'une fonction sur $I\!\!N$ relativement à la mesure de comptage, qui est l'exact analogue de la mesure de Lebesgue): la fonction f (pour le moment continue par morceaux, mais cela s'appliquera à toutes les fonctions Riemann-intégrables sur chaque intervalle borné $[a,b]$) est intégrable pour la mesure de Lebesgue (i.e. appartient à $\mathcal{L}^1(I\!\!R,\mathcal{R},\lambda)$) si et seulement si l'intégrale $\int_{-\infty}^{+\infty} f(x)dx$ est *absolument convergente* (ce qui signifie que $\int_{-\infty}^{+\infty} |f(x)|dx < \infty$), et dans ce cas les intégrales au sens de Lebesgue et de Riemann coïncident et égalent la limite de $\int_{-n}^{n} f(x)dx$ quand $n\to\infty$.

Remarque sur la terminologie: Soit A un borélien de $I\!\!R$. On munit A de la tribu $\mathcal{R}_A$ des parties de $I\!\!R$ qui sont boréliennes et contenues dans A (cette classe de parties est évidemment une tribu, et c'est aussi l'ensemble des parties de A qui, considérées comme parties de $\mathbb{R}$ sont boréliennes).

Il sera commode dans la suite d'appeler "mesure de Lebesgue sur A " la mesure sur $(A, \mathcal{R}_A)$ définie pour tout $B \in \mathcal{R}_A$ par $\mu(B) = \lambda(B)$ (le lecteur comparera cette mesure avec la restriction $\lambda_{|A}$ de λ à A). La mesure ainsi définie sera notée habituellement λ, comme si on était sur l'espace $\mathbb{R}$ tout entier. Remarquer que $\int_A f(x)dx$ ou $\int_A f(x)\lambda(dx)$ (notations du début de la page) signifie alors aussi l'intégrale de f (considérée comme fonction sur A) par rapport à la mesure de Lebesgue sur A: toutes ces notations et cette terminologie sont donc cohérentes.

Le même abus de terminologie s'applique pour la mesure de Lebesgue sur $\mathbb{R}^d$, ou sur une partie borélienne de $\mathbb{R}^d$.

CHAPITRE 3

Intégration: quelques compléments

Ce chapitre est consacré à divers compléments au chapitre 2. Ces compléments tournent autour des ensembles dits "négligeables" et d'une généralisation assez anodine de l'intégration telle qu'elle est exposée au chapitre précédent, et autour des certaines applications assez faciles mais importantes du théorème de convergence dominée. Dans le paragraphe 1 ci-dessous, outre la notion importante d'ensemble négligeable, on introduit celle de tribu complétée qui est nettement moins importante.

1 Ensembles négligeables et complétion de tribus

1) Les ensembles négligeables: Donnons nous un espace mesurable quelconque $(E,\mathcal{E})$, muni d'une mesure μ. Un élément A de $\mathcal{E}$ est dit μ-négligeable si $\mu(A)=0$. A certains égards il est naturel de dire aussi que tout sous-ensemble B de A est μ-négligeable, qu'il appartienne à $\mathcal{E}$ ou non: par exemple sur $I\!R$ muni de la mesure de Lebesgue, toute partie d'un borélien de "longueur" nulle est naturellement qualifié aussi d'une longueur nulle. Cela conduit à la définition suivante:

Définition 1 Une partie B de E est dite *μ-négligeable* (ou négligeable par rapport à μ, ou simplement négligeable s'il n'y a pas d'ambiguïté quant à la mesure μ) s'il existe un ensemble $A\in\mathcal{E}$ tel que $B\subset A$ et que $\mu(A)=0$.

De plus, une propriété $\mathcal{P}$ relative aux points de E est dite *vraie μ-presque partout* si le complémentaire de l'ensemble des points x où elle est réalisée est μ-négligeable; en abrégé on écrit: $\mathcal{P}$ est vraie μ-p.p. $\square$

Par exemple, si f et g sont deux fonctions sur E, on dit que $f=g$ μ-p.p. si l'ensemble $\{f\neq g\}$ est négligeable, ou que $f<g$ μ-p.p. si l'ensemble $\{f\geq g\}$ est négligeable, etc... Si A et B sont deux parties de E, on écrit aussi par abus de notation $A=B$ μ-p.p. (resp. $A\subset B$ μ-p.p.) lorsque l'ensemble $A\Delta B$ est négligeable (resp. l'ensemble $A\cap B^c$ est négligeable), ce qui revient aussi à dire que $1_A=1_B$ μ-p.p. (resp. $1_A\leq 1_B$ μ-p.p.).

Exemples:

1) Supposons que la tribu $\mathcal{E}$ contienne les singletons $\{x\}$. Si μ est la mesure de Dirac au point $a \in E$, un ensemble A est μ-négligeable si et seulement s'il ne contient pas a (en effet le plus grand ensemble de μ-mesure nulle qui soit contenu dans $\mathcal{E}$ est le complémentaire $\{a\}^c$). Noter que cette propriété est vraie quelle que soit la tribu $\mathcal{E}$ contenant les singletons (ou même, quelle que soit la tribu $\mathcal{E}$ contenant le singleton $\{a\}$).

2) Si la tribu est engendrée par une partition finie ou dénombrable $(A_i)_{i\in I}$, une partie de E est négligeable si et seulement si elle est contenue dans la réunion $\cup_{i\in J} A_i$, où J est l'ensemble des indices i pour lesquels $\mu(A_i) = 0$.

3) Si μ est la mesure nulle, toutes les parties de E sont négligeables; cette mesure est clairement la seule pour laquelle E lui-même est négligeable.

Voici quelques propriétés simples de la classe $\mathcal{N}$ des ensembles négligeables:

Proposition 2 *La classe $\mathcal{N}$ vérifie les propriétés suivantes:*

$$\emptyset \in \mathcal{N}, \tag{1}$$

$$B \subset A, \quad A \in \mathcal{N} \quad \Rightarrow \quad B \in \mathcal{N}, \tag{2}$$

$$A_i \in \mathcal{N} \quad \forall i \in I, \quad I \quad \text{fini ou dénombrable} \qquad \Rightarrow \qquad \cup_{i\in I} A_i \in \mathcal{N}, \tag{3}$$

$$A_i \in \mathcal{N} \quad \forall i \in I, \quad I \quad \text{quelconque} \qquad \Rightarrow \qquad \cap_{i\in I} A_i \in \mathcal{N}. \tag{4}$$

Preuve. (1) est évident puisque $\emptyset \in \mathcal{E}$ et $\mu(\emptyset) = 0$. Si $A \in \mathcal{N}$ il existe $A' \in \mathcal{E}$ tel que $A \subset A'$ et $\mu(A') = 0$ par définition. Si alors $B \subset A$ on a aussi $B \subset A'$, et on en déduit que $B \in \mathcal{N}$: d'où (2).

Pour les deux autres propriétés, remarquons que pour chaque i il existe $B_i \in \mathcal{E}$ avec $\mu(B_i) = 0$ et $A_i \subset B_i$. Par suite $\cap_{i\in I} A_i \subset B_j$ pour n'importe quel $j \in I$, de sorte qu'on a (4). On a aussi $\cup_{i\in I} A_i \subset \cup_{i\in I} B_i$; si I est fini ou dénombrable, $\cup_{i\in I} B_i$ est dans $\mathcal{E}$ et de mesure nulle (cf. (1-17)), de sorte qu'on a (3). $\square$

Il découle immédiatement de (3) ci-dessus que

$$f = f' \;\; \mu-\text{p.p. et } \; g = g' \;\; \mu\text{-p.p.} \quad \Rightarrow \quad f+g = f'+g' \;\; \mu-\text{p.p.}, \;\; af = af' \;\; \mu-\text{p.p.} \tag{5}$$

$$f_n = g_n \;\; \mu-\text{p.p.} \;\; \forall n \in I\!N \qquad \Rightarrow \qquad \begin{cases} \sup_n f_n = \sup_n g_n \;\; \mu-\text{p.p.} \\ \inf_n f_n = \inf_n g_n \;\; \mu-\text{p.p.} \\ \limsup_n f_n = \limsup_n g_n \;\; \mu-\text{p.p.} \\ \liminf_n f_n = \liminf_n g_n \;\; \mu-\text{p.p.} \end{cases} \tag{6}$$

2) La tribu complétée: Par définition, on appelle *tribu complétée* de $\mathcal{E}$ par rapport à μ la tribu engendrée par la réunion $\mathcal{E} \cup \mathcal{N}$.

Voici d'abord une description de cette tribu complétée:

Proposition 3 *La tribu complétée de $\mathcal{E}$ par rapport à μ égale chacune des trois classes suivantes de parties de E:*

a) La classe des parties A de E pour lesquelles il existe deux éléments B et C de $\mathcal{E}$ avec

$$B \subset A \subset C, \qquad \mu(C \backslash B) = 0. \tag{7}$$

b) La classe des parties A de E pour lesquelles il existe $B \in \mathcal{E}$ et $N \in \mathcal{N}$ avec

$$A \ = \ B \cup N. \tag{8}$$

c) La classe des parties A de E pour lesquelles il existe $B \in \mathcal{E}$ avec

$$A = B \ \ \mu - \text{p.p.} \quad (\text{i.e. } A \Delta B \in \mathcal{N}). \tag{9}$$

Preuve. Soit $\mathcal{F}$ la tribu complétée; notons $\mathcal{A}$, $\mathcal{B}$ et $\mathcal{C}$ les classes de parties décrites dans (a), (b) et (c). (7) implique que $N = A \backslash B$ est dans $\mathcal{N}$, donc on a aussi (8): par suite $\mathcal{A} \subset \mathcal{B}$. Si on a (8) il vient $A \Delta B \subset N$, donc on a aussi (9) et $\mathcal{B} \subset \mathcal{C}$. Si on a (9) il existe $D \in \mathcal{E}$ avec $A \Delta B \subset D$ et $\mu(D) = 0$: si alors $B' = B \cap D^c$ et $C' = B \cup D$ il vient $B' \subset A \subset C'$ et $B' \in \mathcal{E}$, $C' \in \mathcal{E}$ et $C' \backslash B' \subset D$, donc $\mu(C' \backslash B') = 0$: on a donc (7), de sorte que $\mathcal{C} \subset \mathcal{A}$. Donc finalement $\mathcal{A} = \mathcal{B} = \mathcal{C}$.

Il est clair que $\mathcal{B} \subset \mathcal{F}$, et que $\mathcal{E} \subset \mathcal{B}$ (prendre $N = \emptyset$ dans (8)) et $\mathcal{N} \subset \mathcal{B}$ (prendre $A = \emptyset$ dans (8)). Il reste donc à prouver que $\mathcal{B} = \mathcal{C}$ est une tribu.

Que $E \in \mathcal{C}$ est évident. Si A vérifie (9) avec $B \in \mathcal{E}$, alors A^c vérifie aussi (9) avec B^c (puisque $A^c \Delta B^c = A \Delta B$), tandis que $B^c \in \mathcal{E}$: donc $A^c \in \mathcal{C}$. Si enfin les A_n vérifient (9) avec les $B_n \in \mathcal{E}$, et si $A = \cup_n A_n$ et $B = \cup_n B_n$ on a $B \in \mathcal{E}$, et $A \Delta B \subset \cup_n (A_n \Delta B_n)$; cette dernière réunion est dans $\mathcal{N}$ en vertu de (3), donc également $A \Delta B$ en vertu de (2): par suite $A \in \mathcal{C}$. Cela achève de prouver que $\mathcal{C}$ est une tribu. $\square$

Proposition 4 *Soit $\mathcal{F}$ la tribu complétée de $\mathcal{E}$. Une fonction f sur E à valeurs dans $\mathbb{R}$ ou dans $\overline{\mathbb{R}}$ est $\mathcal{F}$-mesurable si et seulement si l'une des deux conditions équivalentes suivantes est satisfaite:*

a) Il existe une fonction $\mathcal{E}$-mesurable f' telle que $f = f'$ μ-p.p. (i.e. l'ensemble $\{f \neq f'\}$ est μ-négligeable).

b) Il existe deux fonctions $\mathcal{E}$-mesurables g et h telles que

$$g \ \leq \ f \ \leq \ h, \qquad g \ = \ h \ \ \mu - \text{p.p.} \tag{10}$$

Preuve. On a (b)$\Rightarrow$(a): prendre par exemple $f' = g$ ou $f' = h$.

Supposons (a). Pour tout $x \in \mathbb{R}$, on a $\{f < x\} \Delta \{f' < x\} \subset \{f' \neq f\}$, donc $\{f < x\} \Delta \{f' < x\} \in \mathcal{N}$. Comme $\{f' < x\} \in \mathcal{E}$ en vertu de la $\mathcal{E}$-mesurabilité de f', on obtient $\{f < x\} \in \mathcal{F}$ par la proposition précédente. Ceci étant vrai pour tout $x \in \mathbb{R}$, il suffit d'appliquer la proposition 2-5 pour obtenir que f est $\mathcal{F}$-mesurable.

Il reste à montrer que si f est $\mathcal{F}$-mesurable on a (b). Pour cela on considère la classe $\mathcal{U}$ de toutes les fonctions f à valeurs dans $\overline{\mathbb{R}}_+$ et qui vérifient (b). Cette classe est stable par addition: si $f, f' \in \mathcal{U}$ sont associées respectivement aux couples (g, h) et (g', h') par

(10), on peut évidemment supposer que $g \geq 0$ et $g' \geq 0$; alors $g + g'$ et $h + h'$ sont $\mathcal{E}$-mesurables et $g + g' \leq f + f' \leq h + h'$ et $\{g + g' < h + h'\} \subset \{g < h\} \cup \{g' < h'\}$, donc $\mu(\{g + g' < h + h'\}) = 0$, de sorte qu'on a bien $f + f' \in \mathcal{U}$. La classe $\mathcal{U}$ est également stable par multiplication par une constante positive (même démonstration), et aussi par limite croissante: supposons que les $(f_n)_{n\geq 1}$ soient dans $\mathcal{U}$ et croissent vers f; soit (g_n, h_n) le couple associé à f_n par (10); les fonctions $g = \sup_n g_n$ et $h = \sup_n h_n$ sont $\mathcal{E}$-mesurables (proposition 2-8); on a clairement $g \leq f \leq h$; enfin $\{g < h\} \subset \cup_n \{g_n < h_n\}$, qui est négligeable par (3).

Remarquer que tout $A \in \mathcal{F}$ vérifie (7): on a donc $1_B \leq 1_A \leq 1_C$ et $1_B = 1_C$ μ-p.p., de sorte que $1_A \in \mathcal{U}$. En utilisant les propriétés prouvées ci-dessus on en déduit que $\mathcal{U}$ contient toutes les fonctions de la forme $\sum_{i=1}^n a_i 1_{A_i}$ pour $a_i \geq 0$ et $A_i \in \mathcal{F}$: en d'autres termes, $\mathcal{U}$ contient toutes les fonctions $\mathcal{F}$-mesurables étagées positives. A cause de la stabilité de $\mathcal{U}$ par limite croissante, et en utilisant le lemme 2-13, on voit que $\mathcal{U}$ contient toutes les fonctions $\mathcal{F}$-mesurables à valeurs dans $\overline{\mathbb{R}}_+$ (d'après ce qui est montré au début de la preuve, $\mathcal{U}$ est en fait exactement l'ensemble de ces fonctions).

Il reste à examiner le cas où f est $\mathcal{F}$-mesurable de signe quelconque. D'après ce qui précède il existe deux couples de fonctions $\mathcal{E}$-mesurables (g', h') et (g'', h'') tels que $0 \leq g' \leq f^+ \leq h'$ et $0 \leq g'' \leq f^- \leq h''$ et que $g' = f'$ μ-p.p. et $g'' = h''$ μ-p.p.; noter qu'on peut toujours remplacer h'' par la fonction $\mathcal{E}$-mesurable $h'' 1_{\{g'=0\}}$ (car si $g' > 0$ on a $f^+ > 0$, donc $f^- = 0$), ce qui revient à supposer que $h'' = 0$ sur $\{g' = +\infty\}$, et on peut de même supposer que $h' = 0$ sur $\{g' = +\infty\}$. Les fonctions $g = g' - h''$ et $h = h' - g''$ sont $\mathcal{E}$-mesurables et vérifient $g \leq f \leq h$ et $g = h$ μ-p.p.: donc f vérifie (10), et la preuve est terminée. □

3) Extension de la mesure à la tribu complétée: On va maintenant étendre la mesure μ à la tribu complétée $\mathcal{F}$ de $\mathcal{E}$ par rapport à μ. On va commencer par un lemme qui sera amélioré plus loin.

Lemme 5 *a) Si A et B sont deux parties $\mathcal{E}$-mesurables vérifiant $A = B$ μ-p.p., on a $\mu(A) = \mu(B)$.*

b) Si f et g sont deux fonctions $\mathcal{E}$-mesurables vérifiant $f = g$ μ-p.p., alors f admet une intégrale (resp. est intégrable) si et seulement si g admet une intégrale (resp. est intégrable), et on a alors $\int f d\mu = \int g d\mu$.

Preuve. Comme $A = B$ μ-p.p. équivaut à dire que $1_A = 1_B$ μ-p.p., (a) découle de (b) appliqué à $f = 1_A$ et $g = 1_B$.

Comme $f = g$ μ-pp. implique $f^+ = g^+$ μ-pp. et $f^- = g^-$ μ-pp., il suffit clairement de montrer que si f et g sont positives, on a $\int f d\mu = \int g d\mu$. Mais si h est la fonction qui vaut $+\infty$ aux points où $f \neq g$ et qui vaut 0 là où $f = g$, on a $f \leq g + h$, tandis que le fait que h soit étagée avec deux valeurs 0 et $+\infty$ conduit à $\int h d\mu = +\infty \times \mu(\{f \neq g\}) = 0$. Donc $\int f d\mu \leq \int g d\mu$, et l'inégalité inverse se montre de la même manière. □

Proposition 6 *Pour tout $A \in \mathcal{F}$ la formule*

$$\mu'(A) \;=\; \mu(B) \quad \text{si} \quad A = B \cup N \;\; \text{avec} \;\; B \in \mathcal{E} \;\; \text{et} \;\; N \in \mathcal{N}. \tag{11}$$

définit un nombre $\mu'(A)$ qui ne dépend pas de la décomposition $A = B \cup N$ choisie dans (11). L'application $A \mapsto \mu'(A)$ de $\mathcal{F}$ dans $\overline{\mathbb{R}}_+$ définit une mesure μ' sur $(E, \mathcal{F})$ qui est une extension de μ au sens où $\mu'(A) = \mu(A)$ si $A \in \mathcal{E}$. Cette extension est l'unique extension possible de μ à $\mathcal{F}$, et on l'appelle la mesure complétée.

Preuve. Soit $A = B \cup N = B' \cup N'$ deux décompositions de $A \in \mathcal{F}$ avec $B, B' \in \mathcal{E}$ et $N, N' \in \mathcal{N}$. Comme $B \Delta B' \subset N \cup N'$ et comme $N \cup N'$ est négligeable, donc contenu dans un $C \in \mathcal{E}$ avec $\mu(C) = 0$, on a $\mu(B \Delta B') = 0$, ce qui implique $\mu(B) = \mu(B')$: ainsi la formule (11) ne dépend pas de la décomposition choisie pour A.

Il est clair que $\mu'(A) = \mu(A)$ si $A \in \mathcal{E}$, et en particulier $\mu'(\emptyset) = 0$. Pour montrer que μ' est une mesure il reste donc à prouver la σ-additivité. Soit une suite $(A_n)_{n \geq 1}$ d'éléments de $\mathcal{F}$ deux-à-deux disjoints, de décompositions $A_n = B_n \cup N_n$ avec $B_n \in \mathcal{E}$ et $N_n \in \mathcal{N}$. On a $\cup_n A_n = (\cup_n B_n) \cup (\cup_n N_n)$, et $\cup_n B_n \in \mathcal{E}$, et $\cup_n N_n \in \mathcal{N}$, et enfin les B_n sont aussi deux-à-deux disjoints: on a donc

$$\mu'(\cup_n A_n) \ = \ \mu(\cup_n B_n) \ = \ \sum_n \mu(B_n) \ = \ \sum_n \mu'(A_n).$$

Soit enfin μ'' une autre mesure sur $\mathcal{F}$ qui étend μ. Si $A = B \cup N$ est dans $\mathcal{F}$, avec $B \in \mathcal{E}$ et $N \in \mathcal{N}$, il existe $C \in \mathcal{E}$ avec $N \subset C$ et $\mu(C) = 0$. Comme $B \subset A \subset B \cup C$ il vient

$$\mu(B) \ = \ \mu''(B) \ \leq \ \mu''(A) \ \leq \ \mu''(B \cup C) \ = \ \mu(B \cup C) \ \leq \ \mu(B) + \mu(C) \ = \ \mu(B),$$

de sorte que $\mu''(A) = \mu(B)$, qui égale $\mu'(A)$ par (11), donc $\mu'' = \mu'$. $\square$

Voici maintenant un résultat qui contient l'amélioration promise du lemme 5:

Proposition 7 *a) La classe des ensembles négligeables pour μ' est la même que la classe $\mathcal{N}$ des ensembles négligeables pour μ.*

b) Si f est une fonction $\mathcal{F}$-mesurable, pour toute fonction $\mathcal{E}$-mesurable g égale μ-p.p. à f (il en existe d'après la proposition 4)*, on a que f admet une intégrale (resp. est intégrable) par rapport à μ' si et seulement si g admet une intégrale (resp. est intégrable) par rapport à μ, et dans ce cas $\int f d\mu' = \int g d\mu$.*

Preuve. a) Il est clair que la classe $\mathcal{N}$ est contenue dans la classe $\mathcal{N}'$ des ensembles μ'-négligeables. Inversement si $A \in \mathcal{N}'$ il existe $B \in \mathcal{F}$ avec $A \subset B$ et $\mu'(B) = 0$; mais (11) implique alors que $B = C \cup N$ avec $N \in \mathcal{N}$ et $C \in \mathcal{E}$ et $\mu(C) = 0$: on a donc aussi $C \in \mathcal{N}$, donc $B \in \mathcal{N}$; donc $A \in \mathcal{N}$ (appliquer la proposition 2): il s'ensuit que $\mathcal{N} = \mathcal{N}'$.

b) Comme μ' est une extension de μ, on a clairement qu'une fonction $\mathcal{E}$-mesurable g admet une intégrale (resp. est intégrable) par rapport à μ et et seulement si c'est la cas aussi par rapport à μ', et on a alors $\int g d\mu' = \int g d\mu$. Par ailleurs, (a) implique qu'une propriété est vraie μ-p.p. si et seulement si elle est vraie μ'-p.p.: la partie (b) découle alors du lemme 5 appliqué à la mesure μ' et à la tribu $\mathcal{F}$. $\square$

Cette proposition montre qu'il ne sert à rien de "compléter" la tribu $\mathcal{F}$ par rapport à la mesure μ': en effet les ensembles μ'-négligeables sont contenus dans $\mathcal{F}$, de sorte que $\mathcal{F}$ est sa propre complétée.

Notation: Comme μ' est l'unique extension de μ à la tribu $\mathcal{F}$, et comme les intégrales des fonctions $\mathcal{E}$-mesurables sont les mêmes par rapport à μ ou à μ', il est habituel de noter encore μ la mesure précédemment appelée μ'. $\Box$

Exemples:

1) Supposons que $\mu = \varepsilon_a$ soit la masse de Dirac en a, et que la tribu $\mathcal{E}$ contienne le singleton $\{a\}$. On a vu qu'une partie de E est négligeable si et seulement si elle ne contient pas le point a. La tribu complétée $\mathcal{F}$ est alors la tribu $\mathcal{F} = \mathcal{P}(E)$ de toutes les parties de E, et la mesure complétée μ' est la masse de Dirac en a (mais, maintenant, sur l'espace mesurable $(E, \mathcal{P}(E))$).

2) Supposons que $(E, \mathcal{E}) = (I\!R, \mathcal{R})$ soit muni de la mesure de Lebesgue λ. La tribu complétée $\mathcal{F}$ de $\mathcal{R}$ s'appelle la *tribu de Lebesgue*. Elle est strictement plus grande que la tribu borélienne, mais elle est strictement plus petite que la tribu de toutes les parties $\mathcal{P}(I\!R)$.

4) Nous allons terminer ce paragraphe avec quelques résultats en rapport plus ou moins proche avec les ensembles négligeables. Commençons par un lemme qui, connu sous le nom *d'inégalité de Bienaymé-Tchebicheff*, est utile dans de nombreuses applications. Dans ce qui suit on considère l'espace mesuré $(E, \mathcal{E}, \mu)$, mais on pourrait tout aussi bien se placer sur l'espace "complété" $(E, \mathcal{F}, \mu')$.

Lemme 8 *Si f est une fonction mesurable à valeurs dans $\overline{I\!R}$, on a pour tout $a \in]0, \infty[$:*

$$\mu(\{|f| \geq a\}) \;\leq\; \frac{1}{a} \int |f| d\mu. \tag{12}$$

Preuve. La fonction $g = a1_{\{|f| \geq a\}}$ vérifie $g \leq |f|$, donc $\int g d\mu \leq \int |f| d\mu$. Comme $\int g d\mu = a\mu(\{|f| \geq a\})$, on en déduit immédiatement (12). $\Box$

Corollaire 9 *Si f est une fonction mesurable à valeurs dans $\overline{I\!R}$, intégrable, alors l'ensemble $\{|f| = +\infty\}$ est négligeable* (i.e. on a $|f| < +\infty$ μ-p.p.).

Preuve. On a $\mu(\{|f| = +\infty\}) \leq \mu(\{|f| \geq n\}) \leq \frac{1}{n} \int |f| d\mu$ par (12). Comme $\int |f| d\mu < +\infty$, il suffit de faire tendre n vers l'infini pour obtenir le résultat. $\Box$

Pour bien comprendre ce résultat, il faut noter que si la fonction f est intégrable, elle n'est pas nécessairement à valeurs finies: modifier f (par exemple remplacer les valeurs de f par $+\infty$) sur un ensemble négligeable n'altère pas son intégrabilité.

Corollaire 10 *a) Si $(f_n)_{n \geq 1}$ est une suite de fonctions mesurables à valeurs dans $\overline{I\!R}_+$ et si $\sum_n \int f_n d\mu < \infty$, on a $\sum_n f_n < \infty$ μ-p.p.*

b) (Lemme de BOREL-CANTELLI) *Si $(A_n)_{n \geq 1}$ est une suite de parties mesurables de $(E, \mathcal{E})$ vérifiant $\sum_n \mu(A_n) < \infty$, alors $\mu(\limsup_n A_n) = 0$.*

Preuve. a) D'après le corollaire 2-17, la fonction $g = \sum_n |f_n|$ est intégrable, et il suffit donc d'appliquer le corollaire 9.

b) L'assertion découle de (a) appliqué à la suite $f_n = 1_{A_n}$: d'une part on a $\int f_n d\mu = \mu(A_n)$; d'autre part $\limsup_n A_n = \{\sum_n f_n = +\infty\}$. □

Proposition 11 *Si f est une fonction mesurable à valeurs dans $\overline{\mathbb{R}}$, on a l'équivalence:*

$$f = 0 \quad \mu-\text{p.p.} \quad \Leftrightarrow \quad \int |f| d\mu = 0. \tag{13}$$

Preuve. Si $f = 0$ μ-p.p., on a aussi $|f| = 0$ μ-p.p., donc $\int |f| d\mu = 0$ par le lemme 5. Si inversement $\int |f| d\mu = 0$, le lemme 8 implique $\mu(\{|f| \geq \frac{1}{n}\}) = 0$ pour tout n, et comme $\{|f| \geq \frac{1}{n}\}$ croît vers $\{f \neq 0\}$ on en déduit que $\mu(\{f \neq 0\}) = 0$, donc $f = 0$ μ-p.p. □

2 Théorème de convergence dominée: la version définitive

Nous allons donner maintenant les versions "définitives" du théorème de convergence dominée de Lebesgue et du lemme de Fatou. On se place toujours sur un espace mesuré $(E, \mathcal{E}, \mu)$.

Théorème 12 *Soit $(f_n)_{n \geq 1}$ une suite de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$.*

a) Si g est une fonction intégrable, on a les implications:

$$f_n \geq g \quad \mu-\text{p.p.} \quad \forall n \quad \Rightarrow \quad \int (\liminf_n f_n) d\mu \ \leq \ \liminf_n \int f_n d\mu. \tag{14}$$

$$f_n \leq g \quad \mu-\text{p.p.} \quad \forall n \quad \Rightarrow \quad \int (\limsup_n f_n) d\mu \ \geq \ \limsup_n \int f_n d\mu. \tag{15}$$

c) S'il existe une fonction g intégrable telle que $|f_n| \leq g$ μ-p.p. pour tout n, et si la suite (f_n) converge μ-p.p. vers une limite f (ce qui veut dire que f est une fonction telle que l'ensemble des x vérifiant $f_n(x) \to f(x)$ est de complémentaire négligeable), alors

$$\int f_n d\mu \ \to \ \int f d\mu. \tag{16}$$

Il faut remarquer, dans la situation de (c), que $\int f d\mu$ a bien un sens. En effet, si on pose par exemple $h = \limsup_n f_n$, la fonction h est mesurable, et on a $f = h$ μ-p.p.; donc d'après la proposition 4 la fonction f est mesurable par rapport à la tribu complétée de $\mathcal{E}$, et donc $\int f d\mu = \int h d\mu$ par la proposition 7 avec l'abus de notation qui consiste à noter encore μ l'extension de μ à la tribu complétée.

Preuve. Pour (a), considérons $N = \cup_n \{f_n < g\}$, et soit f'_n la fonction définie par $f'_n(x) = g(x)$ si $x \in N$ et $f'_n(x) = f_n(x)$ sinon. On a $f'_n \geq g$, donc 2-(31) implique $\int \liminf_n f'_n d\mu \leq \liminf_n \int f'_n d\mu$. En dehors de l'ensemble négligeable N on a $f'_n = f_n$ et $\liminf_n f_n = \liminf_n f'_n$, de sorte que $\int f_n d\mu = \int f'_n d\mu$ et $\int \liminf_n f_n d\mu = \int \liminf_n f'_n d\mu$ par la proposition 7, d'où (14).

(15) se montre de la même manière. Pour (b) la preuve est du même type: soit $h = \limsup_n f_n$ et $h' = \liminf_n f_n$, puis $N = (\cup_n\{|f_n| > g\}) \cup \{h' < h\}$, puis les fonctions mesurables f'_n et g' définies par $f'_n(x) = g'(x) = 0$ si $x \in N$ et $f'_n(x) = f_n(x)$ et $g'(x) = g(x)$ sinon. On a $f'_n = f_n$ et $f = h$ et $g' = g$ en dehors de l'ensemble négligeable N, donc g' est intégrable et $\int f_n d\mu = \int f'_n d\mu$ et $\int f d\mu = \int h d\mu$. Enfin $|f'_n| \leq g'$ et $f'_n \to h$, donc (16) découle de 2-(33) appliqué à la suite f'_n. □

Exemples:

1) On a $\int_0^1 nxe^{-nx}dx \to 0$ quand $n \to \infty$: cela se vérifie en calculant explicitement cette intégrale, mais on peut aussi appliquer le théorème de Lebesgue à la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{R})$ et aux fonctions $f_n(x) = nxe^{-nx}1_{[0,1]}(x)$, qui convergent vers 0 et vérifient $0 \leq f_n \leq 1_{[0,1]}$, alors que la fonction $1_{[0,1]}$ est intégrable par rapport à la mesure de Lebesgue.

2) On a $\int_0^1 nx^2e^{-nx^2}dx \to 0$: un calcul direct n'est pas possible, mais on peut appliquer le théorème de Lebesgue à la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{R})$ et aux fonctions $f_n(x) = nx^2e^{-nx^2}1_{[0,1]}$, qui convergent vers 0 et vérifient $0 \leq f_n \leq 1_{[0,1]}$.

Corollaire 13 *Soit $(u_{n,i})n \geq 1, i \geq 1$ une double suite de réels. Si d'une part $u_{n,i} \to v_i$ pour tout i lorsque $n \to \infty$, si d'autre part $|u_{n,i}| \leq w_i$ pour tout n, avec $\sum_i w_i < \infty$, alors pour chaque n la série $\sum_i u_{n,i}$ est absolument convergente, et $\lim_n \sum_i u_{n,i} = \sum_i v_i$.*

Preuve. La première assertion est évidente, et pour la seconde il suffit d'appliquer le théorème de Lebesgue à la mesure de comptage μ sur $\mathbb{N}^*$ muni de la tribu de toutes les parties et aux fonctions $f_n(i) = u_{n,i}$: ces fonctions convergent simplement vers $f(i) = v_i$ et vérifient $|f_n| \leq g$ pour la fonction positive $g(i) = w_i$, qui est intégrable par rapport à μ puisque $\int g d\mu = \sum_i w_i < \infty$. □

Ce corollaire est appelé théorème d'inversion de la somme et de la limite pour les séries. Par ailleurs le théorème de Lebesgue permet de justifier dans certains cas le procédé de "dérivation sous le signe somme" pour les intégrales de fonctions dépendant d'un paramètre.

Proposition 14 (Continuité et dérivation sous le signe somme) *Soit une fonction f de $I \times E$ dans $\mathbb{R}$, où I est un intervalle de $\mathbb{R}$. On suppose que pour chaque $t \in I$ la fonction $x \mapsto f(t,x)$ est $\mathcal{E}$-mesurable.*

a) Si d'une part pour tout $t \in I$ on a $|f(t,x)| \leq g(x)$ pour tout x en dehors d'un ensemble négligeable et pour une fonction intégrable g, et si d'autre part la fonction $t \mapsto f(t,x)$ est continue en $t = t_0$ pour tout x en dehors d'un ensemble négligeable, alors la fonction $h(t) = \int f(t,x)\mu(dx)$ est continue au point $t = t_0$.

b) Supposons de plus qu'en dehors d'un ensemble négligeable la fonction $t \mapsto f(t,x)$ soit dérivable sur I et que $|\frac{\partial}{\partial t}f(t,x)| \leq g'(x)$ pour une fonction intégrable g', alors la fonction h définie ci-dessus est dérivable sur I, et sa dérivée est $\int \frac{\partial}{\partial t}f(t,x)\mu(dx)$.

Preuve. Noter d'abord que l'hypothèse $|f(t,.)| \leq g$ μ-p.p. entraine que pour chaque t la fonction $f(t,.)$ est intégrable, donc h est bien définie. Pour (a) il suffit de montrer que si une suite (s_n) de points de I tend vers t_0, alors $h(s_n) \to h(t_0)$: cela provient du théorème de Lebesgue appliqué à la suite $f_n(x) = f(s_n, x)$.

Pour (b) il suffit de montrer que si une suite (s_n) de points de I tend vers t, avec $s_n \neq t$ pour tout n, alors $\frac{h(s_n)-h(t)}{s_n-t}$ converge vers $\int \frac{\partial}{\partial t} f(t,x)\mu(dx)$ (cette dernière intégrale étant bien définie, au vu de la condition de majoration de la dérivée). Pour cela on applique le théorème de Lebesgue à la suite $f_n(x) = \frac{f(s_n,x)-f(t,x)}{s_n-t}$, qui converge vers $\frac{\partial}{\partial t} h(t,x)$, en remarquant que d'après le théorème des accroissements finis on a $|f_n| \leq g'$. $\square$

Exemples: 1) Soit g borélienne bornée sur $I\!R_+$. La fonction $h(t) = \int_0^\infty e^{-tx} g(x)dx$ est bien définie, et indéfiniment dérivable sur $]0,\infty[$: cela se voit par application répétée de la proposition précédente, avec $I =]a,\infty[$ pour $a > 0$ arbitraire (si on montre que h est indéfiniment dérivable sur tout intervalle I de la forme ci-dessus, on aura bien-sûr la même propriété sur $]0,\infty[$).

De manière plus précise soit $f(t,x) = e^{-tx} g(x) 1_{[0,\infty[}(x)$, qui est indéfiniment dérivable en t avec $\frac{\partial^n}{\partial t^n} f(t,x) = (-x)^n e^{-tx} g(x) 1_{[0,\infty[}(x)$; pour tout $n \in I\!N$ on a donc $|\frac{\partial^n}{\partial t^n} f(t,x)| \leq g_n(x)$ pour $t \in I$, avec la fonction $g_n(x) = \alpha_n e^{-ax} 1_{[0,\infty[}$ pour une constante convenable α_n (c'est pour cela qu'on se limite aux intervalles I, et qu'on ne peut pas faire directement la preuve sur $]0,\infty[$ entier); chaque fonction g_n est intégrable par rapport à la mesure de Lebesgue sur $I\!R$. On montre alors par récurrence sur n, à l'aide de la proposition 14, que h est n fois dérivable et que sa dérivée d'ordre n est $\int_0^\infty (-x)^n e^{-tx} g(x)dx$.

2) Soit $(u_n)_{n\geq 1}$ des fonctions dérivables sur l'intervalle I de $I\!R$, avec des dérivées vérifiant $|u_n'(x)| \leq v_n$ où v_n est le terme général d'une série convergente. Supposons aussi la série de terme général $u_n(y)$ absolument convergente, pour un point y de I. La somme $S(x) = \sum_n u_n(x)$ est alors bien définie pour tout x, et la fonction S est dérivable, de dérivée $S'(x) = \sum_n u_n'(x)$.

Pour vérifier ceci, on applique la proposition 14 à la mesure de comptage μ sur $E = I\!N^*$ et aux fonction $f(t,n) = u_n(t)$. $\square$

3 Les mesures avec densité

Lorsqu'on dispose d'une mesure μ sur un espace $(E, \mathcal{E})$, la proposition suivante fournit une méthode permettant de lui associer toute une famille d'autres mesures:

Proposition 15 *Si g est une fonction positive mesurable, la formule*

$$\nu(A) = \int_A g d\mu \quad \text{(ce qui veut dire } \nu(A) = \textstyle\int (g 1_A) d\mu) \quad \forall A \in \mathcal{E} \tag{17}$$

définit une nouvelle mesure ν sur $(E, \mathcal{E})$: la fonction g s'appelle la densité de ν par rapport à μ, et la mesure ν est aussi notée $\nu = g \bullet \mu$.

De plus une fonction mesurable f admet une intégrale (resp. est intégrable) par rapport à ν si et seulement si le produit fg admet une intégrale (resp. est intégrable) par rapport à μ, et on a alors

$$\int f d\nu = \int (fg) d\mu. \tag{18}$$

Preuve. On a clairement $\nu(\emptyset) = 0$, et la σ-additivité de ν découle du fait que si les A_n sont deux-à-deux disjoints on a $1_{\cup_n A_n} = \sum_n 1_{A_n}$ et du corollaire 2-17.

Quant à la seconde partie de la proposition, elle découle immédiatement de la formule (18) lorsque f est positive. Il reste donc à montrer que la classe $\mathcal{A}$ des fonctions mesurables positives f vérifiant (18) contient toutes les fonctions mesurables positives.

D'abord, lorsque $f = 1_A$, (18) n'est autre que (17): ainsi, $\mathcal{A}$ contient les indicatrices d'ensembles mesurables. Par "linéarité" (cf. (i,ii) du théorème 2-16) on en déduit que $\mathcal{A}$ contient les fonctions de la forme $\sum_{i=1}^n a_i 1_{A_i}$ pour $n \in I\!N^*$, $a_i \geq 0$ et $A_i \in \mathcal{E}$, c'est-à-dire contient les fonctions mesurables étagées positives. Enfin d'apès (iv) du théorème 2-16 $\mathcal{A}$ contient les limites croissantes de fonctions étagées mesurables positives, c'est-à-dire toutes les fonctions mesurables positives. $\square$

En particulier si $(E, \mathcal{E}) = (I\!R^d, \mathcal{R}^d)$ et si $\mu = \lambda_d$ est la mesure de Lebesgue, la mesure ν construite ci-dessus est appelée *la mesure sur* $I\!R^d$ *de densité* g.

Exemples:

1) Si I est un intervalle de $I\!R$, la restriction à I de la mesure de densité 1_I est ce qu'on a appelé la mesure de Lebesgue sur I à la fin du chapitre 2.

2) La mesure sur $I\!R$ de densité $g(x) = \theta e^{-\theta x} 1_{[0,\infty[}(x)$ s'appelle la loi de probabilité exponentielle de paramètre θ: c'est une mesure de probabilité, c'est-à-dire une mesure de masse totale égale à 1 puisque $\int_0^\infty \theta e^{-\theta x} dx = 1$. Plus généralement, toute mesure sur $I\!R$ de densité g vérifiant $\int_{-\infty}^{+\infty} g(x)dx = 1$ est une mesure de probabilité.

3) Revenons au cas d'un espace mesuré quelconque $(E, \mathcal{E}, \mu)$, et soit g et h deux fonctions mesurables positives sur E. On vérifie immédiatement que $h \bullet (g \bullet \mu) = (gh) \bullet \mu$.

4 Les fonctions intégrables au sens de Riemann

On va terminer ce chapitre en montrant que les fonctions intégrables au sens de Riemann, sur un intervalle borné $I = [a, b]$ de $I\!R$, sont également intégrables au sens de Lebesgue. Ces fonctions ne sont pas nécessairement boréliennes, et il faut donc prendre quelques précautions. De manière précise, on a le résultat suivant:

Théorème 16 *Soit f une fonction bornée sur l'intervalle $I = [a, b]$, intégrable au sens de Riemann. Elle est alors mesurable par rapport à la tribu de Lebesgue (i.e., la tribu complétée de la tribu borélienne par rapport à la mesure de Lebesgue), et son intégrale de Riemann est égale à l'intégrale de Lebesgue de $f 1_I$ par rapport à la mesure (complétée de la mesure) de Lebesgue.*

Preuve. Pour chaque n on considère la subdivision $a = t(n,0) < t(n,1) < \ldots < t(n,2^n) = b$ de $[a,b]$ définie par $t(n,i) = a + (b-a)i2^{-n}$ pour $i = 0, 1, \ldots, 2^n$. On pose

$$\begin{aligned} u(n,i) &= \inf(f(x) : t(n,i-1) \leq x \leq t(n,i)), \\ v(n,i) &= \sup(f(x) : t(n,i-1) \leq x \leq t(n,i)), \end{aligned}$$

$$I_-(n) \;=\; \frac{b-a}{2^n}\sum_{i=1}^{2^n} u(n,i), \qquad I_+(n) \;=\; \frac{b-a}{2^n}\sum_{i=1}^{2^n} v(n,i).$$

Comme f est Riemann-intégrable, on sait que les deux suites $(I_-(n)_{n\geq 1}$ et $(I_+(n))_{n\geq 1}$ convergent vers l'intégrale de Riemann $\int_a^b f(x)dx$.

Par ailleurs, considérons les fonctions boréliennes suivantes:

$$g_n(x) \;=\; \begin{cases} u(n,1) & \text{si } \; t(n,0) \leq x \leq t(n,1) \\ u(n,i) & \text{si } \; t(n,i-1) < x \leq t(n,i) \;\text{ et }\; i=2,3,\ldots,2^n \\ 0 & \text{si } \; x < a \;\text{ ou }\; x > b \end{cases},$$

$$h_n(x) \;=\; \begin{cases} v(n,1) & \text{si } \; t(n,0) \leq xt(n,1) \\ v(n,i) & \text{si } \; t(n,i-1) < x \leq t(n,i) \;\text{ et }\; i=2,3,\ldots,2^n \\ 0 & \text{si } \; x < a \;\text{ ou }\; x > b \end{cases}.$$

On a bien-sûr $g_n \leq f \leq h_n$. Par ailleurs la suite (g_n) est croissante et la suite (h_n) est décroissante: on note g et h leurs limites respectives, qui sont boréliennes et vérifient $g \leq f \leq h$.

Si M désigne la borne supérieure de $|f|$ et si $k(x) = M1_{[a,b]}(x)$, on a $|g_n| \leq k$ et $|h_n| \leq k$, et k est intégrable par rapport à la mesure de Lebesgue. Donc le théorème de Lebesgue implique que $I_-(n)$ et $I_+(n)$ convergent respectivement vers $\int g d\lambda$ et $\int h d\lambda$, qui sont donc toutes deux égales à l'intégrale de Riemann $\int_a^b f(x)dx$ (on ne peut pas appliquer directement le théorème de convergence monotone ici, car les fonctions g_n (resp. h_n) ne sont pas nécessairement positives (resp. négatives)). Donc la fonction positive $h-g$ est d'intégrale nulle, et (13) implique que $g = h$ λ-p.p. Il suffit alors d'utiliser les propositions 4 et 7 pour obtenir le résultat. $\square$

CHAPITRE 4

Produits de mesures

Le cœur de ce chapitre est consacré à la définition du produit de deux (ou de plusieurs) mesures, ce qui va permettre la définition des intégrales "doubles" ou "multiples". Auparavant il nous faut revenir sur les fondements de la théorie de la mesure: plus précisément, nous développons des critères d'unicité très utiles et dont le prototype est le suivant: si μ est une mesure sur $(I\!R, \mathcal{R})$ telle que $\mu(]a,b]) = b - a$ pour tout intervalle borné $]a,b]$, alors μ est la mesure de Lebesgue. La construction proprement dite des mesures est laissée de coté, et le lecteur intéressé pourra consulter l'un des nombreux livres de théorie de l'intégration pour ce sujet.

1 Quelques résultats d'unicité

1) Ci-dessous, $(E, \mathcal{E})$ désigne un espace mesurable quelconque. Le résultat essentiel de ce paragraphe est le suivant:

Théorème 1 *Soit μ et ν deux mesures sur $(E, \mathcal{E})$, et $\mathcal{C}$ une classe de parties de E vérifiant les propriétés suivantes:*

(i) la tribu engendrée par $\mathcal{C}$ est $\mathcal{E}$;

(ii) $\mu(A) = \nu(A) < \infty$ pour tout $A \in \mathcal{C}$;

(iii) la classe $\mathcal{C}$ est stable par intersection finie (i.e. $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$);

(iv) il existe une suite croissante $(E_n)_{n \geq 1}$ d'éléments de $\mathcal{C}$ telle que $E = \lim_n E_n$.

Les mesures μ et ν sont alors égales.

Noter que (ii) et (iv) impliquent que les mesures μ et ν sont σ-finies. En vue de prouver ce théorème nous énonçons d'abord un lemme qui sera utilisé plusieurs fois dans la suite et qui concerne la notion suivante: Une classe $\mathcal{D}$ de parties de E est appelée un λ-*système* si elle vérifie les deux propriétés suivantes:

$$A, B \in \mathcal{D}, \quad A \subset B \quad \Rightarrow \quad B \backslash A \in \mathcal{D}, \tag{1}$$

$$(A_p)_{p \geq 1} \text{ est une suite croissante d'éléments de } \mathcal{D} \quad \Rightarrow \quad \cup_p A_p \in \mathcal{D}. \tag{2}$$

L'intersection d'un nombre quelconque de λ-systèmes est un λ-système (vérification immédiate), et le λ-système *engendré* par une classe $\mathcal{A}$ de parties de E est par définition le plus petit λ-système contenant $\mathcal{A}$ (= l'intersection de tous les λ-systèmes contenant $\mathcal{A}$). Le lemme suivant est souvent appelé *Théorème des classes monotones*, ou plutôt il s'agit d'une des versions de ce théorème.

Lemme 2 *Si $\mathcal{C}$ est une classe de parties de E stable par intersection finie et contenant E lui-même, le λ-système engendré par $\mathcal{C}$ est aussi la tribu engendrée par $\mathcal{C}$.*

Preuve. Soit $\mathcal{E}$ (resp. $\mathcal{F}$) la tribu (resp. le λ-système) engendrée par $\mathcal{C}$. Comme toute tribu est un λ-système, on a $\mathcal{F} \subset \mathcal{E}$, et pour montrer l'inclusion inverse il suffit de prouver que $\mathcal{F}$ est une tribu.

Pour tout $C \in \mathcal{C}$ on note $\mathcal{G}_C$ la classe des $A \in \mathcal{F}$ tels que $A \cap C \in \mathcal{F}$. Comme $(B\backslash A) \cap C = (B \cap C)\backslash(A \cap C)$ et $(\cup_p A_p) \cap C = \cup_p (A_p \cap C)$, il est clair que $\mathcal{G}_C$ est un λ-système. $\mathcal{C}$ étant stable par intersection, on a $\mathcal{C} \subset \mathcal{G}_C$, donc $\mathcal{G}_C = \mathcal{F}$ par définition même de $\mathcal{F}$.

Pour tout $F \in \mathcal{F}$ on note $\mathcal{H}_F$ la classe des $A \in \mathcal{F}$ tels que $A \cap F \in \mathcal{F}$. Exactement comme ci-dessus on voit que $\mathcal{H}_F$ est un λ-système. De plus $\mathcal{C} \subset \mathcal{H}_F$ (en effet si $C \in \mathcal{C}$, et comme $F \in \mathcal{F} = \mathcal{G}_C$, on a $F \cap C \in \mathcal{F}$), de sorte que $\mathcal{H}_F = \mathcal{F}$ par définition de $\mathcal{F}$.

Ce qui précède implique que pour tous $A, B \in \mathcal{F}$ on a $A \cap B \in \mathcal{F}$. Par ailleurs on a $E \in \mathcal{C} \subset \mathcal{F}$, donc (1) implique que si $A \in \mathcal{F}$ on a aussi $A^c \in \mathcal{F}$: ainsi, $\mathcal{F}$ est une algèbre. Pour montrer que c'est une tribu, il reste donc à montrer que $\mathcal{F}$ est stable par réunion dénombrable. Mais si les B_p sont dans $\mathcal{F}$ on a vu (puisque $\mathcal{F}$ est une algèbre) que $A_p = B_1 \cup \ldots B_p$ est dans $\mathcal{F}$, de sorte que (2) entraine $\cup_{p\geq 1} B_p \in \mathcal{F}$, et cela achève la preuve que $\mathcal{F}$ est une tribu. $\square$

Preuve du théorème 1. Notons μ_n et ν_n les restrictions de μ et ν à E_n: rappelons par exemple que $\mu_n(A) = \mu(A \cap E_n)$. Vu le théorème 1-14, on a $\mu(A) = \lim_n \mu_n(A)$ et $\nu(A) = \lim_n \nu_n(A)$ pour tout $A \in \mathcal{E}$: il suffit donc de montrer que $\mu_n = \nu_n$ pour tout n.

Dans la suite, on fixe n. Pour tout $A \in \mathcal{C}$ on a $A \cap E_n \in \mathcal{C}$ par (iii), donc $\mu_n(A) = \nu_n(A) < \infty$. On a aussi $\mu_n(E) = \nu_n(E) < \infty$, puisque $E \cap E_n = E_n \in \mathcal{C}$: en d'autres termes, $\mu_n(A) = \nu_n(A) < \infty$ pour tout A dans la classe $\mathcal{C}' = \mathcal{C} \cup \{E\}$. Par ailleurs la classe $\mathcal{C}'$ engendre la tribu $\mathcal{E}$ et est stable par intersection.

Soit $\mathcal{D}$ la classe des $A \in \mathcal{E}$ tels que $\mu_n(A) = \nu_n(A)$ (rappelons que n est fixé). Cette classe vérifie (1) car on peut écrire $\mu_n(B) = \mu_n(A) + \mu_n(B\backslash A)$ par additivité, donc $\mu_n(B\backslash A) = \mu_n(B) - \mu_n(A)$ puisque la mesure μ_n est finie, et on a des relations analogues pour ν_n; elle vérifie (2) car on a $\mu_n(\cup_p A_p) = \lim_p \mu_n(A_p)$ et une relation analogue pour ν_n. Par suite $\mathcal{D}$ est un λ-système, qui contient $\mathcal{C}'$. En vertu du lemme 2, et comme $\mathcal{D} \subset \mathcal{E}$ par construction, on a en fait $\mathcal{D} = \mathcal{E}$, ce qui veut dire que $\mu_n(A) = \nu_n(A)$ pour tout $A \in \mathcal{E}$, et par suite $\mu_n = \nu_n$. $\square$

Comme première application de ce résultat on obtient l'unicité de la mesure de Lebesgue dans les théorèmes 1-19 et 1-20: en effet toutes les mesures candidates à être la mesure de Lebesgue prennent la même valeurs finie pour tout élément A de la classe $\mathcal{C}$ des rectangles bornés, et cette classe vérifie (i) (par définition des boréliens), (iii) et (iv) ci-dessus.

Voici une autre application:

Corollaire 3 *Soit μ et ν deux mesures σ-finies sur $(E,\mathcal{E})$. Si elles coïncident sur une algèbre engendrant la tribu $\mathcal{E}$, elles sont égales.*

2) Les fonctions de répartition: Dans ce sous-paragraphe nous introduisons une notion relative aux mesures sur $I\!R$. Elle est particulièrement utile pour les probabilités, et nous commençons par ce cas.

Définition 4 La *fonction de répartition* d'une probabilité μ sur $I\!R$ (i.e. une mesure de masse totale $\mu(I\!R) = 1$) est la fonction F sur $I\!R$ définie par

$$F(x) \;=\; \mu(]-\infty,x]). \tag{3}$$

Proposition 5 *La fonction de répartition F d'une probabilité μ sur $I\!R$ vérifie les propriétés suivantes:*

$$\left.\begin{array}{l} F \text{ est croissante et continue à droite,} \\ \lim_{x\uparrow+\infty} F(x) = 1, \qquad \lim_{x\downarrow-\infty} F(x) = 0. \end{array}\right\} \tag{4}$$

De plus, en notant $F(x-)$ la limite à gauche de F au point x, et avec les conventions $F(-\infty) = 0$ et $F(+\infty-) = 1$ (naturelles au vu de (4))*, on a:*

$$\left.\begin{array}{lcll} \mu(]a,b]) &=& F(b) - F(a) & \text{si} \quad -\infty \le a < b < +\infty \\ \mu([a,b]) &=& F(b) - F(a-) & \text{si} \quad -\infty < a \le b < +\infty \\ \mu(]a,b[) &=& F(b-) - F(a) & \text{si} \quad -\infty \le a < b \le +\infty \\ \mu([a,b[) &=& F(b-) - F(a-) & \text{si} \quad -\infty < a < b \le +\infty. \end{array}\right\} \tag{5}$$

Preuve. Comme $]-\infty,x] \subset]-\infty,y]$ si $x \le y$, la croissance de F est évidente, et ma première égalité (5) découle de ce que $]-\infty,b] =]-\infty,a] \cup]a,b]$ si $a < b$ et de ce que la mesure de n'importe quel borélien est finie.

Pour montrer la continuité à droite, il suffit de vérifier que si x_n décroit vers x on a $F(x_n) \to F(x)$. Mais la première égalité (5) implique $F(x_n) = F(x) + \mu(]x,x_n])$ et $]x,x_n] \downarrow \emptyset$, de sorte que le résultat découle du théorème 1-14-(b). De même si $x_n \downarrow -\infty$ on a $]-\infty,x_n] \downarrow \emptyset$, donc $F(x_n) \downarrow 0$, et si $x_n \uparrow +\infty$ on a $]-\infty,x_n] \uparrow I\!R$, donc $F(x_n) \uparrow 1$: cela achève de prouver (4).

Enfin les trois dernières égalités de (5) se montrent de la même manière. Montrons par exemple la seconde: On a $]a-1/n,b] \downarrow [a,b]$, donc d'après le théorème 1-14-(b) on a $\mu([a,b]) = \lim_n \mu(]a-1/n,b]) = \lim_n(F(b) - F(a-1/n)) = F(b) - F(a-)$. $\square$

Exemples:

1) Si μ est la masse de Dirac au point a, sa fonction de répartition F est

$$F(x) \;=\; \begin{cases} 0 & \text{si } \; x < a, \\ 1 & \text{si } \; x \geq a. \end{cases}$$

2) Soit $(a_n)_{n\geq 1}$ une suite de réels, et $(b_n)_{n\geq 1}$ une suite de réels positifs de somme 1. Considérons la mesure $\mu = \sum_n b_n \varepsilon_{a_n}$, qui est une probabilité sur $I\!\!R$ puisque $\sum_n b_n = 1$ (on a $\mu(A) = \sum_{n: a_n \in A} b_n$ pour tout borélien A). La fonction de répartition F est alors

$$F(x) \;=\; \sum_{n: a_n \leq x} b_n. \tag{6}$$

Noter que cette fonction F, clairement croissante, est discontinue en tout point a_n tel que $b_n > 0$, et continue partout ailleurs.

3) Soit f une fonction positive d'intégrale $\int f d\lambda = 1$ par rapport à la mesure de Lebesgue λ, et considérons la mesure μ de densité f (rappelons que $\mu(A) = \int_A f d\lambda$ pour tout borélien A). La fonction de répartition est alors

$$F(x) \;=\; \int_{-\infty}^{x} f(y)dy. \tag{7}$$

Noter que si f est continue, alors F est dérivable, de dérivée f.

Lorsque μ est une mesure finie sur $I\!\!R$, sa fonction de répartition est encore définie par (3), et la proposition 5 est encore vraie: il faut simplement remplacer $\lim_{x\uparrow+\infty} F(x) = 1$ dans (4) par $\lim_{x\uparrow+\infty} F(x) = \mu(I\!\!R)$.

Pour les mesures infinies la situation est un peu différente, puisque la formule (3) peut fort bien donner $F(x) = \infty$ pour tout x, de sorte que dans ce cas la définition 4 n'offre aucun intérêt. Il y a cependant une notion analogue, pour les mesures dites *de Radon*: ce sont les mesures qui vérifient $\mu([-n,n]) < \infty$ pour tout entier n.

Définition 6 Soit μ une mesure sur $I\!\!R$ vérifiant $\mu([-n,n]) < \infty$ pour tout entier n. Sa *fonction de répartition généralisée* est la fonction G sur $I\!\!R$ définie par:

$$G(x) \;=\; \begin{cases} -\mu(]x,0[) & \text{si } \; x < 0 \\ \mu([0,x]) & \text{si } \; x \geq 0. \end{cases} \tag{8}$$

Proposition 7 *Soit μ une mesure sur $I\!\!R$ vérifiant $\mu([-n,n]) < \infty$ pour tout entier n. Sa fonction de répartition généralisée G est une fonction croissante, continue à droite, vérifiant $G(0-) = 0 \leq G(0)$, et on a encore (5) pour tous a, b finis, avec G au lieu de F.*

Preuve. D'abord, le fait que G vérifie (5) lorsque $-\infty < a < b < +\infty$ découle de l'additivité de μ et des propriétés suivantes:

$$0 \leq a < b \quad \Rightarrow \quad]a,b] = [0,b]\backslash[0,a], \quad \mu([0,b]) < \infty,$$

$$a < 0 \leq b \quad \Rightarrow \quad]a,b] = [a,0[\cup[0,b], \quad [a,0[\cap[0,b] = \emptyset,$$
$$a < b < 0 \quad \Rightarrow \quad]a,b] =]a,0[\backslash]b,0[, \quad \mu(]a,0[) < \infty.$$

Cela montre en particulier que G est croissante, et $G(0-) \leq 0 \leq G(0)$ est évident. Mais $]-1/n,0[\downarrow \emptyset$ et les ensembles $]-1/n,0[$ sont tous contenus dans l'ensemble $[-1,0]$, qui est de mesure finie: donc le théorème 1-14 entraine que $G(-1/n) = -\mu(]-1/n,0[) \to 0$, de sorte que $G(0-) = 0$. Les autres propriétés se montrent exactement comme dans la proposition 5. $\square$

Exemple: Si $\mu = \lambda$ est la mesure de Lebesgue, sa fonction de répartition généralisée est $G(x) = x$.

Lorsque μ est une probabilité, ou une mesure finie, les rapports entre la fonction de répartition F et la fonction de répartition généralisée G sont:

$$G(x) \;=\; F(x) - F(0-), \qquad F(x) \;=\; G(x) - \lim_{y\to-\infty} G(y). \tag{9}$$

Voici enfin le résultat d'unicité qui montre qu'une mesure de Radon sur $I\!R$ est entièrement caractérisée par sa fonction de répartition généralisée:

Théorème 8 *Deux mesures μ et ν sur $(I\!R,\mathcal{R})$, finies sur les ensembles $[-n,n]$ pour tout entier n, et qui ont même fonction de répartition généralisée sont égales. Le même résultat est vrai si elles sont finies et ont même fonction de répartition.*

Preuve. Il suffit d'apliquer le théorème 1 avec la classe/$\mathcal{C}$ constituée de tous les intervalles de la forme $]x,y]$ pour $-\infty < x < y < +\infty$: on a évidemment (i), (iii) et (iv), tandis que (ii) vient de ce que $\mu(]x,y]) = G(y) - G(x) = \nu(]x,y])$. $\square$

Nous terminons ce paragraphe en énonçant un résultat, qui avec le théorème précédent implique le théorème 1-19, et qui sera démontré à la fin du cours:

Théorème 9 *Si G est une fonction de $I\!R$ dans $I\!R$, croissante, continue à droite, telle que $G(0-) = 0$, il existe une mesure μ* (et une seule d'après le théorème précédent) *qui admet G pour fonction de répartition généralisée.*

2 Produit d'espaces mesurables

1) La tribu produit: Nous considérons ci-dessous une famille d'espaces mesurables $(E_i,\mathcal{E}_i)_{1\leq i\leq d}$, avec un entier $d \geq 2$. Soit le produit $F = \prod_{i=1}^d E_i$, c'est-à-dire l'ensemble des suites à d éléments $(x_1,\ldots,x_d)$ (on dit aussi les "d-uplets") où, pour chaque i, x_i parcourt l'ensemble E_i. L'exemple le plus courant est celui où $(E_i,\mathcal{E}_i) = (I\!R,\mathcal{R})$, auquel cas $F = I\!R^d$.

On appelle $j^{\text{ème}}$ *application coordonnée* l'application

$$Y_j : \; F \;\to\; E_j \quad \text{définie par} \quad Y_j(x_1,\ldots,x_d) \;=\; x_j. \tag{10}$$

Un *pavé mesurable* est une partie de F la forme $A = \prod_{i=1}^d A_i$, où $A_i \in \mathcal{E}_i$ pour tout i. La *base* du pavé A est l'ensemble J des indices i tels que $A_i \neq E_i$, et sa *dimension* est le nombre de points de J.

Définition 10 La *tribu produit* des $\mathcal{E}_i$ est la plus petite tribu $\mathcal{F}$ de F telle que chaque application coordonnée Y_i soit mesurable de $(F, \mathcal{F})$ dans $(E_i, \mathcal{E}_i)$, c'est-à-dire la tribu de F engendrée par la réunion de tribus $\cup_{i=1}^d Y_i^{-1}(\mathcal{E}_i)$. On la note aussi $\mathcal{F} = \otimes_{i=1}^d \mathcal{E}_i = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_d$.

Lorsque tous les $(E_i, \mathcal{E}_i)$ sont égaux à un même espace $(E, \mathcal{E})$ on écrit aussi $F = E^d$ et $\mathcal{F} = \mathcal{E}^{\otimes d}$.

Proposition 11 *La tribu produit $\mathcal{F}$ est aussi engendrée par chacune des classes suivantes de parties de F:*

a) la classe des pavés mesurables;

b) la classe des pavés mesurables de dimension 1.

Preuve. Soit $\mathcal{A}$ la classe de tous les pavés mesurables, et $\mathcal{B}$ celle des pavés mesurables de dimension 1. Si $A = \prod_{i=1}^d A_i$ est dans $\mathcal{A}$, on a aussi $A = \cap_{i=1}^d Y_i^{-1}(A_i)$ (vérification immédiate), donc $A \in \mathcal{F}$ et finalement $\mathcal{A} \subset \mathcal{F}$. On a aussi $\mathcal{B} \subset \mathcal{A}$, de sorte qu'il reste à montrer que $\sigma(\mathcal{B})$ contient $\mathcal{F}$. Pour cela, il suffit clairement de montrer, vu la définition de $\mathcal{F}$, que chaque tribu $Y_i^{-1}(\mathcal{E}_i)$ est contenue dans $\mathcal{B}$; mais si $A_i \in \mathcal{E}_i$ l'image réciproque $Y_i^{-1}(A_i)$ est le pavé mesurable B de dimension 1 donné par $B = \prod_{i=1}^d B_i$, avec $B_i = A_i$ et $B_j = E_j$ si $j \neq i$: comme $B \in \mathcal{B}$, cela achève la démonstration. $\square$

Corollaire 12 *La tribu borélienne $\mathcal{R}^d$ de $\mathbb{R}^d$ égale la tribu produit $\mathcal{R}^{\otimes d}$.*

Preuve. D'après la définition 1-10 la tribu borélienne $\mathcal{R}^d$ est engendrée par la classe des pavés $A = \prod_{i=1}^d A_i$ avec des A_i qui sont des ouverts: on a donc $\mathcal{R}^d \subset \mathcal{R}^{\otimes d}$.

Pour montrer l'inclusion inverse, vu la définition 10, il suffit de vérifier que chaque application Y_i est mesurable de $(\mathbb{R}^d, \mathcal{R}^d)$ dans $(\mathbb{R}, \mathcal{R})$, i.e. est borélienne; mais comme Y_i est continue, elle est aussi borélienne (cf. la proposition 2-4), d'où le résultat. $\square$

Un autre résultat important est *l'associativité* du produit de tribus. Soit k un entier entre 1 et $d-1$. Soit le produit $F_1 = E_1 \times \ldots \times E_k$ des k premiers facteurs, muni de la tribu produit $\mathcal{F}_1 = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_k$ (si $k = 1$, cela se réduit à $F_1 = E_1$ et $\mathcal{E}_1 = \mathcal{F}_1$), et de même $F_2 = E_{k+1} \times \ldots \times E_d$ avec la tribu $\mathcal{F}_2 = \mathcal{E}_{k+1} \otimes \ldots \otimes \mathcal{E}_d$. On a bien-sûr $F = F_1 \times F_2$ (en identifiant le "couple" $((x_1, \ldots, x_k), (x_{k+1}, \ldots, x_d))$ et le "d-uplet" $(x_1 \ldots, x_d)$), ainsi que:

Proposition 13 *Les tribus produits $\mathcal{F}_1 \otimes \mathcal{F}_2$ et $\mathcal{F} = \otimes_{i=1}^d \mathcal{E}_i$ sont égales.*

A titre d'exemple, on déduit de cette proposition et du corollaire précédent que $\mathcal{R}^{n+m} = \mathcal{R}^n \otimes \mathcal{R}^m$

Preuve. Soit $A = \prod_{i=1}^d A_i$ avec $A_i \in \mathcal{E}_i$ un pavé mesurable de F. On peut écrire $A = B_1 \times B_2$, avec $B_1 = \prod_{i=1}^k A_i$ et $B_2 = \prod_{i=k+1}^d A_i$. La proposition 11 entraine $B_1 \in \mathcal{F}_1$ et $B_2 \in \mathcal{F}_2$, donc aussi $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$; une nouvelle application de cette proposition entraine que $\mathcal{F} = \otimes_{i=1}^d \mathcal{E}_i$ est contenue dans $\mathcal{F}_1 \otimes \mathcal{F}_2$.

Il reste à montrer que $\mathcal{F}_1 \otimes \mathcal{F}_2 \subset \mathcal{F}$. Pour cela, notons $\mathcal{F}'$ la classe de tous les ensembles $A \subset F_1$ tels que $A \times F_2 \in \mathcal{F}$. Il est immédiat de vérifier que $\mathcal{F}'$ est une tribu. Par ailleurs si C est un pavé mesurable de F_1, le produit $C \times F_2$ est un pavé mesurable de F, donc $C \times F_2 \in \mathcal{F}$, donc $C \in \mathcal{F}'$: on déduit de la proposition 11 que $\mathcal{F}'$ contient la tribu $\mathcal{F}_1$, ce qui veut dire que $A \times F_2 \in \mathcal{F}$ pour tout $A \in \mathcal{F}_1$; on montre de même que $F_1 \times B \in \mathcal{F}$ dès que $B \in \mathcal{F}_2$. Par suite $A \times B = (A \times F_2) \cap (F_1 \times B)$ est dans $\mathcal{F}$ dès que $A \in \mathcal{F}_1$ et $B \in \mathcal{F}_2$: une dernière application de la proposition 11 entraine alors que $\mathcal{F}_1 \otimes \mathcal{F}_2 \subset \mathcal{F}$, et la preuve est achevée. $\square$

2) Les fonctions mesurables: Passons maintenant à l'étude des applications mesurables. On suppose toujours que $F = \prod_{i=1}^d E_i$ est muni de la tribu produit $\mathcal{F} = \otimes_{i=1}^d \mathcal{E}_i$. Il y a deux aspects, selon qu'on considère une application f d'un espace G dans le produit F, ou une application f du produit F dans un espace G.

Commençons par le cas où f est une application de G dans F. De manière équivalente on peut la considérer comme une collection $(f_1, \ldots, f_d)$, où chaque f_i est une application de G dans E_i: f_i est appelée la $i^{\text{ème}}$ application coordonnée de f (une autre manière d'écrire ceci est $f_i = Y_i \circ f$, avec la notation (10)).

Proposition 14 *Soit $(G, \mathcal{G})$ un espace mesurable. Une application f de G dans F est mesurable relativement aux tribus $\mathcal{G}$ et $\mathcal{F}$ si et seulement si chaque application coordonnée f_i est mesurable de $(G, \mathcal{G})$ dans $(E_i, \mathcal{E}_i)$.*

Preuve. Comme $f_i = Y_i \circ f$ et comme la composée de deux applications mesurables est mesurable (proposition 2-3), si f est mesurable chaque f_i est aussi mesurable.

Supposons inversement chaque f_i mesurable. Pour montrer la mesurabilité de f il suffit (cf. proposition 2-2) de montrer que $f^{-1}(A) \in \mathcal{G}$ pour tout A dans une classe $\mathcal{A}$ de parties de F qui engendre la tribu $\mathcal{F}$. On va prendre pour $\mathcal{A}$ la classe des pavés mesurables de dimension 1 (cf. proposition 11): un tel pavé s'écrit $A = Y_i^{-1}(B)$ pour un i et un $B \in \mathcal{E}_i$. Mais $f_i = Y_i \circ f$ entraîne $f^{-1}(A) = f_i^{-1}(B)$, qui appartient à $\mathcal{G}$ par la mesurabilité de f_i: on a donc le résultat. $\square$

A l'inverse on considère maintenant, dans le cas où $d = 2$ seulement pour simplifier, une application f de $F = E_1 \times E_2$ dans un espace G. On lui associe les familles $(f_{x_1}^{(2)} : x_1 \in E_1)$ et $(f_{x_2}^{(1)} : x_2 \in E_2)$ d'applications de E_2 et E_1 respectivement dans G, définies par

$$f_{x_1}^{(2)}(x_2) = f(x_1, x_2), \qquad f_{x_2}^{(1)}(x_1) = f(x_1, x_2). \tag{11}$$

Proposition 15 *Si f est une application mesurable de $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ dans $(G, \mathcal{G})$, pour tout $x_1 \in E_1$ (resp. $x_2 \in E_2$) l'application $f_{x_1}^{(2)}$ (resp. $f_{x_2}^{(1)}$) est mesurable de $(E_2, \mathcal{E}_2)$ (resp. $(E_1, \mathcal{E}_1)$) dans $(G, \mathcal{G})$.*

Preuve. On va montrer, par exemple, que $g = f^{(2)}_{x_1}$ pour un $x_1 \in E_1$ fixé est mesurable de $(E_2, \mathcal{E}_2)$ dans $(G, \mathcal{G})$.

Soit $B \in \mathcal{G}$. Nous devons montrer que $g^{-1}(B) \in \mathcal{E}_2$. Si à toute partie A de $E_1 \times E_2$ on associe la partie A' de E_2 définie par $A' = \{x_2 \in E_2 : (x_1, x_2) \in A\}$ (rappelons que x_1 est fixé), on a $g^{-1}(B) = C'$ si $C = f^{-1}(B)$, et on sait que $C \in \mathcal{E}_1 \otimes \mathcal{E}_2$. Il reste donc à montrer que si $A \in \mathcal{E}_1 \otimes \mathcal{E}_2$, alors $A' \in \mathcal{E}_2$.

Pour cela, soit $\mathcal{C}$ la classe des parties A du produit $E_1 \times E_2$ telles que $A' \in \mathcal{E}_2$. Cette classe est évidemment une tribu, et elle contient les ensembles $A = A_1 \times A_2$ où $A_i \in \mathcal{E}_i$ (car alors $A' = A_2$ si $x_1 \in A_1$ et $A' = \emptyset$ sinon), donc elle contient la tribu $\mathcal{E}_1 \otimes \mathcal{E}_2$ par la proposition 11: la preuve est achevée. $\square$

En combinant cette proposition et la proposition 13, on voit que si f est une application mesurable de $(\prod_{i=1}^d E_i, \otimes_{i=1}^d \mathcal{E}_i)$ dans $(G, \mathcal{G})$, si $k \in \{1, \ldots d-1\}$ et si les $x_i \in E_i$ sont fixés pour $i = k+1, \ldots, d$, alors l'application

$$(x_1, \ldots, x_k) \;\mapsto\; f(x_1, \ldots, x_k, x_{k+1}, \ldots, x_d)$$

est mesurable de $(\prod_{i=1}^k E_i, \otimes_{i=1}^k \mathcal{E}_i)$ dans $(G, \mathcal{G})$. En particulier, si f est une fonction borélienne sur $\mathbb{R}^d$, la fonction ci-dessus (avec $x_{k+1}, \ldots, x_d$ fixés) est borélienne sur $\mathbb{R}^k$.

Remarque: La "réciproque" de la proposition précédente est **fausse**: les applications $f^{(2)}_{x_1}$ et $f^{(1)}_{x_2}$ peuvent être mesurables pour tous x_1, x_2 sans que l'application f soit mesurable par rapport à la tribu produit $\mathcal{E}_1 \otimes \mathcal{E}_2$. Par exemple si $E_1 = E_2 = \mathbb{R}$ est muni de la tribu $\mathcal{E}$ engendrée par les singletons $\{x\}$ (c'est une tribu "beaucoup plus petite" que la tribu borélienne, puisqu'elle ne contient aucun intervalle de longueur finie et non nulle), la fonction $f = 1_\Delta$ indicatrice de la diagonale $\Delta = \{(x,x) : x \in \mathbb{R}\}$ sur $\mathbb{R}^2$ n'est pas mesurable par rapport à $\mathcal{E} \otimes \mathcal{E}$, alors que les fonctions $f^{(2)}_{x_1}$ et $f^{(1)}_{x_2}$ sont $\mathcal{E}$-mesurables. $\square$

3 Produit de mesures

1) Le produit de deux mesures: Soit $(E_1, \mathcal{E}_1, \mu_1)$ et $(E_2, \mathcal{E}_2, \mu_2)$ deux espaces mesurés. On va construire le "produit" des deux mesures μ_1 et μ_2 sur l'espace $F = E_1 \times E_2$ muni de la tribu $\mathcal{F} = \mathcal{E}_1 \otimes \mathcal{E}_2$. Les résultats sont rassemblés dans deux théorèmes, qu'on démontrera simultanément:

Théorème 16 *Si les deux mesures μ_1 et μ_2 sont σ-finies, il existe une mesure μ et une seule sur $(F, \mathcal{F})$, qu'on note aussi $\mu = \mu_1 \otimes \mu_2$ et qu'on appelle la* mesure produit, *qui vérifie*

$$\mu(A_1 \times A_2) \;=\; \mu_1(A_1)\mu_2(A_2) \qquad \forall A_1 \in \mathcal{E}_1, \quad A_2 \in \mathcal{E}_2. \tag{12}$$

Théorème 17 (THEOREME DE FUBINI) *Supposons que les deux mesures μ_1 et μ_2 soient σ-finies, et soit $\mu = \mu_1 \otimes \mu_2$.*

a) Si f est une fonction mesurable sur $(F, \mathcal{F})$ à valeurs dans $\overline{\mathbb{R}}_+$, les fonctions

$$f_1(x_1) \;=\; \int f(x_1, x_2)\mu_2(dx_2), \qquad f_2(x_2) \;=\; \int f(x_1, x_2)\mu_1(dx_1) \tag{13}$$

(en vertu de la proposition 15 ces intégrales sont bien définies) *sont mesurables sur* $(E_1,\mathcal{E}_1)$ *et* $(E_2,\mathcal{E}_2)$ *respectivement, et on a*

$$\int f d\mu \;=\; \int \mu_1(dx_1)\left(\int f(x_1,x_2)\mu_2(dx_2)\right) \;=\; \int \mu_2(dx_2)\left(\int f(x_1,x_2)\mu_1(dx_1)\right). \quad (14)$$

b) Si f est une fonction mesurable f sur $(F,\mathcal{F})$ à valeurs dans $\overline{I\!R}$, les trois assertions suivantes sont équivalentes:

(i) f est intégrable par rapport à μ;

(ii) la fonction $x_1 \mapsto \int |f(x_1,x_2)|\mu_2(dx_2)$ est intégrable par rapport à μ_1;

(iii) la fonction $x_2 \mapsto \int |f(x_1,x_2)|\mu_1(dx_1)$ est intégrable par rapport à μ_2.

Dans ce cas, l'ensemble $B_1 = \{x_1 : \int |f(x_1,x_2)|\mu_2(dx_2) < \infty\}$ est $\mathcal{E}_1$-mesurable et vérifie $\mu_1((B_1)^c) = 0$ et l'ensemble $B_2 = \{x_2 : \int |f(x_1,x_2)|\mu_1(dx_1) < \infty\}$ est $\mathcal{E}_2$-mesurable et vérifie $\mu_2((B_2)^c) = 0$. La fonction f_1 (resp. f_2) de (13) est alors bien définie sur B_1 (resp. B_2), et on a (14).

Il semble utile de faire d'emblée quelques commentaires. Considérons par exemple la première des formules (14): en toute rigueur, il faudrait l'écrire

$$\int f d\mu \;=\; \int f_1 d\mu_1, \qquad \text{avec } f_1 \text{ définie par (13).} \quad (15)$$

Lorsque $f \geq 0$ la fonction f_1 est bien définie, mesurable et positive, de sorte que les deux membres de (14) ont un sens. Lorsque f est de signe quelconque, mais intégrable par rapport à μ, (13) définit $f_1(x_1)$ pour $x_1 \in B_1$, tandis que $f_1(x_1)$ risque de ne pas avoir de sens si $x \notin B_1$; toutefois la fonction f_1' égale à f_1 sur B_1 et (par exemple) à 0 sur $(B_1)^c$ est $\mathcal{E}_1$-mesurable, et l'intégrale $\int f_1' d\mu_1$ ne dépend pas des valeurs de f_1' sur l'ensemble μ_1-négligeable $(B_1)^c$ (cf. la proposition 3-7): il est alors naturel de l'écrire $\int f_1 d\mu_1$ (par un abus - anodin - de notation), et c'est le sens qu'on donne au second membre de (15).

Preuve. 1) Par hypothèse il existe des suites $(C_n)_{n\geq 1}$ dans $\mathcal{E}_1$ et $(D_n)_{n\geq 1}$ dans $\mathcal{E}_2$, telles que $C_n \uparrow E_1$, $D_n \uparrow E_2$, $\mu_1(C_n) < \infty$ et $\mu_2(D_n) < \infty$ pour tout n.

2) Nous allons maintenant montrer que si f est une fonction mesurable positive sur $(F,\mathcal{F})$, les fonctions f_1 et f_2 de (13) sont mesurables. On va traiter, par exemple, le cas de f_1.

Par limite croissante (cf. le lemme 2-13 et (iv) du théorème 2-16), il suffit de montrer le résultat lorsque f est étagée; par linéarité (cf. la proposition 2-11) il suffit même de le montrer lorsque $f = 1_A$ est l'indicatrice d'un $A \in \mathcal{F}$.

Soit μ_2^n la restriction de μ_2 à D_n (donc $\mu_2^n(B) = \mu_2(B \cap D_n)$). On a $\mu(B) = \lim_n \uparrow \mu_2^n(B)$, de sorte que si $f = 1_A$ la quantité $f_1(x_1)$ est la limite croissante des intégrales de la fonction $x_2 \mapsto 1_A(x_1,x_2)$ par rapport aux μ_2^n. Il suffit donc de montrer la mesurabilité de f_1 lorsqu'on remplace μ_2 par μ_2^n: en d'autres termes on peut supposer que la mesure μ_2 est finie.

Notons $\mathcal{D}$ la classe des $A \in \mathcal{F}$ tels que la fonction f_1 associée à $f = 1_A$ soit $\mathcal{E}_1$-mesurable. Comme μ_2 est supposée finie, il est évident de vérifier que cette classe vérifie (1) et (2), c'est-à-dire est un λ-système. Par ailleurs si $A = A_1 \times A_2$ est un pavé mesurable, on a $f_1 = \mu_2(A_2)1_{A_1}$, qui est $\mathcal{E}_1$-mesurable, de sorte que $\mathcal{D}$ contient la classe $\mathcal{C}$ des pavés mesurables. Comme la classe $\mathcal{C}$ est stable par intersection et contient F lui-même, une application du lemme 2 montre que $\mathcal{D} = \mathcal{F}$, et a prouvé le résultat cherché.

3) Montrons maintenant l'existence d'une mesure μ sur $(F, \mathcal{F})$ vérifiant (12). D'après 2) on peut poser pour tout $A \in \mathcal{F}$:

$$\mu(A) \;=\; \int \mu_1(dx_1) \left(\int 1_A(x_1, x_2)\mu_2(dx_2) \right). \tag{16}$$

Il est clair que $\mu(\emptyset) = 0$, et la σ-additivité de μ découle d'une double application du corollaire 2-17. Le fait que μ vérifie (12) est évident.

4) Passons à l'unicité. Soit μ et μ' deux mesures vérifiant (12). Elles coïncident donc sur les pavés mesurables. Pour obtenir que $\mu = \mu'$ il suffit alors d'appliquer le théorème 1 à la classe $\mathcal{C}$ des pavés mesurables $A = A_1 \times A_2$ tels que $\mu(A) < \infty$ (i.e. $\mu_i(A_i) < \infty$ pour $i = 1, 2$): cette classe vérifie évidemment les conditions (ii) et (iii) de ce théorème; elle vérifie (iv) avec la suite $F_n = C_n \times D_n$; enfin elle vérifie (i), puisque tout pavé mesurable A est réunion des pavés $A \cap F_n$ qui appartiennent à $\mathcal{C}$, de sorte que tout pavé mesurable est dans la tribu $\sigma(\mathcal{C})$, et donc $\sigma(\mathcal{C}) = \mathcal{F}$ par la proposition 11.

5) Pour le moment on a prouvé le théorème 16, et la première partie de (a) du théorème 17. Montrons maintenant (14) lorsque f est positive. Quand $f = 1_A$ la première de ces formules est exactement (16). Par linéarité on en déduit la première formule (14) pour toute fonction étagée, puis par limite croissante pour toute fonction mesurable positive. L'égalité entre les membres extrêmes de (14) se montre de la même manière.

6) Il reste à montrer la partie (b) du théorème 17. L'équivalence de (i), (ii) et (iii) découle immédiatement de (14) appliquée à $|f|$. Le fait que $B_1 \in \mathcal{E}_1$ vient de la mesurabilité de la fonction $x_1 \mapsto \int |f(x_1, x_2)|\mu_2(dx_2)$, et $\mu_1((B_1)^c) = 0$ vient de (ii) et du corollaire 3-9. On a de même les résultats concernant B_2. Enfin la validité de (14) pour f provient de l'application de (14) aux fonctions positives f^+ et f^- et du fait que $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$. $\square$

Exemples:

1) Lorsque $(E_1, \mathcal{E}_1) = (E_2, \mathcal{E}_2) = (\mathbb{R}, \mathcal{R})$, on a vu que $(F, \mathcal{F}) = (\mathbb{R}^2, \mathcal{R}^2)$. Si de plus $\mu_1 = \mu_2 = \lambda$ est la mesure de Lebesgue, le produit $\mu_1 \otimes \mu_2$ est alors la mesure de Lebesgue λ_2 sur $\mathbb{R}^2$, et les théorèmes 1-20 et 1-21 découlent du théorème 16 lorsque $d = 2$. L'intégrale d'une fonction f sur $\mathbb{R}^2$ par rapport à λ_2 se note aussi

$$\int f d\lambda_2 \;=\; \int\int f(x, y)dxdy$$

et la formule (14) est ainsi une version améliorée du résultat selon lequel une intégrale double se calcule comme une succession de deux intégrales “simples”, dans l'ordre qu'on veut: attention toutefois aux hypothèses sur f pour que cette formule soit exacte.

2) Lorsque $(E_1, \mathcal{E}_1) = (E_2, \mathcal{E}_2) = (\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$ et lorsque $\mu_1 = \mu_2$ est la mesure de comptage sur $\mathbb{N}^*$, le produit $\mu = \mu_1 \otimes \mu_2$ est la mesure de comptage sur $(\mathbb{N}^*)^2$.

L'intégrale d'une fonction (positive ou intégrable) par rapport à la mesure de comptage étant la somme des valeurs prises par cette fonction, la formule (14) devient dans ce cas:

$$\sum_{n,m\in \mathbb{N}^*} u_{n,m} = \sum_{n=1}^{\infty}\left(\sum_{m=1}^{\infty} u_{n,m}\right) = \sum_{m=1}^{\infty}\left(\sum_{n=1}^{\infty} u_{n,m}\right), \tag{17}$$

à condition que $u_{n,m} \geq 0$ pour tous n, m, ou que que $\sum_{n,m\in \mathbb{N}^*} |u_{n,m}| < \infty$ si les $u_{n,m}$ sont de signe quelconque. On retrouve en particulier la formule 2-(26).

3) Soit $(E_1, \mathcal{E}_1, \mu_1)$ un espace mesuré quelconque avec une mesure μ_1 σ-finie, et soit $(E_2, \mathcal{E}_2) = (\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$ muni de la mesure de comptage μ_2. Une fonction f sur $F = E_1 \times E_2$ peut être considérée comme une suite $(f_n)_{n\geq 1}$ de fonctions sur E_1 par les formules $f_n(x) = f(x, n)$, et on vérifie aisément que f est mesurable par rapport à $\mathcal{F} = \mathcal{E}_1 \otimes \mathcal{E}_2$ si et seulement si les fonctions f_n sont $\mathcal{E}_1$-mesurables. La fonction mesurable f est intégrable par rapport à $\mu = \mu_1 \otimes \mu_2$ si et seulement si on a

$$\int (\sum_{n\geq 1} |f_n -) f \mu_1 = \sum_{n\geq 1} \int |f_n| d\mu_1 < \infty \tag{18}$$

(appliquer (14) à $|f|$; la première égalté vient du corollaire 2-17). Si on on a (18), la série $\sum_{n\geq 1} f_n$ est donc μ_1-a.s. absolument convergente, de somme μ_1-intégrable, et la formule (14) appliquée à f donne alors

$$\int f d\mu = \sum_{n\geq 1} \int f_n d\mu_1 = \int \left(\sum_{n\geq 1} f_n\right) d\mu_1. \tag{19}$$

Ainsi, sous (18), on peut intervertir somme et intégrale: on obtient ainsi une version un peu différente du corollaire 2-17, avec des f_n de signe quelconque mais vérifiant (18).

Remarque 1: La mesurabilité de f par rapport à la tribu produit est essentielle dans le théorème 17. On peut trouver des fonctions positives f qui ne sont pas $\mathcal{F}$ mesurables mais qui sont "séparément" mesurables en chacune des variables (cf. la remarque de la fin du paragraphe 2), et telles que les fonctions f_i de (13) soient également mesurables: les deux derniers membres de (14) sont alors bien définis, mais pas nécessairement égaux, tandis que le premier n'a pas de sens. $\square$

Remarque 2: Même lorsque f est mesurable, il faut faire très attention quand on utilise (14), qui n'est vraie que si f est de signe constant, ou est intégrable.

Illustrons ceci dans le cadre de l'exemple 1 ci-dessus. Soit

$$f(x,y) = \begin{cases} \frac{x-y}{(x+y)^3} & \text{si } x, y \in]0,1] \\ 0 & \text{sinon.} \end{cases}$$

On a alors

$$\int dx \left(\int f(x,y)dy\right) = \int_0^1 dx \int_0^1 \left(\frac{2x}{(x+y)^3} - \frac{1}{(x+y)^2}\right) dy = \int_0^1 \frac{1}{(1+x)^2} dx = \frac{1}{2},$$

et un calcul analogue conduit à $\int dy\,(\int f(x,y)dx) = -\frac{1}{2}$. Les deux derniers membres de (14) sont donc différent (bien-sûr la fonction borélienne f sur $\mathbb{R}^2$ n'est pas λ_2-intégrable.

Pire: les deux derniers membres de (14) peuvent être égaux, alors que l'intégrale de f n'a pas de sens. Prenons par exemple la fonction g sur $]0,\infty[$ définie par $g(x) = x^{-1/2}$ si $x \leq 1$ et $g(x) = x^{-2}$ si $x > 1$, de sorte que $a = \int_0^\infty g(x)dx$ est finie. Soit

$$f(x,y) \;=\; \begin{cases} g(x-y) & \text{si} \quad x > y \\ 0 & \text{si} \quad x = y \\ -g(y-x) & \text{si} \quad x < y. \end{cases}$$

Il est clair que $\int f(x,y)dx = \int f(x,y)dy = a - a = 0$, donc les deux derniers membres de (14) sont nuls. Cependant $f^+(x,y) = g(x-y)1_{\{x>y\}}$, donc (14) appliqué à la fonction positive f^+ donne

$$\int f^+ d\lambda_2 \;=\; \int_{-\infty}^{+\infty} dy \int_y^{+\infty} g(x-y)dx \;=\; \int_{-\infty}^{+\infty} a dy \;=\; +\infty,$$

et de même pour f^-: donc l'intégrale de f par rapport à λ_2 n'a pas de sens.

2) Le produit de plusieurs mesures On considère maintenant une famille *finie* d'espaces mesurés $(E_i, \mathcal{E}_i, \mu_i)$, pour $i = 1, \ldots, n$. On pose $F = \prod_{i=1}^n E_i$, muni de la tribu produit $\otimes_{i=1}^n \mathcal{E}_i$.

Théorème 18 *Si les mesures μ_i sont toutes σ-finies, il existe une mesure μ et une seule sur $(F, \mathcal{F})$, qu'on note aussi $\mu = \mu_1 \otimes \ldots \otimes \mu_n = \otimes_{i=1}^n \mu_i$ et qu'on appelle la* mesure produit, *qui vérifie*

$$\mu(\prod_{i=1}^n A_i) \;=\; \prod_{i=1}^n \mu_i(A_i) \qquad \forall A_i \in \mathcal{E}_i. \tag{20}$$

Preuve. On fait une récurrence sur n (le résultat étant vrai pour $n = 2$ d'après le théorème 16). Supposons le résultat vrai pour $n-1$: sur l'espace $F' = \prod_{i=1}^{n-1} E_i$ muni de la tribu $\mathcal{F}' = \otimes_{i=1}^{n-1} \mathcal{E}_i$ on a construit la mesure produit μ', qui est l'unique mesure vérifiant

$$\mu'(\prod_{i=1}^{n-1} A_i) \;=\; \prod_{i=1}^{n-1} \mu_i(A_i) \quad \forall A_i \in \mathcal{E}_i.$$

On a $F = F' \times E_n$ et, par la proposition 13, $\mathcal{F} = \mathcal{F}' \otimes \mathcal{E}_n$. Le théorème 16 permet de construire sur $(F, \mathcal{F})$ la mesure produit $\mu = \mu' \otimes \mu_n$, qui vérifie clairement (20). Enfin, l'unicité de μ se montre exactement comme pour le théorème 16. $\square$

Exemple: La mesure de Lebesgue λ_d sur $\mathbb{R}^d$ est ainsi la mesure produit - d fois - de la mesure de Lebesgue sur $\mathbb{R}$, et les théorèmes 1-20 et 1-21 découlent du théorème précédent.

Nous avons vu l'associativité du produit des tribus (proposition 13). La même propriété est vraie pour les produits de mesure, en utilisant les notations $F_1 = \prod_{i=1}^k E_i$, $\mathcal{F}_1 = \otimes_{i=1}^k \mathcal{E}_i$ et $\nu_1 = \otimes_{i=1}^k \mu_i$, ainsi que $F_2 = \prod_{i=k+1}^n$, $\mathcal{F}_2 = \otimes_{i=k+1}^n \mathcal{E}_i$ et $\nu_2 = \otimes_{i=k+1}^n \mu_i$:

Corollaire 19 *Les mesures produits $\nu_1 \otimes \nu_2$ et $\otimes_{i=1}^n \mu_i$ sont égales.*

Preuve. Il suffit de remarquer que ces deux mesures coïncident sur les pavés mesurables de $(F, \mathcal{F})$, donc sont égales d'après l'unicité dans le théorème précédent. $\square$

Etant donné ce corollaire, le théorème de Fubini se généralise immédiatement au produit fini $\mu = \otimes_{i=1}^n \mu_i$ par une récurrence immédiate. Plus précisément, si f est une fonction mesurable sur $(F, \mathcal{F})$, on a

$$\int f d\mu \;=\; \int \mu(dx_1) \left(\int \mu(dx_2) \left(\ldots \int f(x_1, \ldots, x_n) \mu(dx_n) \ldots \right) \right), \tag{21}$$

lorsqu'en plus f est positive ou intégrable par rapport à μ, et de plus f est intégrable si et seulement si le membre de droite de (21) écrit pour $|f|$ est fini.

Lorsque la fonction f se met sous la forme $f(x_1, \ldots, x_n) = \prod_{i=1}^n f_i(x_i)$ (on écrit aussi $f = \otimes_{i=1}^n f_i$, et c'est d'ailleurs là l'origine de la notation $\otimes$ pour les produits de tribus ou de mesures), (21) prend une forme bien plus agréable:

Proposition 20 *Soit f_i des fonctions mesurables sur $(E_i, \mathcal{E}_i)$, et supposons les mesures μ_i σ-finies. Soit $f(x_1, \ldots, x_n) = \prod_{i=1}^n f_i(x_i)$ et $\mu = \otimes_{i=1}^n \mu_i$.*

a) La fonction f est μ-intégrable si et seulement si on a l'une des deux conditions suivantes:

(i) la fonction f_i est μ_i-intégrable pour tout $i = 1, \ldots, n$;

(ii) il existe un indice i tel que la fonction f_i soit μ_i-p.p. égale à 0.

b) Si toutes les fonctions f_i sont positives, ou si l'une des deux conditions de (a) sont remplies, on a

$$\int f d\mu \;=\; \prod_{i=1}^n \int f_i d\mu_i. \tag{22}$$

Preuve. D'abord, lorsque les f_i sont positives la formule (22) découle immédiatement de (21) (on peut aussi faire une preuve "directe": (22) n'est autre que (20) lorsque les f_i sont des fonctions indicatrices; par linéarité la formule (22) est donc vraie lorsque les f_i sont étagées, puis par limite croissante lorsque les f_i sont mesurables positives).

L'assertion (a) découle de la formule (22) appliquée aux valeurs absolues $|f_i|$ (en se rappelant que l'intégrale d'une fonction positive est nulle si et seulement si cette fonction est presque partout nulle), et (22) pour f intégrable de signe quelconque se déduit de (22) appliqué à toutes les combinaisons possibles des f_i^+ et f_i^-. $\square$

Voici une remarque évidente: la masse totale de la mesure produit égale le produit des masses totales (appliquer (20) avec $A_i = E_i$). Par exemple, le produit d'un nombre fini de probabilités est une probabilité.

Mais cette remarque explique pourquoi on ne fait pas en général de produit infini de mesures, sauf lorsqu'il s'agit de probabilités: si on se donne une suite infinie $(\mu_n)_{n \geq 1}$ de mesures σ-finies (chacune définie sur un espace mesurable $(E_n, \mathcal{E}_n)$), et si on cherche à définir la mesure produit sur les pavés mesurables de $F = \prod_{n\geq 1} E_n$ par la formule (20), le second membre devient un produit infini qui, en général, diverge. Cependant, si les μ_n sont toutes des probabilités, il est possible de définir le produit infini $\otimes_{n\geq 1}\mu_n$ par cette formule (nous nous contentons de cette remarque un peu informelle; la démonstration du résultat est en fait difficile).

4 La formule de changement de variable

Ce paragraphe est essentiellement consacré à la démonstration de la formule "de changement de variable" dans les intégrales par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$. Cela permettra d'étudier la mesure image d'une mesure sur $\mathbb{R}^n$ ayant une densité.

Le cadre est le suivant: soit D et Δ deux ouverts de $\mathbb{R}^n$, et h un C^1-*difféomorphisme de* Δ *dans* D, c'est-à-dire une application h de Δ dans D qui est bijective et continuement différentiable et dont l'application réciproque h^{-1} (de D dans Δ) est aussi continuement différentiable. On note $h_i(x) = h_i(x_1, \dots, x_n)$ la $i^{\text{ème}}$ coordonnée de $h(x)$. On appelle *matrice jacobienne* en $x \in \Delta$ la matrice des dérivées partielles $(\partial h_i / \partial x_j)_{1\leq i,j\leq n}$ prise au point x, et *jacobien* de h le déterminant de cette matrice: ce déterminant est noté $Dh(x)$.

En dérivant les deux membres de l'égalité $h^{-1} \circ h(x) = x$ on vérifie immédiatement que les matrices jacobiennes de h en x et de h^{-1} en $h(x)$ sont inverses l'une de l'autre. Par suite on a

$$Dh(x) Dh^{-1}(h(x)) \;=\; 1 \qquad \forall x \in \Delta. \tag{23}$$

Rappelons enfin que l'intégrale d'une fonction f sur $\mathbb{R}^n$ par rapport à la mesure de Lebesgue est notée $\int f(x)\lambda_d(dx)$, notation qu'on abrège en $\int f(x)dx$, ou qu'on remplace aussi par $\int f(x_1, \dots, x_n)dx_1 \dots dx_n$; l'intégrale de la fonction $f1_A$ lorsque $A \in \mathcal{R}^d$ est aussi notée $\int_A f(x)dx$ ou $\int_A f(x_1, \dots, x_n)dx_1 \dots dx_n$.

Théorème 21 *Sous les hypothèses précédentes, pour toute fonction borélienne f sur $\mathbb{R}^d$ telle que $f1_D$ soit intégrable par rapport à la mesure de Lebesgue, on a*

$$\int_D f(x)dx \;=\; \int_\Delta f \circ h(x) |Dh(x)| dx. \tag{24}$$

Attention à la valeur absolue du jacobien! Cette formule s'appelle la formule du changement de variable, car elle revient à faire dans la seconde intégrale le changement de variable $x = (x_1, \dots, x_n) \mapsto y = (y_1, \dots, y_n) = h(x)$. Souvent $Dh(x)$ est noté

$$Dh(x) \;=\; \frac{D(y_1, \dots, y_n)}{D(x_1, \dots, x_n)}, \tag{25}$$

de sorte que (24) devient

$$\int_D f(y_1, \dots, y_n) dy_1 \dots dy_n \;=\; \int_\Delta f \circ h(x_1, \dots, x_n) \left| \frac{D(y_1, \dots, y_n)}{D(x_1, \dots, x_n)} \right| dx_1 \dots dx_n. \tag{26}$$

La notation (25), cohérente avec (23), permet de se rappeler que dans le changement de variable "l'élément différentiel" $dy_1 \dots dy_n$ est remplacé par $\left| \frac{D(y_1, \dots, y_n)}{D(x_1, \dots, x_n)} \right| dx_1 \dots dx_n$.

Exemples:

1) Supposons que $n = 1$, et que $D =]a, b[$ et $\Delta =]c, d[$ avec $a < b$ et $c < d$ (ces nombres peuvent être infinis). Un C^1-difféomorphisme est donc une application dérivable h ayant l'une des deux propriétés suivantes (h' est la dérivée de h):

   (i) on a $h'(x) > 0$ pour tout $x \in \Delta$, et $\lim_{x\downarrow c} h(x) = a$ et $\lim_{x\uparrow d} h(x) = b$, ou

   (ii) on a $h'(x) < 0$ pour tout $x \in \Delta$, et $\lim_{x\downarrow c} h(x) = b$ et $\lim_{x\uparrow d} h(x) = a$.

   (24) s'écrit alors:

$$\left.\begin{array}{lcl} h' > 0 \ \text{ sur } \ ]c, d[& \Rightarrow & \int_a^b f(x)dx = \int_c^d f \circ h(x) h'(x) dx \\ h' < 0 \ \text{ sur } \ ]c, d[& \Rightarrow & \int_a^b f(x)dx = -\int_c^d f \circ h(x) h'(x) dx \end{array}\right\}. \tag{27}$$

   et la seconde formule s'écrit aussi souvent $\int_a^b f(x)dx = \int_d^c f \circ h(x)h'(x)dx$, avec la convention $\int_d^c = -\int_c^d$: on retrouve donc la formule bien connue de changement de variable sur $\mathbb{R}$.

   Noter d'ailleurs que lorsque $n = 1$ la formule (31) ne se ramène pas toujours à (27): en effet, un ouvert D n'est pas forcément un intervalle ouvert. La forme générale de (24) lorsque $n = 1$ est en fait la suivante: soit $(]a_i, b_i])_{i\in I})$ et $(]c_i, d_i[)_{i\in I}$ deux familles d'intervalles ouverts respectivement deux-à-deux disjoints, avec I fini ou dénombrable. Pour chaque i soit h_i une bijection dérivable de $]c_i, d_i[$ dans $]a_i, b_i[$ dont la dérivée est toujours soit strictement positive, soit strictement négative. On a alors dès que $f1_D$ est intégrable, avec $D = \cup_i]a_i, b_i[$:

$$\int_D f(x)dx \ = \ \sum_{i\in I} \int_{c_i}^{d_i} f \circ h_i(x) |h_i'(x)| dx. \tag{28}$$

   Cette dernière formule est d'ailleurs vraie dès que les $]a_i, b_i[$ sont deux-à-deux disjoints (même si ce n'est pas le cas des $]c_i, d_i[$).

2) Soit f une fonction Lebesgue-intégrable sur $\mathbb{R}^n$, et $y \in \mathbb{R}^n$. On a alors

$$\int f(x)dx \ = \ \int f(x + y)dx. \tag{29}$$

   Il suffit d'appliquer (24) avec $D = \Delta = \mathbb{R}^n$ et $h(x) = x + y$: cette application est un C^1-difféomorphisme de $\mathbb{R}^n$ dans lui-même qui vérifie $Dh(x) = 1$ (sa matrice jacobienne est en fait la matrice identité)

Nous allons commencer par un lemme, dans lequel on fait les hypothèses du théorème 21.

Lemme 22 *Pour tout $x \in \Delta$ il existe une boule fermée B de centre x et de rayon $\varepsilon(x) > 0$, contenue dans Δ, telle que pour toute fonction borélienne positive f on ait, si C désigne l'image $\{h(x) : x \in B\}$ de B par h:*

$$\int_C f(y)dy = \int_B f \circ h(y) |Dh(y)| dy. \tag{30}$$

Preuve. La preuve se fait par récurrence sur la dimension n.

a) Soit $n = 1$ et $x \in \Delta$. Comme Δ est ouvert, il existe $\varepsilon(x) > 0$ tel que l'intervalle $B = [c,d] = [x - \varepsilon(x), x + \varepsilon(x)]$ soit contenu dans Δ. L'image C est un intervalle $[a,b]$, de sorte que (30) s'écrit en fait (27): cette formule, connue lorsque f est continue (pour l'intégrale de Riemann), doit être démontrée dans le cas où f est seulement borélienne positive.

Exactement comme dans l'exemple ci-dessus, deux cas sont possibles selon que la dérivée h' est positive ou négative sur $[c,d]$, et on va par exemple traiter le cas où $h'(y) < 0$ pour tout $y \in [c,d]$ (l'autre cas est un peu plus simple).

D'abord, par linéarité et limite croissante il suffit (comme on l'a déjà vu plusieurs fois) de montrer (34) lorsque $f = 1_A$ est l'indicatrice d'un borélien A. Mais si on pose $\mu(A) = \int_a^b 1_A(y)dy$ et $\nu(A) = -\int_c^d 1_A(h(y))h'(y)dy$, on définit clairement deux mesures finies μ et ν, de sorte qu'il nous faut montrer que ces deux mesures sont égales. D'après le théorème 8 il suffit donc de vérifier $\mu(A) = \nu(A)$ pour $A =]-\infty, \beta]$. Comme on a aussi de manière évidente $\mu(A) = \nu(A) = 0$ si $A \cap [a,b] = \emptyset$ il suffit de montrer que $\mu(]-\infty,\beta]) = \nu(]-\infty,\beta])$ pour $a \leq \beta \leq b$; comme dans ce cas il existe un unique point $\alpha \in B$ tel que $\beta = h(\alpha)$, on a alors $y \in [c,d], h(y) \leq \beta \Leftrightarrow \alpha \leq y \leq d$ et donc

$$\mu(A) \;=\; \beta - a, \qquad \nu(A) \;=\; -\int_\alpha^d h'(y)dy \;=\; h(\alpha) - h(d) \;=\; \beta - a$$

(par une propriété bien connue des intégrales de Riemann; ici h' est continue, donc Riemann-intégrable sur $[a,\alpha]$): on a donc le résultat.

b) Supposons (30) vraie pour $n-1$. Soit $x \in \Delta$. D'après (23) on a $Dh(x) \neq 0$, donc $\partial h_1/\partial x_i(x) \neq 0$ pour au moins un i. La numérotation des coordonnées n'ayant pas d'importance, on peut supposer que ceci est vrai pour $i = 1$. Soit θ l'application de Δ dans $I\!R^n$, continuement différentiable, définie de la manière suivante par ses coordonnées:

$$\theta_1(y) \;=\; h_1(y), \qquad \theta_j(y_1,\ldots,y_n) \;=\; y_j \quad \text{pour} \;\; j \geq 2.$$

Comme $\partial h_1/\partial x_1(x) \neq 0$, le théorème des fonctions implicites montre qu'il existe une boule fermée B de centre x et de rayon $\varepsilon(x) > 0$, contenue dans Δ, et une fonction continuement différentiable ρ de B dans $I\!R$, tels que

$$\theta_1(\rho(y), y_2, \ldots, y_n) \;=\; y_1 \qquad \forall y = (y_1,\ldots,y_n) \in B.$$

Notons C et F les images de la boule B par h et θ. On peut considérer aussi h (resp. θ) comme une application de B dans C (resp. dans F): la première est bijective par hypothèse, la seconde l'est également puisqu'elle admet clairement comme application réciproque $\theta^{-1}(y) = (\rho(y), y_2, \ldots, y_n)$, et on pose $\varphi = h \circ \theta^{-1}$ qui est bijective de F dans C et vérifie $\varphi_1(y_1,\ldots,y_n) = y_1$.

Introduisons quelques notations: si $y = (y_1,\ldots,y_n)$ on note $y' = (y_2,\ldots,y_n)$, de sorte qu'on peut écrire $y = (y_1, y')$. Soit $B'_{y'} = \{y_1 : (y_1,y') \in B\}$ et $B' = \{y' : B_{y'} \neq \emptyset\}$, et associons de même $F'_{y'}$ et F' à F. Remarquons que si $y \in B$ on a $\theta(y) = (h_1(y_1,y'), y')$, de sorte que $F' = B'$ et que $F'_{y'}$ est l'image de $B'_{y'}$ par l'application $y_1 \mapsto h_1(y_1,y')$. Par ailleurs par composition des dérivées et par $h = \varphi \circ \theta$ il vient $Dh(y) = D\theta(y)D\varphi(\theta(y))$, tandis que d'après la définition de θ on voit que $D\theta = \partial h_1/\partial x_1$. Par suite, en appliquant le théorème de Fubini, puis (30) pour $n = 1$, puis de nouveau le théorème de Fubini, on obtient:

$$\begin{aligned}
\int_B f \circ h(y)|Dh(y)|dy &= \int_{B'} dy' \int_{B'_{y'}} f \circ \varphi(h_1(t,y'),y') \left| D\varphi(h_1(t,y'),y') \tfrac{\partial}{\partial x_1} h_1(t,y') \right| dt \\
&= \int_{F'} dy' \int_{F'_{y'}} f \circ \varphi(z,y')|D\varphi(z,y')|dz \\
&= \int_F f \circ \varphi(y)|D\varphi(y)|dy
\end{aligned}$$

Maintenant, on note $F''_{y_1} = \{y' : (y_1, y) \in F\}$ et $F'' = \{y_1 : F''_{y_1} \neq \emptyset\}$, et on associe de même C''_{y_1} et C'' à C. Soit également $\varphi''_{y_1}(y') = (\varphi_i(y') : 2 \leq i \leq n)$. On a $\varphi_1(y_1, y') = y_1$, de sorte que la première ligne de la matrice jacobienne de φ est $(1, 0, \ldots, 0)$: par suite $D\varphi(t, y') = D\varphi''_t(y')$. Enfin $C'' = F''$ et $C''_t = F''_t$. Donc d'après le théorème de Fubini, puis (30) appliqué à $n-1$, puis de nouveau le théorème de Fubini, il vient

$$\begin{aligned}
\int_F f \circ \varphi(y)|D\varphi(y)|dy &= \int_{F''} dt \int_{F''_t} f(t, \varphi''_t(y'))|D\varphi''_t(y')|dy' \\
&= \int_{C''} dt \int_{C''_t} f(t,y')dy' \ = \ \int_C f(x)dx.
\end{aligned}$$

On a donc montré (30) pour n. $\square$

Preuve du théorème 21. Il suffit (par différence) de prouver (24) pour $f \geq 0$. A chaque $x \in \Delta$ on associe une boule B_x de centre x et de rayon strictement positif, tel que si C_x désigne l'image de B_x par h on ait (30). Cette égalité s'écrit aussi

$$\int_D f(y) 1_{C_x}(y)dy \ = \ \int_\Delta f \circ h(y) 1_{B_x}(y)|Dh(y)|dy.$$

Soit maintenant $(x(i) : i = 1, 2, \ldots)$ une énumération des points de Δ qui sont à coordonnées rationnelles (l'ensemble de ces points est dénombrable). Soit $A_1 = C_{x(1)}$ et, pour $i = 2, \ldots$, $A_i = C_{x(i)} \cap (\cup_{1 \leq j \leq i-1} A_j)^c$: les A_i forment une partition de Δ, et les images G_i de A_i par h forment une partition de D, avec $A_i \subset C_{x(i)}$ et $G_i \subset B_{x(i)}$. En appliquant l'égalité ci-dessus à $x = x(i)$ et à $f 1_{G_i}$, et comme $1_{G_i} \circ h = 1_{A_i}$, il vient

$$\int_D f(y) 1_{G_i}(y)dy \ = \ \int_\Delta f \circ h(y) 1_{A_i}(y)|Dh(y)|dy.$$

Il suffit de sommer sur i pour obtenir (24). $\square$

Corollaire 23 *Si μ est une mesure sur $I\!R^n$ admettant une densité f (par rapport à la mesure de Lebesgue), si h est un C^1-difféomorphisme de $I\!R^n$ dans lui-même, et si ν désigne la mesure image de μ sur $I\!R^n$ par l'application h, alors la mesure ν admet aussi une densité g, qui est donnée par la formule*

$$g(x) \ = \ f \circ h^{-1}(x)|Dh^{-1}(x)|; \tag{31}$$

5 Le produit de convolution

1) Dans ce paragraphe nous introduisons une "multiplication" des mesures sur $I\!R^d$, qui s'appelle le produit de convolution. Toutes les mesures dont on parle ci-dessous sont des mesures sur $I\!R^d$ muni de la tribu borélienne $\mathcal{R}^d$.

Définition 24 Si μ et ν sont deux mesures σ-finies sur $\mathbb{R}^d$, on appelle *produit de convolution* de μ et ν et on note $\mu \star \nu$ l'image de la mesure $\mu \otimes \nu$ par l'application de $\mathbb{R}^d \times \mathbb{R}^d$ dans $\mathbb{R}^d$ définie par $(x, y) \mapsto x + y$. $\square$

Ainsi, $\mu \star \nu$ est une mesure sur $\mathbb{R}^d$, qui d'après 2-(15) est donnée par

$$\mu \star \nu(A) \;=\; \int 1_A(x+y) d(\mu \otimes \nu)(x, y). \tag{32}$$

En utilisant le théorème de Fubini, on peut aussi écrire

$$\mu \star \nu(A) \;=\; \int \mu(dx) \int 1_A(x+y)\nu(dy) \;=\; \int \nu(dy) \int 1_A(x+y)\mu(dx). \tag{33}$$

On en déduit que le produit de convolution est *commutatif.* D'après le corollaire 19 il est aussi *associatif*, i.e. $(\mu \star \nu) \star \eta = \mu \star (\nu \star \eta)$, à condition bien entendu que les deux mesures $\mu \star \nu$ et $\nu \star \eta$ soient elles-mêmes σ-finies (ce qui n'est pas toujours vrai, comme l'exemple 3 ci-dessous le montre !).

Exemples.

1) Si $\mu = \varepsilon_0$ est la masse de Dirac en 0, on a $\mu \star \nu = \nu$ d'après (33): en d'autres termes, la masse de Dirac en 0 est un *élément neutre* pour le produit de convolution.

2) La masse totale de $\mu \star \nu$ est $\mu(\mathbb{R}^d)\nu(\mathbb{R}^d)$. En particulier, le produit de convolution de deux probabilités est encore une probabilité.

3) Si $\mu = \nu = \lambda_d$ est la mesure de Lebesgue sur $\mathbb{R}^d$, le produit $\eta = \mu \star \nu$ est la mesure donnée par $\eta(A) = 0$ si $\lambda_d(A) = 0$ et $\eta(A) = +\infty$ si $\lambda_d(A) > 0$: cela découle immédiatement de (33). Noter que cette mesure n'est pas σ-finie.

Proposition 25 *Si f est une fonction borélienne, positive ou intégrable par rapport au produit de convolution $\mu \star \nu$, on a*

$$\int f d(\mu \star \nu) \;=\; \int \mu(dx) \int f(x+y)\nu(dy) \;=\; \int \nu(dy) \int f(x+y)\mu(dx). \tag{34}$$

Preuve. Lorsque $f \geq 0$ cette formule se déduit de (33) selon le schéma habituel: par linéarité, puis limite croissante. Lorsque f est de signe quelconque et intégrable par rapport au produit de convolution, les formules (34) sont vraies pour f^+ et f^-, et donnent des valeurs finies, donc on a (34) pour f par différence. $\square$

2) Mesures signées avec densité. En vue de définir le produit de convolution d'une fonction et d'une mesure ou de deux fonctions, nous allons d'abord introduire le concept de "mesure signée", ce qui veut dire mesure non nécessairement positive. En vue d'éviter une théorie générale un peu lourde, nous nous contentons du cas des mesures admettant une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$.

Si f est une fonction borélienne positive sur $\mathbb{R}^d$, Lebesgue-intégrable, on sait qu'on peut définir la mesure $\mu = f \bullet \lambda_d$ de densité f par la formule $\mu(A) = \int_A f(x)dx$ (pour

$A \in \mathcal{R}^d$). Cette mesure est de masse totale $\mu(I\!R^d) = \int f(x)dx$ finie. Si maintenant f est Lebesgue-intégrable, mais de signe quelconque, on a les deux mesures $\mu^+ = f^+ \bullet \lambda_d$ et $\mu^- = f^- \bullet \lambda_d$. On pose alors

$$\mu \ = \ \mu^+ - \mu^- \qquad \text{(i.e. } \mu(A) \ = \ \mu^+(A) - \mu^-(A) \ \ \forall A \in I\!R^d), \tag{35}$$

et on note aussi $\mu = f \bullet \lambda_d$. (lL formule ci-dessus a bien un sens, puisque $\mu^+(A)$ et $\mu^-(A)$ sont finies). On dit que μ est une *mesure signée*, car elle vérifie $\mu(\emptyset) = 0$ et la σ-additivité, mais les nombres (finis) $\mu(A)$ sont a priori de signe quelconque. La théorie de l'intégration par rapport à de telles "mesures" est facile, et basée sur la formule $\int gd(f \bullet \lambda_d) = \int f(x)g(x)dx$ qu'on a vue dans la proposition 3-15 pour $f \geq 0$. Plus précisément, on pose la

Définition 26 Si f est une fonction borélienne sur $I\!R^d$, Lebesgue-intégrable, et si $\mu = f \bullet \lambda_d$, la fonction borélienne g est dite μ-intégrable si et seulement si la fonction fg est λ_d-intégrable. Dans ce cas on pose $\int gd\mu = \int f(x)g(x)dx$. $\square$

Notons aussi la propriété immédiate suivante: si $\mu = f \bullet \lambda_d$ et $\nu = g \bullet \lambda_d$ (avec f et g boréliennes Lebesgue-intégrables), la formule $\eta(A) = \mu(A) - \nu(A)$ définit une nouvelle mesure signée, qui n'est autre que $\eta = (f - g) \bullet \lambda_d$.

3) Avec cette définition, on a alors la proposition suivante. On rappelle que

Proposition 27 *Soit μ une mesure finie sur $I\!R^d$, et f une fonction borélienne sur $I\!R^d$, Lebesgue-intégrable. La formule*

$$(f \star \mu)(x) \ = \ \int f(x-y)\mu(dy) \tag{36}$$

définit λ_d-p.p. une fonction qui est Lebesgue-intégrable. Si $\nu = f \bullet \lambda_d$, cette fonction est la densité de la mesure signée $\mu \star \nu = \mu \star \nu^+ - \mu \star \mu^-$.

Preuve. En raisonnant sur f^+ et sur f^- et en faisant la différence, et compte tenu de la remarque suivant la définition 26, on voit qu'il suffit de montrer le résultat lorsque $f \geq 0$.

Dans ce cas, la fonction $f \star \mu$ est définie partout (à valeurs dans $\overline{I\!R}_+$). D'après le théorème de Fubini, elle est borélienne et vérifie

$$\int (f \star \mu)(x) 1_A(x)dx \ = \ \int \mu(dy) \int f(x-y)1_A(x)dx \ = \ \int \mu(dy) \int f(u)1_A(y+u)du$$

(on fait le changement de variable $x \mapsto h(x) = x - y$ et on applique (29) dans la dernière intégrale). Ceci vaut $\int \mu(dy) \int 1_A(y+u)\nu(du)$ par définition de ν, et une nouvelle application du théorème de Fubini entraine que

$$\int (f \star \mu)(x)1_A(x)dx \ = \ \int 1_A(y+u)d(\mu \otimes \nu)(y,u) \ = \ (\mu \star \nu)(A)$$

par (32). Donc $\mu \star \nu$ admet la densité $f \star \mu$. Enfin μ et ν étant deux mesures de masse totale finie, il en est de même de $\mu \star \nu$, donc $f \star \mu$ est Lebesgue-intégrable. $\square$

Enfin, si dans la proposition précédente la mesure μ admet elle aussi une densité, disons g, la fonction $f \star \mu$ est aussi notée $f \star g$, et elle vaut

$$f \star g(x) \;=\; \int f(x-y)g(y)dy \;=\; \int f(y)g(x-y)dy. \tag{37}$$

Il n'y a d'ailleurs pas de raison de supposer $g \geq 0$ ci-dessus: si g est de signe quelconque, cette formule définit la densité de la mesure signée $\mu \star \nu = \mu^+ \star \nu^+ + \mu^- \star \nu^- - \mu^+ \star \nu^- - \mu^- \star \nu^+$. Cela conduit à poser la définition suivante:

Définition 28 Si f et g sont deux fonctions boréliennes Lebesgue-intégrables sur $\mathbb{R}^d$, leur *produit de convolution* $f \star g$ est la fonction Lebesgue-intégrable définie par (37). □

Cette définition est un peu restrictive, et dans les livres d'analyse on voit parfois une définition plus générale du produit de convolution de deux fonctions: il suffit en fait que la formule (37) ait un sens.

En vertu de ce qui suit la définition 24, on voit que le produit de convolution des fonctions (Lebesgue-intégrables) est commutatif et associatif.

CHAPITRE 5

Les espaces L^p

1 Les définitions

Dans tout ce chapitre, l'espace mesuré $(E, \mathcal{E}, \mu)$ est fixé. Nous avons déjá rencontré l'espace $\mathcal{L}^1 = \mathcal{L}^1(E, \mathcal{E}, \mu)$ de toutes les fonctions mesurables sur $(E, \mathcal{E})$, à valeurs réelles, qui sont μ-intégrable (cf. chapitre 2). Mais, plus généralement, il existe toute une famille $\mathcal{L}^p$ d'espaces de fonctions mesurables, ainsi définis:

Définition 1 Si $p \in [1, \infty[$, on note $\mathcal{L}^p = \mathcal{L}^p(E, \mathcal{E}, \mu)$ l'ensemble de toutes les fonctions mesurables sur $(E, \mathcal{E})$, à valeurs réelles, telles que la fonction $|f|^p$ soit μ-intégrable. Si $f \in \mathcal{L}^p$, on pose

$$||f||_p \;=\; \left(\int |f|^p d\mu\right)^{1/p}. \tag{1}$$

Proposition 2 *Chaque espace $\mathcal{L}^p$ est un espace vectoriel.*

Preuve. D'abord, si $f \in \mathcal{L}^p$ et $a \in \mathbb{R}$, il est évident que le produit af appartient aussi à $\mathcal{L}^p$. Il nous suffit donc de montrer que si $f, g \in \mathcal{L}^p$, alors $f + g \in \mathcal{L}^p$.

On vérifie facilement que $(1+x)^p \leq 2^p(1+x^p)$ pour tout $x \geq 0$, donc aussi $(x+y)^p \leq 2^p(x^p + y^p)$ si $x, y \geq 0$. Il s'ensuit que $|f+g|^p \leq 2^p(|f|^p + |g|^p)$: si $f, g \in \mathcal{L}^p$, la fonction $|f+g|^p$ est intégrable et $f + g \in \mathcal{L}^p$. $\square$

Rappelons que si F désigne un espace vectoriel, on appelle *norme* sur F une application $u \mapsto ||u||$ de F dans $\mathbb{R}_+$ qui vérifie:

$$\left.\begin{array}{lll} \text{(i)} & ||u|| = 0 \;\Leftrightarrow\; u = 0, & \\ \text{(ii)} & a \in \mathbb{R},\; u \in F \;\Rightarrow\; ||au|| \;=\; |a|\;||u|| & \text{(homogénéité)}, \\ \text{(iii)} & ||u+v|| \;\leq\; ||u|| + ||v|| & \text{(inégalité triangulaire)}. \end{array}\right\} \tag{2}$$

Si on pose alors $d(u, v) = ||u - v||$, on définit une *distance* sur F, et la topologie associée est compatible avec la structure d'espace vectoriel, ce qui signifie que si $u_n \to u$ et $v_n \to v$

pour cette topologie (i.e. $d(u_n,u) \to 0$ et $d(v_n,v) \to 0$), et si $a_n \to a$ dans $I\!R$, alors $u_n + v_n \to u + v$ et $a_n u_n \to au$. On dit alors que F, ou plus précisément $(F, ||.||)$, est un *espace vectoriel normé*.

Revenons aux espaces $\mathcal{L}^p$. L'application $f \mapsto ||f||_p$ de $\mathcal{L}^p$ dans $I\!R_+$ vérifie clairement (ii) ci-dessus, ainsi que $||0||_p = 0$, et on verra plus tard que (iii) est aussi vérifié (c'est un résultat non évident, sauf pour $p = 1$). En revanche, $||f||_p = 0$ implique seulement que $f = 0$ μ-p.p., en vertu de 3-(13), de sorte que $||.||_p$ n'est en général pas une norme sur $\mathcal{L}^p$ (voir cependant l'exemple 2 ci-dessous).

Pour pallier ce problème, on opère ainsi: d'abord, si f et g sont deux fonctions réelles mesurables, on écrit $f \sim g$ si et seulement si $f = g$ μ-p.p., ce qui définit clairement une relation d'équivalence. En vertu du lemme 3-5, si $f \in \mathcal{L}^p$ et si $g \sim f$, on a aussi $g \in \mathcal{L}^p$ et $||g||_p = ||f||_p$. On peut donc poser la

Définition 3 Si $p \in [1, \infty[$, on note $L^p = L^p(E, \mathcal{E}, \mu)$ l'ensemble des classes d'équivalence des fonctions de $\mathcal{L}^p$, pour la relation d'équivalence "égalité μ-presque partout" rappelée ci-dessus. Si $f \in L^p$, on note $||f||_p$ la valeur commune des $||g||_p$ pour les fonctions g appartenant à la classe f.

Une autre manière d'exprimer cette définition consiste à dire que L^p est le *quotient* de $\mathcal{L}^p$ par la relation d'équivalence "égalité μ-presque partout". Si $f \in L^p$, on appelle *représentant* de f toute fonction mesurable $f' \in \mathcal{L}^p$ qui appartient à la classe d'équivalence f.

Soit alors $f, g \in L^p$ et $a \in I\!R$. Si f' et f'' (resp. g' et g'') sont deux représentants quelconques de f (resp. g), on a $f' + g' = f'' + g''$ μ-p.p. et $af' = af''$ μ-p.p.: on peut alors définir la *somme* $f + g$ (resp. le *produit* af) comme la classe d'équivalence de la somme $f' + g'$ (resp. du produit af') pour des représentants quelconques f' et g' de f et g: cela munit l'ensemble L^p d'une structure d'espace vectoriel, appelée structure quotient. En particulier l'élément nul (noté encore 0) de L^p est la classe d'équivalence de la fonction nulle, et une fonction mesurable f' est dans la classe 0 si et seulement si $f' = 0$ μ-p.p. D'après la proposition 3-11, on voit qu'on a alors l'équivalence:

$$||f||_p = 0 \quad \Leftrightarrow \quad f = 0 \qquad (\text{si } f \in L^p). \tag{3}$$

En d'autres termes, $||.||_p$ vérifie (2-(i)) sur l'espace L^p.

Les définitions de $\mathcal{L}^\infty$ et de L^∞ sont un peu plus délicates. L'idée est que $\mathcal{L}^\infty$ est l'ensemble des fonctions mesurables et "presque partout" bornées, proposition dont la traduction rigoureuse est la suivante:

Définition 4 a) On note $\mathcal{L}^\infty = \mathcal{L}^\infty(E, \mathcal{E}, \mu)$ l'ensemble de toutes les fonctions mesurables f sur $(E; \mathcal{E})$, à valeurs réelles, qui sont *essentiellement bornées*, ce qui signifie qu'il existe un réel $a \in I\!R_+$ (dépendant de f, bien entendu), tel que $|f| \leq a$ μ-p.p. Pour une telle fonction, on pose

$$||f||_\infty \;=\; \inf(a \in I\!R_+ : |f| \leq a \;\; \mu\text{-p.p.}). \tag{4}$$

b) On note $L^\infty = L^\infty(E, \mathcal{E}, \mu)$ l'ensemble des classes d'équivalence des fonctions de $\mathcal{L}^\infty$, pour la relation d'équivalence "égalité μ-presque partout": là encore, si $f \in \mathcal{L}^\infty$ et si

$g = f$ μ-p.p., alors $g \in \mathcal{L}^\infty$, et on a clairement $||f||_\infty = ||g||_\infty$, de sorte que si $h \in L^\infty$ on peut noter $||h||_\infty$ la valeur commune des $||f||_\infty$ lorsque f parcourt la classe h. $\square$

Remarquer que si $|f| \le A$ μ-p.p. et $|g| \le B$ μ-p.p. et si $a \in I\!R$, on a $|f+g| \le A+B$ μ-p.p. et $|af| \le |a|A$ μ-p.p.: on en déduit immédiatement que $\mathcal{L}^\infty$ est un espace vectoriel, et exactement comme ci-dessus on munit L^∞ de la structure vectorielle quotient induite par la relation d'équivalence "égalité μ-presque partout". La propriété (2-(i)) est alors satisfaite par $||.||_\infty$, sur l'espace L^∞.

Puisque $|f| \le a$ μ-p.p. pour tout $a > ||f||_\infty$, on a aussi la propriété suivante:

$$f \in \mathcal{L}^\infty \qquad \Rightarrow \qquad |f| \le ||f||_\infty \ \ \mu - \text{p.p.} \tag{5}$$

Dans toute la suite, on oubliera les $\mathcal{L}^p$ et on ne considèrera en fait que les L^p. Cependant, les éléments de L^p seront implicitement considérés comme des fonctions (ce qui revient en fait à confondre une classe d'équivalence avec l'un quelconque de ses représentants): cette identification d'une classe avec un représentant est en fait anodine, dans la mesure où les intégrales (par rapport à μ) sont les mêmes pour tous les représentants de la même classe. Attention, toutefois: lorsqu'on considère simultanément deux mesures μ et ν, les classes d'équivalence ne sont pas les mêmes relativement à chacune de ces mesures, et l'identification d'une classe à l'un quelconque de ses représentants ne peut plus se faire.

Exemples:

1) Si E est fini et si $\mu(E) < \infty$, tous les espaces L^p (resp. tous les espaces $\mathcal{L}^p$) pour $1 \le p \le +\infty$ sont les mêmes.

2) Si E est fini ou dénombrable, si $\mathcal{E} = \mathcal{P}(E)$, et si $\mu(\{x\}) > 0$ pour tout $x \in E$, alors $L^p = \mathcal{L}^p$ pour tout $p \in [1, \infty]$.

3) Soit $E = I\!N$ avec $\mathcal{E} = \mathcal{P}(E)$ et μ la mesure de comptage; on note ℓ^p l'espace $L^p(E, \mathcal{E}, \mu) = \mathcal{L}^p(E, \mathcal{E}, \mu)$. Cet espace est l'espace des suites $(u_n)_{n \in I\!N}$ telles que:

$$p \in [1, \infty[\qquad \Rightarrow \qquad \sum_n |u_n|^p < \infty \quad \text{et} \quad ||(u_n)||_p = \left(\sum_n |u_n|^p\right)^{1/p}, \tag{6}$$

$$p = \infty \qquad \Rightarrow \qquad \sup_n |u_n| < \infty \quad \text{et} \quad ||(u_n)||_\infty = \sup_n |u_n|. \tag{7}$$

Lemme 5 *Si μ est une mesure finie et si $1 \le p \le q \le +\infty$, on a $L^q \subset L^q$.*

Preuve. Si $q < \infty$, on a $|f|^p \le 1 + |f|^q$, donc

$$\int |f|^p d\mu \ \le \ \int (1 + |f|^q) d\mu \ = \ \mu(E) + \int |f|^q d\mu,$$

qui est fini si $f \in \mathrm{L}^q$. Si maintenant $f \in L^\infty$ et si $a = ||f||_\infty$, on a $|f|^p \le a^p$ (on peut négliger d'écrire "μ-p.p.", puisqu'on considère des classes d'équivalence). Donc $\int |f|^p d\mu \le a^p \mu(E) < \infty$. $\square$.

Remarque. Ce résultat est faux si $\mu(E) = \infty$: par exemple si $(E, \mathcal{E}, \mu) = (\mathbb{R}, \mathcal{R}, \lambda)$, la fonction $f(x) = 1$ est dans L^∞, mais pas dans L^p si $p < \infty$. La fonction $f(x) = x^{-a} 1_{[1,\infty[}(x)$ pour $a > 0$ est dans L^p si $p < 1/a$, mais pas si $p \geq 1/a$.

L'inclusion peut même être en sens inverse: en reprenant l'exemple 3 ci-dessus, on voit que $\ell^p \subset \ell^q$ si $p \leq q$.

2 Les espaces L^p pour $1 \leq p \leq \infty$

1) Nous allons commencer par une inégalté faisant intervenir les fonctions convexes, et dont nous déduirons ensuite deux inégalités sur les normes pour les espaces L^P.

Rappelons d'abord que si F est un espace vectoriel, une partie A de F est dite *convexe* si pour tous $x, y \in A$ on a $ax + (1-a)y \in A$ pour tout $a \in [0,1]$ (en d'autres termes, le "segment" de F d'extrémités x et y est tout entier contenu dans A). Ensuite, si I est un intervalle de $\mathbb{R}_+$ (borné ou non), une fonction ψ de I dans $\mathbb{R}$ est dite *concave* (resp. *convexe*) si l'ensemble $\{(x,y) \in \mathbb{R}^2 : x \in I, y \leq \psi(x)\}$ (resp. $\{(x,y) \in \mathbb{R}^2 : x \in I, y \geq \psi(x)\}$ est un ensemble convexe de $\mathbb{R}^2$. Remarquer que ψ est convexe si et seulement si $-\psi$ est concave. Noter aussi que si ψ est deux fois dérivable dans l'intérieur de I, elle est convexe (resp. concave) si et seulement si sa dérivée seconde est positive (resp. négative).

Lemme 6 (Inégalité de Jensen) *Soit ν une probabilité sur $(E, \mathcal{E})$, soit ψ une fonction concave sur un intervalle I de $\mathbb{R}$, soit enfin f une fonction réelle ν-intégrable, telle que $f(x) \in I$ pour tout $x \in E$. On a alors $\int f d\nu \in I$, et*

$$\int \psi(f) d\nu \;\leq\; \psi\left(\int f d\nu\right). \tag{8}$$

Preuve. Posons $m = \int f d\nu$. Soit a l'extrémité gauche de I. Si $a = -\infty$ on a $m > a$. Si $a > -\infty$ on a $f \geq a$ par hypothèse, donc $m \geq \int a d\nu = a$ puisque ν est une probabilité. De même si b est l'extrémité droite de I, on a $m < b$ si $b = \infty$, et $m \leq b$ si $b < \infty$: cela prouve que $m \in I$.

Comme ψ est concave, il existe au moins une droite de $\mathbb{R}^2$ d'équation $y = \alpha(x-m) + \psi(m)$ qui est située entièrement au dessus de graphe de ψ, i.e. $\alpha(x-m) + \psi(m) \geq \psi(x)$ pour tout $x \in I$. Par suite

$$\int \psi(f) d\nu \;\leq\; \int (\alpha(f-m) + \psi(m))\, d\nu \;=\; \alpha \int f d\nu - \alpha m + \psi(m) \;=\; \psi(m) \quad \square$$

Lemme 7 (Inégalité de Hölder) *Soit p, q, r des nombres de $[1, \infty]$ vérifiant $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$ (avec la convention $\frac{1}{\infty} = 0$). Si $f \in L^p$ et $g \in L^q$, le produit fg appartient à L^r, et on a*

$$||fg||_r \;\leq\; ||f||_p ||g||_q. \tag{9}$$

Preuve. Si $p = q = r = \infty$, ou si $p = r < \infty$ et $q = \infty$, le résultat est évident. On suppose donc que p, q, r sont finis. Comme les normes de f, g et fg ne font intervenir que les valeurs absolues de ces fonctions, on peut aussi supposer que f et g sont positives. Par ailleurs si $||f||_p = 0$ on a $f = 0$ μ-p.p., donc aussi $fg = 0$ μ-p.p., donc $||fg||_r = 0$. On peut donc enfin supposer que le nombre $C = \int f^p d\mu$ est strictement positif.

On pose alors $f' = f^p/C$, et on note $\nu = f' \bullet \mu$ la mesure qui admet la densité f' par rapport à μ. Noter que ν est une probabilité, et que $f > 0$ ν-p.p. (puisque $f' = 0$ sur l'ensemble $\{f = 0\}$, donc $\nu(f = 0) = \int f' 1_{\{f=0\}} d\mu = 0$). Etant donnés les rapports entre μ et ν, on a

$$\int f^r g^r d\mu \;=\; \int \frac{g^r}{f^{p-r}} f^p d\mu \;=\; C \int \frac{g^r}{f^{pr/q}} d\nu,$$

puisque $p - r = pr/q$. Comme $r < q$, la fonction $x \mapsto |x|^{q/r}$ est clairement convexe, et le lemme précédent entraine que

$$\int f^r g^r d\mu \;\leq\; C \left(\frac{1}{C} \int \frac{g^q}{f^p} f^p d\mu \right)^{r/q} \;=\; C^{r/p} \left(\int g^q d\mu \right)^{r/q}$$

(en utilisant que $1 - r/q = r/p$). Mais $\int g^q d\mu = ||g||_q^q$ et $C = ||f||_p^p$, de sorte que l'inégalité précédente est exactement (9). $\square$

Lemme 8 (Inégalité de Minkowski) *Soit $p \in [1, \infty]$, et f et g dans L^p. On a*

$$||f + g||_p \;\leq\; ||f||_p + ||g||_p. \tag{10}$$

Preuve. Si $p = 1$ le résultat est très simple: en effet, en identifiant (comme on l'a souligné ci-dessus) un élément f de L^p (i.e. une classe d'équivalence) avec l'un quelconque de ses représentants, on a

$$||f + g||_1 \;=\; \int |f + g| d\mu \;\leq\; \int (|f| + |g|) d\mu \;=\; \int |f| d\mu + \int |g| d\mu \;=\; ||f||_1 + ||g||_1.$$

Dans le cas $p = \infty$, on a $|f| \leq ||f||_\infty$ μ-p.p. et $|g| \leq ||g||_\infty$ μ-p.p., donc aussi $|f + g| \leq ||f||_\infty + ||g||_\infty$ μ-p.p., de sorte qu'on a (10).

Passons au cas où $1 < p < \infty$. Soit q le réel tel que $\frac{1}{p} + \frac{1}{q} = 1.$, et $h = |f + g|$. En utilisant d'abord que $h^p \leq (|f| + |g|) h^{p-1}$, puis l'inégalité (9) avec $r = 1$, on obtient:

$$\begin{aligned} \int |h|^p d\mu \;&\leq\; \int |f| h^{p-1} d\mu + \int |g| h^{p-1} d\mu \;\leq\; ||f||_p ||h^{p-1}||_q + ||g||_p ||h^{p-1}||_q \\ &=\; (||f||_p + ||g||_p) \left(\int h^{(p-1)q} d\mu \right)^{1/q}, \end{aligned}$$

ce qui donne finalement $\int h^p d\mu \leq (||f||_p + ||g||_p) \left(\int h^p d\mu\right)^{1/q}$, puisque $q(p-1) = p$. Comme on a déjà vu que L^p est un espace vectoriel, on a aussi $h \in L^p$, de sorte que $\int h^p d\mu < \infty$: on déduit alors de l'inégalité précédente que $\left(\int h^p d\mu\right)^{1-1/q} \leq ||f||_p + ||g||_p$. Comme $1 - 1/q = 1/p$, on en déduit le résultat. $\square$

2) Nous sommes maintenant prêt à démontrer les résultats principaux de ce paragraphe:

Théorème 9 *Si $p \in [1, \infty]$, l'espace $(L^p, ||.||_p)$ est un espace vectoriel normé.*

Preuve. Nous avons déjà vu que L^p est un espace vectoriel, et que sur cet espace l'application $f \mapsto ||f||_p$ vérifie (i) et (ii) de (2). La propriété (iii) de (2) n'est autre que (10).

Dans la suite, on dit qu'une suite $(f_n)_{n\geq 1}$ de L^p converge vers une limite f *dans* L^p, et on écrit $f_n \;\to^{L^p}\; f$, si $||f_n - f||_p \to 0$. Rappelons qu'on a

$$f_n \;\to^{L^p}\; f \quad \Rightarrow \quad ||f_n||_p \;\to\; ||f||_p. \tag{11}$$

(C'est en fait fait un résultat général sur la convergence associée à une norme, qui se démontre ainsi: on a $||u|| \leq ||u-v|| + ||v||$ par l'inégalité triangulaire, donc $||u|| - ||v|| \leq ||u-v||$ et on a de même $||v|| - ||u|| \leq ||u-v||$, de sorte que $|\; ||u|| - ||v|| \;| \leq ||u-v||$).

Signalons aussi les propriétés évidentes suivantes:

$$f \in L^p \quad \Leftrightarrow \quad |f| \in L^p, \quad \text{et alors} \quad ||\;|f|\;||_p = ||f||_p. \tag{12}$$

$$|f| \leq g \in L^p \quad \Rightarrow \quad f \in L^p \quad \text{et} \quad ||f||_p \leq ||g||_p. \tag{13}$$

Exemples:

1) Si E est un ensemble fini, avec la tribu de toutes ses parties, et si μ est une mesure telle que $0 < \mu(\{x\}) < \infty$ pour tout $x \in E$, on a déjà vu que $L^p = \mathcal{L}^p$ ne dépend pas de p, et il est clair que cet espace peut s'identifier à $\mathbb{R}^E$: une fonction est simplement une famille finie de réels $u = (u_x : x \in E)$. On a alors $||u||_p = \left(\sum_{x\in E} |u_x|^p \mu(\{x\})\right)^{1/p}$, et cette norme coïncide avec la norme euclidienne usuelle si $p = 2$ et si μ est la mesure de comptage. Sinon, c'est une norme différente, mais la topologie associée est la même dans tous les cas: c'est la topologie usuelle sur $\mathbb{R}^E$.

2) Si on considère l'espace ℓ^p décrit dans l'exemple 3 du paragraphe 1, la suite $(u^{(m)} = (u_n^{(m)} : n \in \mathbb{N}))_{m\geq 1}$ converge dans ℓ^p (i.e. pour la distance associée à la norme $||.||_p$) vers la limite (u_n) si et seulement si $\sum_n |u_n^{(m)} - u_n|^p \to 0$ quand $m \to \infty$, lorsque $p \in [1,\infty[$; si $p = \infty$, il y a convergence dans ℓ^∞ si et seulement si $\sup_n |u_n n^{(m)} - u_n| \to 0$. Ces conditions entrainent toutes que $u_n^{(m)} \to u_n$ pour tout n.

Le second résultat important concerne les rapports entre la convergence μ-presque partout d'une suite $(f_n)_{n\geq 1}$ de fonctions (qui est aussi, comme l'appartenance à $\mathcal{L}^p$, une propriété des classes d'équivalence), et la convergence dans L^p: pour étudier ces rapports, on supposera que $p \in [1,\infty[$, le cas $p = \infty$ étant de nature très différente. Supposons d'abord que $f_n \to f$ μ-p.p. (rappelons que cela veut dire que l'ensemble des $x \in E$ pour lesquels $f_n(x)$ ne converge pas vers $f(x)$ est μ-négligeable). On ne peut évidemment pas conclure que $f_n \;\to^{L^p}\; f$, ne serait-ce, par exemple, que parce que les fonctions f_n ou f n'appartiennent pas nécessairement à L^p. Cependant, on a:

$$p \in [1,\infty[, \quad f_n \to f \;\; \mu-\text{p.p.}, \quad |f_n| \leq g \in L^p \;\; \forall n \qquad \Rightarrow \qquad f_n \to^{L^p} f \tag{14}$$

(appliquer le théorème de convergence dominée de Lebesgue à la suite $|f_n - f|^p$, qui converge p.p. vers 0 et vérifie $|f_n - f|^p \leq (2g)^p$ μ-p.p.).

Dans le sens opposé, on a la

Proposition 10 *Soit $p \in [1,\infty[$. Si $f_n \to^{L^p} f$, il existe une suite $(n_k)_{k\geq 1}$ strictement croissante d'entiers telle que $f_{n_k} \to f$ μ-p.p.* (on dit aussi: on peut extraire de la suite (f_n) une sous-suite qui converge p.p. vers f).

Preuve. On pose $n_0 = 0$, et on définit par récurrence la suite n_k ainsi: si on connait n_{k-1} pour un $k \in \mathbb{N}^*$, on peut trouver un $n_k \in \mathbb{N}$ tel que $n_k > n_{k-1}$ et que $||f_{n_k} - f||_p \leq 2^{-k}$. Posons $A(k,q) = \{|f_{n_k} - f| > \frac{1}{q}\}$ (pour $q \in \mathbb{N}^*$). D'après l'inégalité de Bienaymé-Tchebicheff 3-(12) appliquée à la fonction $|f_{n_k} - f|^p$, on a $\mu(A(k,q)) \leq q^p \int |f_{n_k} - f|^p d\mu \leq q^p 2^{-pk}$. On a donc $\sum_{k\geq 1} \mu(A(k,q)) < \infty$, et le lemme de Borel-Cantelli (corollaire 3-10) implique que l'ensemble $B(q) = \limsup_k A(k,q)$ est μ-négligeable pour tout q. Il en est donc de même de $B = \cup_{q\geq 1} B(q)$.

Soit $x \notin B$. Pour tout $q \geq 1$ on a $x \notin B(q)$, ce qui veut dire qu'il y a (au plus) un nombre fini d'entiers k tels que $x \in A(k,q)$. Notons $K(x,q)$ le plus grand des entiers k tels que $x \in A(k,q)$. Pour tout $k > K(x,q)$ on a alors $|f_{n_k}(x) - f(x)| \leq \frac{1}{q}$: comme q est arbitrairement grand, cela veut exactement dire que $f_{n_k}(x) \to f(x)$. On a donc montré que $f_{n_k}(x) \to f(x)$ si $x \notin B$, et le résultat est démontré. $\square$

Lorsque $p = \infty$, on a un résultat bien plus fort: si $f_n \to^{L^\infty} f$, alors en dehors d'un ensemble négligeable on a que la suite $(f_n)_{n\geq 1}$ converge uniformément vers f.

Corollaire 11 *Soit $p \in [1,\infty]$. Si la suite $(f_n)_{n\geq 1}$ converge dans L^p vers une limite f, et μ-p.p. vers une limite g, on a $f = g$ μ-p.p.*

Preuve. Le résultat découle immédiatement de la remarque précédant l'énoncé, lorsque $p = \infty$. Si maintenant $p \in [1,\infty[$, on a vu plus haut qu'il existe une suite (n_k) telle que $f_{n_k} \to f$ μ-p.p., et comme $f_n \to g$ μ-p.p. on a a fortiori $f_{n_k} \to g$ μ-p.p.: la propriété $f = g$ μ-p.p. est alors évidente. $\square$

Remarques: 1) On ne peut pas faire mieux que la proposition 10. Soit par exemple $E = [0,1[$, muni de la tribu borélienne $\mathcal{E}$ et de la mesure de Lebesgue λ. Soit $u_n = \sum_{i=1}^n \frac{1}{i}$. On note A_n l'ensemble des $x \in E$ qui sont de la forme $x = y + p$, avec $p \in \mathbb{Z}$ et $u_n \leq y \leq u_{n+1}$ (c'est à dire l'ensemble des points de $[u_n, u_{n+1}]$ "modulo 1"). Soit aussi $f_n = 1_{A_n}$. On a $\int f_n d\lambda = \frac{1}{n+1}$, de sorte que $f_n \to^{L^p} 0$ pour tout $p \in [1,\infty[$. Cependant, comme $u_n \uparrow \infty$, on voit que les ensembles A_n "glissent" le long de E une infinité de fois, de sorte que $\limsup_n f_n = 1$ et $\liminf_n f_n = 0$: on n'a donc pas $f_n \to 0$ λ-p.p.

2) A l'inverse, si on a $f_n \to f$ μ-p.p. et si les fonctions f_n et f sont dans L^p, il n'est pas sûr que $f_n \to^{L^p} f$: Sur le même espace que dans la remarque précédente, soit $f_n(x) = n1_{[0,1/n]}(x)$. La suite f_n converge p.p. vers $f = 0$, mais $\int f_n^p d\lambda = n^{p-1}$ ne tend pas vers 0 (bien-sûr, l'hypothèse de (14) n'est pas satisfaite dans cette situation). $\square$

Proposition 12 *Soit $p \in [1,\infty]$ et $(f_n)_{n\geq 1}$ des fonctions de L^p telles que $\sum_{n\geq 1} ||f_n||_p < \infty$. La série $\sum_n f_n$ est alors presque partout absolument convergente, et convergente dans L^p, et on a*

$$||\sum_n f_n||_p \ \leq \ \sum_n ||f_n||_p. \tag{15}$$

Voici quelques commentaires sur la signification de cet énoncé. D'abord, dire que la série $\sum_n f_n$ est p.p. absolument convergente signifie que pour tout x en dehors d'un ensemble négligeable N on a $\sum_n |f_n(x)| < \infty$, donc la série numérique $\sum_n f_n(x)$ converge pour ces valeurs de x. La convergence dans L^p signifie que les fonctions $g_n = \sum_{i=1}^n f_i$ convergent dans L^p vers une limite g. En vertu du corollaire 11, on a donc $g(x) = \sum_n f_n(x)$ pour tout x en dehors d'un ensemble négligeable, et il est alors naturel de noter $\sum_n f_n$ la fonction g.

Preuve. Posons comme ci-dessus $g_n = \sum_{i=1}^n f_i$, et aussi $h_n = \sum_{i=1}^n |f_i|$ et $h = \lim_n \uparrow h_n$. Supposons d'abord $p = \infty$. Il existe un ensemble négligeable N tel que si $x \in N^c$ on a $|f_n(x)| \leq ||f_n||_\infty$. Donc si $x \in N^c$ on a $h(x) \leq \sum_n ||f_n||_\infty < \infty$, donc la série $\sum_n f_n(x)$ est absolument convergente et sa somme $g(x)$ vérifie $|g(x) - g_n(x)| \leq \sum_{m>n} ||f_m||_\infty$: toutes les assertions sont alors évidentes.

Supposons ensuite $p < \infty$. D'après l'inégalité triangulaire et (12) on a $||h_n||_p \leq \sum_{i=1}^n ||f_i||_p \leq a$, si a désigne la somme $a = \sum_n ||f_n||_p$, qui est finie par hypothèse. D'après le théorème de limite monotone, on a

$$\int h^p d\mu \;=\; \lim_n \uparrow \int h_n^p d\mu \;=\; \lim_n \uparrow ||h_n||_p^p \;\leq\; a^p.$$

On en déduit que h^p, étant μ-intégrable, est μ-p.p. finie, et il en est évidemment de même de h. En d'autres termes la série numérique $\sum_n f_n(x)$ est absolument convergente, et a fortiori convergente, sur l'ensemble $\{x : h(x) < \infty\}$ dont le complémentaire est négligeable.

Posons $g(x) = \sum_n f_n(x)$ pour tout point x tel que la série soit absolument convergente, et (de manière arbitraire) $g(x) = 0$ ailleurs. On a bien-sûr $|g| \leq h$, donc $\int |g|^p d\mu \leq \int h^p d\mu \leq a^p$ d'après ce qui précde: on en déduit que $g \in L^p$ et qu'on a (15).

Il reste à montrer que $g_n \to^{L^p} g$. Si $h(x) < \infty$, on a $g(x) - g_n(x) = \sum_{i=n+1}^\infty f_i$, de sorte qu'en appliquant (15) à la série commençant à l'indice $n+1$ (au lieu de 1), on obtient $||g - g_n||_p \leq \sum_{i=n+1}^\infty ||f_i||_p$. Cette dernière quantité est le reste d'une série numérique convergente, donc tend vers 0: cela achève la démonstration. $\square$

Passons enfin au troisième et dernier résultat important. Rappelons qu'un espace métrique est *complet* si *toute suite de Cauchy converge*: cela signifie que, avec d désignant la distance, toute suite (x_n) de points vérifiant $d(x_n, x_m) \to 0$ lorsque n et m tendent vers l'infini est convergente (inversement, une suite convergente est toujours une suite de Cauchy, que l'espace soit complet ou non). Un espace vectoriel normé complet est appelé *espace de Banach*.

Théorème 13 *Si $p \in [1, \infty]$, l'espace $(L^p, ||.||_p)$ est un espace de Banach.*

Compte tenu du théorème 9, il suffit d'appliquer la proposition 12 et le lemme général suivant:

Lemme 14 *Soit F un espace vectoriel normé, de norme $||.||$. Si toute série $\sum_n u_n$ vérifiant $\sum_n ||u_n|| < \infty$ converge dans F (i.e., les sommes partielles $v_n = \sum_{i \leq n} u_i$ vérifient $||v_n - v|| \to 0$ pour un certain $v \in F$), alors F est un espace de Banach.*

Preuve. Soit $(u_n)_{n\geq 1}$ une suite de Cauchy. Pour tout $k \in \mathbb{N}$ on note p_k le plus petit entier tel que $||u_n - u_m|| \leq 2^{-k}$ pour tous $n, m \geq p_k$: d'après la définition des suites de Cauchy, p_k existe, et on a évidemment $p_k \leq p_{k+1}$.

Posons alors $w_0 = u_{p_0}$ et $w_k = u_{p_k} - u_{p_{k-1}}$ pour $k \geq 1$. On a $||w_0|| < \infty$, et $||w_k|| \leq 2^{-(k-1)}$ pour $k \geq 1$ par définition de p_{k-1} et le fait que $p_k \geq p_{k-1}$. Par suite $\sum_{k\geq 0} ||w_k|| < \infty$, et l'hypothèse implique que $u_{p_k} = \sum_{i=0}^k w_i$ converge (en norme) vers une limite w.

Enfin, on a

$$n \geq p_k \quad \Rightarrow \quad ||u_n - w|| \ \leq \ ||u_n - u_{p_k}|| + ||u_{p_k} - w|| \ \leq \ 2^{-k} + ||u_{p_k} - w||.$$

Comme $||u_{p_k} - w|| \to 0$ quand $k \to \infty$, on en déduit que $||u_n - w|| \to 0$ quand $n \to \infty$, d'où le résultat. $\square$

3 L'espace L^2 et les espaces de Hilbert

3-1) Soit H un espace vectoriel (réel). Un *produit scalaire* est une application de $H \times H$ dans $\mathbb{R}$, notée $(u, v) \mapsto \langle u, v\rangle$, qui vérifie

$$\left.\begin{array}{lll} \text{(i)} & \langle u,u\rangle \geq 0 \ \ \text{(positivité)}, \\ \text{(ii)} & \langle u,v\rangle = \langle v,v\rangle \ \ \text{(symétrie)}, \\ \text{(iii)} & u \mapsto \langle u,v\rangle \ \ \text{est linéaire.} \end{array}\right\} \tag{16}$$

On dit aussi que $\langle .,.\rangle$ est une *forme bilinéaire symétrique positive*. Elle est dite *strictement positive* si au lieu de (i) on a

$$\text{(i')} \quad u \neq 0 \quad \Rightarrow \quad \langle u,u\rangle \ > \ 0. \tag{17}$$

Lorsque on a (17), on dit que l'espace H muni du produit scalaire $\langle .,.\rangle$ est un *espace pré-hilbertien.*

Lemme 15 *a) Si $\langle .,.\rangle$ est un produit scalaire, l'application $u \mapsto ||u|| = \langle u,u\rangle^{1/2}$ vérifie (ii) et (iii) de (2), et on a l'inégalité de Schwarz: $|\langle u,v\rangle| \ \leq \ ||u|| \ ||v||$.*

b) Si de plus on a (17), l'application $u \mapsto ||u||$ est une norme.

Preuve. (16) implique que pour tout $x \in \mathbb{R}$:

$$0 \ \leq \ \langle u + xv, u + xv\rangle \ = \ x^2||v||^2 + 2x\langle u,v\rangle + ||u||^2.$$

Le membre de droite est un trinôme du second degré qui est toujours positif, donc son discriminant $\langle u,v\rangle^2 - ||u||^2||v||^2$ est négatif ou nul: on en déduit l'inégalité de Schwarz. En particulier si $x = 1$ on obtient

$$||u+v||^2 \ = \ ||v||^2 + 2\langle u,v\rangle + ||u||^2 \ \leq \ ||v||^2 + 2||u|| \ ||v|| + ||v||^2 \ = \ (||u|| + ||v||)^2,$$

de sorte que $||.||$ vérifie l'inégalité triangulaire. L'homogénéité de $||.||$ est évidente, ainsi que la condition (i) de (2) lorsqu'on a (17). $\square$

Définition 16 Un *espace de Hilbert* est un espace vectoriel muni d'un produit scalaire vérifiant (17), et qui muni de la norme associée comme ci-dessus est un espace complet.

Exemple: L'espace $\mathbb{R}^d$ muni du produit scalaire usuel (qui au couple $x = (x_i)_{1\leq i\leq d}, y = (y_i)_{1\leq i\leq d}$ associe $\langle x,y\rangle = \sum_{i=1}^d x_i y_i$), est un espace de Hilbert. La norme associée est la norme euclidienne usuelle.

3-2) Nous en venons maintenant à un théorème très important:

Théorème 17 *L'espace $L^2 = L^2(E,\mathcal{E},\mu)$ est un espace de Hilbert pour le produit scalaire*

$$\langle f,g\rangle \;=\; \int (fg)d\mu, \tag{18}$$

et la norme associée est la norme $||.||_2$. En outre, on a l'inégalité de Cauchy-Schwarz:

$$||fg||_1 \;\leq\; ||f||_2||g||_2. \tag{19}$$

Preuve. Comme $|fg| \leq f^2 + g^2$, on voit en premier lieu que si $f,g \in L^2$ alors $fg \in L^1$, de sorte que la formule (18) a un sens. Il est immédiat (à cause de la linéarité et de la positivité de l'intégrale) que $\langle .,.\rangle$ vérifie (16), et aussi que $\langle f,f\rangle = ||f||_2^2$. On a donc (17), grâce à (3). On a vu au théorème 13 que $(L^2, ||.||_2)$ est complet, donc c'est un espace de Hilbert. Enfin (19) n'est autre que l'inégalité de Schwarz appliquée aux fonctions $|f|$ et $|g|$, pour le produit scalaire ci-dessus (c'est aussi un cas particulier de l'inégalité de Hölder). $\square$

Lorsque $f_n \to^{L^2} f$ on dit aussi que f_n converge vers f *en moyenne quadratique.*

Corollaire 18 *a) Si $f_n \to^{L^2} f$ et $g_n \to^{L^2} g$, on a $f_n g_n \to^{L^1} fg$.*

b) Si μ est une mesure finie, on a $L^2 \subset L^1$ et l'injection canonique de L^2 dans L^1 est continue, et on a

$$f \in L^2 \quad\Rightarrow\quad ||f||_1 \leq \sqrt{\mu(E)}||f||_2. \tag{20}$$

Preuve. a) On a $f_n g_n - fg = (f_n - f)g + f(g_n - g) + (f_n - f)(g_n - g)$, donc

$$\begin{aligned} ||f_n g_n - fg||_1 &\leq ||(f_n - f)g||_1 + ||f(g_n - g)||_1 + ||(f_n - f)(g_n - g)||_1 \\ &\leq ||f_n - f||_2||g||_2 + ||f||_2||g_n - g||_2 + ||f_n - f||_2||g_n - g||_2 \end{aligned}$$

en utilisant (19). On déduit alors $||f_n g_n - fg||_2 \to 0$ des hypothèses.

b) On a déjà vu l'inclusion $L^2 \subset L^1$ (lemme 5), et la continuité de l'injection canonique découle de (20), qui elle-même résulte de (19) appliquée à f et à $g = 1$. $\square$

3-3) Géométrie des espaces de Hilbert. Dans ce sous-paragraphe, on considère un espace de Hilbert H, muni du produit scalaire $\langle .,.\rangle$ et de la norme associée $||.||$. Nous allons donner quelques éléments sur la "géométrie" de H: il faut bien-sûr penser à l'exemple fondamental d'espace de Hilbert $H = \mathbb{R}^d$ donné après la définition 19: les principales propriétés de la géométrie euclidienne se transposent aux espaces de Hilbert sans modification.

Un élément de H sera appelé souvent un “vecteur”. Rappelons que $u_n \to u$ (sous-entendu: dans H) si $||u_n - u|| \to 0$; rappelons aussi (cf. après (11)) que si $u_n \to u$ on a $||u_n|| \to ||u||$, c'est à dire que l'application $u \mapsto ||u||$ de H dans $I\!R_+$ est continue. Plus généralement l'application $(u,v) \mapsto \langle u,v\rangle$ de $H \times H$ dans $I\!R$ est aussi continue: si $u_n \to u$ et $v_n \to v$, on a $\langle u_n, v_n\rangle \to \langle u,v\rangle$ (cela se démontre exactement comme la partie (a) du corollaire 18).

Commençons par la notion d'orthogonalité:

Définition 19 Deux vecteurs u et v de H sont dits *orthogonaux* si $\langle u,v\rangle = 0$ (on écrit aussi $u \perp v$). Si K est une partie de H on appelle *orthogonal* de K, et on note $K^\perp$, l'ensemble des vecteurs $u \in H$ qui sont orthogonaux à tous les vecteurs de K. Deux parties K et L de H sont dites *orthogonales* si $K \subset L^\perp$ ($\Leftrightarrow \quad L \subset K^\perp$). $\square$

Le résultat suivant est très intuitif en dimension finie (faire, par exemple, un dessin dans le cas de la dimension 2).

Proposition 20 *a) L'orthogonal $K^\perp$ de toute partie K de H est un sous-espace vectoriel fermé de H, et est donc lui-même un espace de Hilbert* (fermé signifie que la limite d'une suite quelconque de vecteurs de $K^\perp$ appartient aussi à $K^\perp$).

b) (Théorème de projection) *Si K est une partie convexe fermée de H* (cf. avant le lemme 6 pour la définition de la convexité), *et si $u \in H$, il existe un vecteur et un seul, noté $\Pi_K u$ de K et appelé* projection orthogonale *de u sur K, qui minimise l'application $v \mapsto ||v-u||$ sur K. On a $\Pi_K u = u$ si $u \in K$.*

Preuve. a) Pour tous $u, v \in K^\perp$ et $a \in I\!R$ on a $\langle au, w\rangle = a\langle u,w\rangle = 0$ et $\langle u+v, w\rangle = \langle u,w\rangle + \langle v,w\rangle = 0$ si $w \in K$: par suite au et $u+v$ sont dans $K^\perp$, qui est donc un espace vectoriel. Si $u_n \to u$ et $u_n \in K^\perp$ et $w \in K$ on a $\langle u,w\rangle = \lim_n \langle u_n, w\rangle = 0$: donc u appartient à $K^\perp$, qui est donc fermé. Enfin la restriction du produit scalaire à $K^\perp$ est encore un produit scalaire, et si $(u_n)_{n\geq 1}$ est une suite de Cauchy dans $K^\perp$, c'est aussi une suite de Cauchy dans H, donc elle converge vers une limite u qui appartient à $K^\perp$ d'après ce qui précède: cela prouve que $K^\perp$ est aussi un espace de Hilbert.

b) Soit $a = \inf_{v\in K} ||v-u||$. Il existe une suite $(v_n)_{n\geq 1}$ dans K telle que $||v_n - u|| \to a$. Montrons que cette suite est de Cauchy. En effet, il est facile de voir à partir de (16) et de $||w||^2 = \langle w,w\rangle$ que $||w+w'||^2 + ||w-w'||^2 = 2||w||^2 + 2||w'||^2$. Donc

$$||v_n + v_m - 2u||^2 + ||v_n - v_m||^2 \ = \ 2||v_n - u||^2 + ||v_m - u||^2.$$

Par ailleurs la convexité de K implique $\frac{1}{2}(v_n + v_m) \in K$, donc $||v_n + v_m - 2u||^2 = 4||\frac{1}{2}(v_n + v_m) - u||^2 \geq 4a^2$, et il vient

$$||v_n - v_m||^2 \ \leq \ 2||v_n - u||^2 + 2||v_m - u||^2 - 4a^2.$$

Comme $||v_n - u||^2 \to a^2$ on en déduit que $||v_n - v_m||^2 \to 0$ lorsque n et m tendent vers ∞: la suite (v_n) est donc de Cauchy, de sorte qu'elle converge vers une limite v qui vérifie $||v-u|| = \lim_n ||v_n - u|| = a$, et qui appartient à K puisque K est fermé.

Il reste à montrer l'unicité de v. Si $v' \in K$ vérifie également $||v' - u|| = a$, posons $v'_{2n} = v$ et $v'_{2n+1} = v'$. On a $||v'_n - u|| = a$ pour tout n, donc d'après ce qui précède la suite (v'_n) est une suite de Cauchy, qui converge; comme elle admet les deux points limite v et v', il faut donc que $v' = v$. Enfin si $u \in K$, il est clair que $v = u$ minimise $v \mapsto ||v - u||$ sur K. $\square$

Proposition 21 *Soit K un sous-espace vectoriel fermé de H.*

a) $\Pi_K u$ est l'unique vecteur v de K tel que $u - v \in K^\perp$.

b) Π_K est une application linéaire continue, contractant la norme (i.e. $||\Pi_K u|| \leq ||u||$). Son image est K et son noyau est $K^\perp$, et on l'appelle l'opérateur projection (orthogonale) sur K.

c) Tout vecteur u de H se décompose de manière unique en une somme $u = v + w$ avec $v \in K$ et $w \in K^\perp$, et on a $v = \Pi_K u$ et $w = \Pi_{K^\perp} u$ (donc les sous-espaces K et $K^\perp$ sont supplémentaires dans H).

d) On a $(K^\perp)^\perp = K$.

Preuve. a) Soit $v = \Pi_K u$. Pour tout $w \in K$ et tout $x \in I\!R$ on a $v + xw \in K$, donc

$$||v + xw - u||^2 \;=\; ||v - u||^2 + 2x\langle w, v - u\rangle + x^2||w||^2 \;\geq\; ||v - u||^2$$

pour tout $x \in I\!R$, ce qui n'est possible que si $\langle w, v - u\rangle = 0$: cela montre que $v - u \in K^\perp$. Si $v' \in K$ vérifie aussi $v' - u \in K^\perp$, le vecteur $v - v'$ est à la fois dans K et dans $K^\perp$; étant orthogonal à lui-même, il est nul (par (17)).

b) Le fait que Π_K soit une application linéaire découle immédiatement de la caractérisation (a). Il est clair que l'image de H par Π_K est contenue dans K, et comme $\Pi_K u = u$ si $u \in K$, elle est exactement K. D'après (a) on a $\Pi_K u = 0$ si et seulement si $u \in K^\perp$, donc cet ensemble est le noyau de Π_K. Enfin, toujurs d'après (a), on a $u = v + w$ avec $v = \Pi_K u$ et $w \perp v$, de sorte que $||u||^2 = ||v||^2 + ||w||^2$ et $||\Pi_K u||^2 \leq ||u||^2$: ainsi, Π_K est une contraction, et est donc en particulier continue.

c) On a vu ci-dessus que $u = v + w$ avec $v = \Pi_K u$ et $w \in K^\perp$. Comme $K^\perp$ est aussi un sous-espace vectoriel fermé, et comme $u - w \in K$ et que tout vecteur de K est orthogonal à $K^\perp$ (propriété évidente), la caractérisation (a) pour $\Pi_{K^\perp}$ implique que $w = \Pi_{K^\perp} u$. Si $u = v' + w'$ est une autre décomposition avec $v' \in K$ et $w' \in K^\perp$, par différence $v - v' = w' - w$ est dans $K \cap K^\perp$, et on a déjà vu que cela implique $v - v' = 0$: on a donc achevé de prouver (c).

(d) On a déjà vu que $K \subset (K^\perp)^\perp$, et l'inclusion inverse découle de (c). $\square$

Soit K une partie de H. L'espace vectoriel *engendré* par K, et noté $e(K)$, est le plus petit espace vectoriel contenant K (il existe, car d'une part $K \subset H$, d'autre part une intersection quelconque d'espaces vectoriels est un espace vectoriel). Noter que, de manière évidente, $e(K)$ est ausi l'ensemble des combinaisons linéaires finies de vecteurs de K.

La fermeture de $e(K)$ (i.e. l'ensemble des limites des suites convergentes de vecteurs de $e(K)$) est encore clairement un espace vectoriel, appelé *l'espace vectoriel fermé engendré par K*. Enfin, on dit que K *est total* dans H si l'espace vectoriel fermé engendré par K égale H.

Corollaire 22 *Une partie K de H est totale si et seulement si $K^\perp = \{0\}$.*

Preuve. Soit H' l'espace vectoriel fermé engendré par K. Il est évident que $H'^\perp \subset K^\perp$. Si $u \in K^\perp$, alors u est aussi orthogonal à tous les éléments de $e(K)$ (utiliser (16)-(iii)); si alors $v \in H'$ il existe des $v_n \in e(K)$ avec $v_n \to v$, et comme $\langle u, v_n\rangle = 0$ pour tout n on a aussi $\langle u, v\rangle = 0$ et par suite $u \in H'^\perp$: on a donc $H'^\perp = K^\perp$. Comme $H' = H$ équivaut à $H'^\perp = \{0\}$ par (c) de la proposition 21, on a le résultat. □

Le second sujet important est celui de la dualité. Rappelons que si $(F, ||.||)$ est un espace vetoriel normé, son *dual* est l'ensemble F' des applications linéaires $\psi : F \mapsto I\!R$ telles que $|\psi(u)| \leq C||u||$ pour tout $u \in F$, pour une certaine constante C (cette dernière propriété est en fait équivalente à la continuité de ψ). Il est clair que F' est un espace vectoriel, qu'on munit d'une norme $||.||'$ définie ainsi:

$$||\psi||' \;=\; \sup(|\psi(u)| : u \in F, ||u|| \leq 1) \;=\; \sup(\frac{|\psi(u)|}{||u||} : u \in F, u \neq 0). \tag{21}$$

Lorsque $(F, ||.||)$ est un espace de Banach, on peut montrer qu'il en est de même de $(F', ||.||')$.

Théorème 23 *Soit H un espace de Hilbert. On peut identifier le dual $(H', ||.||')$ avec $(H, ||.||)$, en associant à tout $v \in H$ l'application linéaire ψ_v définie par $\psi_v(u) = \langle u, v\rangle$.*

Preuve. Si $v \in H$ l'application ψ_v définie ci-dessus est linéaire continue et vérifie $||\psi_v||' \leq ||v||$ d'après l'inégalité de Schwarz. Comme $\psi_v(v) = \langle v, v\rangle = ||v||^2$, (21)) implique $||\psi_v||' = ||v||$. Remarquer aussi que si $\psi_v = \psi_{v'}$, le vecteur $v - v'$ est orthogonal à tout $u \in H$, donc orthogonal en particulier à lui-même, de sorte que $v = v'$.

Il reste à montrer qu'inversement, si $\psi \in H'$ il existe un $v \in H$ tel que $\psi = \psi_v$. Si $\psi = 0$, $v = 0$ répond à la question. Supposons donc que $\psi \neq 0$. Le noyau K de ψ est un sous-espace vectoriel de H, fermé à cause de la continuité de ψ, et $K^\perp$ n'est pas réduit à $\{0\}$ (sinon on aurait $K = H$ d'après le corollaire 22, donc $\psi = 0$). Soit alors $w \in K^\perp$, $w \neq 0$, de sorte que $\psi(w) \neq 0$. Posons $v = \frac{\psi(w)}{||w||}w$.

Pour tout $u \in H$ on pose $u' = u - \frac{\psi(u)}{\psi(w)}w$. On a $\psi(u') = 0$, donc $u' \in K$, donc $\langle u', v\rangle = 0$ et

$$\langle u', v\rangle \;=\; \langle u, v\rangle - \frac{\psi(u)}{\psi(w)}\langle w, v\rangle \;=\; \langle u, v\rangle - \psi(u)$$

est donc nul: par suite $\psi(u) = \langle u, v\rangle = \psi_v(u)$. □

Le troisième sujet important est celui des bases orthonormales. Commençons par une définition:

Définition 24 Un *système orthonormal* est une famille $(u_i)_{i \in I}$ de vecteurs de l'espace de Hilbert H qui vérifie $\langle u_i, u_j\rangle = 0$ si $i \neq j$ et $\langle u_i, u_i\rangle = 1$. Une *base orthonormale* est un système orthonormal total dans H. □

D'après le corollaire 22, un système orthonormal $(u_i)_{i\in I}$ est une base si et seulement si

$$\langle v, u_i\rangle = 0 \quad \forall i\in I \qquad \Rightarrow \qquad v=0. \tag{22}$$

Attention: une base orthonormale n'est pas une base "algébrique", au sens où tout vecteur serait une combinaison linéaire finie de vecteurs de la base, sauf bien-sûr si H est de dimension finie.

Soit $(u_i)_{1\leq i\leq d}$ un système orthonormal fini, et K l'espace vectoriel fermé qu'il engendre. K contient évidemment l'ensemble des combinaisons linéaires finies $u=\sum_{i=1}^d a_iu_i$ ($a_i\in I\!\!R$) et, comme ce dernier ensemble est à l'évidence fermé il est en fait égal à K. Noter que si $u=\sum_{i=1}^d a_iu_i$ et $v=\sum_{i=1}^d b_iu_i$, alors

$$\langle u,v\rangle \;=\; \sum_{1\leq i\leq d, 1\leq j\leq d} a_ib_j\langle u_i,u_j\rangle \;=\; \sum_{i=1}^d a_ib_i.$$

Ainsi, K peut être identifié à l'espace $I\!\!R^d$ muni de la norme euclidienne, par la correspondance $u\leftrightarrow (a_i)_{1\leq i\leq d}$. Cela se généralise:

Proposition 25 *Soit $(u_n)_{n\in I\!\!N}$ un système orthonormal dénombrable, et K l'espace vectoriel fermé engendré par ce système.*

a) K est isomorphe, en tant qu'espace de Hilbert, à l'espace ℓ^2 des suites réelles $a=(a_n)_{n\in I\!\!N}$ telles que $\sum_n (a_n)^2<\infty$. Plus précisément si $a=(a_n)$ est dans ℓ^2, la série $\sum_n a_nu_n$ converge dans H et définit un vecteur $u(a)$ de K; l'application $a\mapsto u(a)$ est linéaire bijective de ℓ^2 dans K et préserve le produit scalaire (donc la norme, donc elle est continue ainsi que son inverse):

$$\langle \sum_n a_nu_n, \sum_n b_nu_n\rangle \;=\; \sum_n a_nb_n. \tag{23}$$

b) Si $u\in H$ et $a_n=\langle u,u_n\rangle$, alors $a=(a_n)_{n\in I\!\!N}$ appartient à ℓ^2 et on a $\sum_n a_nu_n = \Pi_K u$, et en particulier

$$\sum_n \langle u, u_n^2\rangle \;\leq\; ||u||^2, \tag{24}$$

avec égalité si et seulement si $u\in K$.

Commençons par un lemme, qui a un intérêt propre:

Lemme 26 *Si $(v_n)_{n\in I\!\!N}$ est une suite de vecteurs deux-à-deux orthogonaux, la série $\sum_n v_n$ converge dans H si et seulement si $\sum_n ||v_n||^2<\infty$, et on a alors*

$$||\sum_n v_n||^2 \;=\; \sum_n ||v_n||^2. \tag{25}$$

Preuve. Soit $w_n=\sum_{i=0}^n v_i$ et $S_n=\sum_{i=0}^n ||v_i||^2$. Si $n<m$ on a

$$||w_m-w_n||^2 \;=\; \langle \sum_{i=n+1}^m v_i, \sum_{i=n+1}^m v_i\rangle \;=\; \sum_{n<i,j\leq m}\langle v_i,v_j\rangle \;=\; \sum_{i=n+1}^m ||v_i||^2 \;=\; S_m-S_n$$

puisque $\langle v_i, v_j\rangle = 0$ si $i \neq j$. La suite (w_m) converge dans H si et seulement si elle est de Cauchy, donc d'après ce qui précède si et seulement si la suite $(S_n)_n$ est de Cauchy dans $I\!R$, donc si et seulement si $\sum_i ||v_i||^2 < \infty$. Enfin sous ces conditions, on note w la limite de la suite (w_n); exactement comme ci-dessus on a $||w_n||^2 = S_n$, et en passant à la limite on obtient (25). $\square$

Preuve de la proposition 25. a) Soit $a = (a_n) \in \ell^2$. Comme $||a_n u_n|| = a_n$, le lemme 26 entraine que la série $\sum_n a_n u_n$ converge, et on note $u(a)$ sa somme. Il est clair que $u(a) \in K$, et que $a \mapsto u(a)$ est linéaire. (25) implique $||u(a)|| = ||a||_2$ (on note $||.||_2$ et $\langle .,.\rangle_2$ la norme et le produit scalaire de ℓ^2). On a $\langle u,v\rangle = \frac{1}{2}\left(||u+v||^2 + ||u-v||^2\right)$, et une relation analogue entre $||.||_2$ et $\langle .,.\rangle_2$: donc l'application linéaire $a \mapsto u(a)$, qui préserve la norme, préserve aussi le produit scalaire, et on a (23). Enfin, l'image K' de ℓ^2 est un espace vectoriel contenant les u_n et contenu dans K; si $v_n \in K'$ et $v_n \to v$, alors (v_n) est une suite de Cauchy dans H, donc les inverses $u^{-1}(v_n)$ forment une suite de Cauchy dans ℓ^2, convergeant donc vers une limite a, et évidemment $v = u(a)$: ainsi K' est fermé, donc $K' = K$ et $u(.)$ est bijective de ℓ^2 dans K.

b) Soit $u \in H$ et $v = \Pi_K u$. Il existe $a = (a_n) \in \ell^2$ avec $v = \sum_n a_n u_n$ et $||a||_2 = ||v||$. Si $v_n = \sum_{i=0}^n a_i u_i$, on a $\langle v_n, u_m\rangle = a_m$ si $n \geq m$, et comme $v_n \to v$ on en déduit que $a_m = \langle v, u_m\rangle$ pour tout m. Pour terminer il suffit de remarquer que $||u||^2 = ||v||^2 + ||u-v||^2$ ("théorème de Pythagore"), donc $||u|| \geq ||v||$, avec égalité si et seulement si $u = v$, donc si et seulement si $u \in K$. $\square$

Revenons pour terminer à l'espace $L^2 = L^2(E, \mathcal{E}, \mu)$. On peut énoncer le théorème 23 dans ce cadre, ce qui donne:

Théorème 27 *L'espace L^2 est son propre dual, ce qui revient à dire qu'à toute application linéaire continue ψ de L^2 dans $I\!R$ on peut associer une fonction $g \in L^2$ telle que $\psi(f) = \int fgd\mu$ pour toute $f \in L^2$.*

On a aussi le théorème suivant, que nous énonçons sans démonstration:

Théorème 28 *Si μ une mesure σ-finie sur $(E, \mathcal{E}) = (I\!R^d, \mathcal{R}^d)$, l'espace L^2 admet une base orthonormale dénombrable.*

Un exemple de base orthonormale: Supposons que $E = [0,1]$ soit muni de la tribu borélienne $\mathcal{E}$ et de la mesure de Lebesgue λ. La suite de fonctions ci-dessous constitue une base orthonormale de L^2, appelée la *base de Haar*:

$$f_n(x) = \begin{cases} 1 & \text{si } k2^{-n} \leq x < (k+1)2^{-n} \quad \text{pour un } k \text{ impair} \\ -1 & \text{si } k2^{-n} \leq x < (k+1)2^{-n} \quad \text{pour un } k \text{ pair.} \end{cases}$$

4 Le théorème de Radon-Nikodym

Nous commençons ce paragraphe par quelques compléments sur les mesures avec densité par rapport à une mesure donnée. L'espace $(E, \mathcal{E})$ est fixé. Rappelons que si μ est une mesure sur $(E, \mathcal{E})$ et si f et f' sont deux fonctions mesurables à valeurs dans $[0, \infty]$, les deux mesures $f \bullet \mu$ et $f' \bullet \mu$ sont égales dès que $f = f'$ μ-p.p. Ce qui suit est une série de variations sur la réciproque de ce résultat.

Lemme 29 *Si μ est une mesure σ-finie et si f est une fonction mesurable à valeurs dans $[0, \infty]$, la mesure $\nu = f \bullet \mu$ est σ-finie si et seulement si f est μ-presque partout finie* (on a alors aussi $\nu = f' \bullet \mu$ avec la fonction finie $f' = f 1_{\{f<\infty\}}$).

Preuve. Si ν est σ-finie il existe une suite $(E_n)_{n\geq 1}$ d'ensembles mesurables croissant vers E et avec $\nu(E_n) = \int (f 1_{E_n}) d\mu < \infty$. Par le corollaire 9 on a $f 1_{E_n} < \infty$ μ-p.p., et comme $E_n \uparrow E$ on en déduit que $f < \infty$ μ-p.p.

Supposons inversement que $f < \infty$ μ-p.p. On pose $F_0 = \{f = \infty\}$ et, pour $n \geq 1$, $F_n = F_0 \cup \{f \leq n\}$. Les F_n sont mesurables et croissent vers E. Par hypothèse il existe aussi une suite $(G_n)_{n \in \mathbb{N}}$ d'ensembles mesurables croissant vers E et tels que $\mu(G_n) < \infty$. La suite $E_n = F_n \cap G_n$ croît vers E, et

$$\nu(E_n) \;=\; \int (f 1_{F_0 \cap G_n}) d\mu + \int (f 1_{\{f \leq n\} \cap G_n}) d\mu \;\leq\; 0 + n\mu(G_n) \;<\; \infty$$

puisque $\mu(F_0) = 0$: donc ν est σ-finie. $\square$

Lemme 30 *Soit μ une mesure sur $(E, \mathcal{E})$ et f et f' deux fonctions mesurables.*

a) Si les fonctions f et f' sont positives, et si les mesures $f \bullet \mu$ et $f' \bullet \mu$ sont égales et σ-finies, on a $f = f'$ μ-p.p.

b) Si les fonctions f et f' sont μ-intégrables et si $\int_A f d\mu = \int_A f' d\mu$ pour tout A dans une classe $\mathcal{C}$ de parties mesurables qui est stable par intersection ($A, B \in \mathcal{C} \;\Rightarrow\; A \cap B \in \mathcal{C}$), qui contient une suite $(E_n)_{n \geq 1}$ croissant vers E, et qui engendre la tribu $\mathcal{E}$. Alors $f = f'$ μ-p.p.

c) Si $\int_A f d\mu \geq 0$ pour tout $A \in \mathcal{E}$, on a $f \geq 0$ μ-p.p.

Preuve. a) Soit $\nu = f \bullet \mu = f' \bullet \mu$, et $(E_n)_{n\geq 1}$ une suite d'ensembles mesurables croissant vers E avec $\nu(E_n) < \infty$. Si $A = \{f < f'\}$ on a $\int_{A \cap E_n} f d\mu = \int_{A \cap E_n} f' d\mu < \infty$, de sorte que $\int (f' - f) 1_{A \cap E_n} d\mu = 0$ et comme l'intégrand est positif ou nul on déduit de la proposition 3-11 que $(f' - f) 1_{A \cap E_n} = 0$ μ-p.p. On en déduit que $f' \leq f$ μ-p.p. sur chaque E_n, donc aussi sur E. On montre de même que $f \leq f'$ μ-p.p., donc finalement $f = f'$ μ-p.p.

b) Posons $\nu_+ = f^+ \bullet \mu$, $\nu_- = f^- \bullet \mu$, $\nu'_+ = f'^+ \bullet \mu$ et $\nu'_- = f'^- \bullet \mu$. Ces quatre mesures sont finies (car f et f' sont intégrables), et l'hypothèse implique que $\nu_+(A) + \nu'_-(A) = \nu_-(A) + \nu'_+(A)$ pour tout $A \in \mathcal{C}$: Le théorème 4-1 entraine alors que $\nu_+ + \nu'_- = \nu_- + \nu'_+$, et (a) implique $f^+ + f'^- = f^- + f'^+$ μ-p.p., donc aussi $f = f'$ μ-p.p.

c) Si $A = \{f < 0\}$ on a $\int (f 1_A) d\mu \geq 0$ et $f 1_A \geq 0$, ce qui implique $f 1_A = 0$ μ-p.p.: par suite $f \geq 0$ μ-p.p. $\square$

Remarque: Le résultat (a) ci-dessus est en défaut sans l'hypothèse de σ-finitude. Si par exemple $\mu(A) = \infty$ si $A \neq \emptyset$ et $\mu(\emptyset) = 0$ la mesure $f \bullet \mu$ égale μ lorsque $f > 0$ partout.

Nous allons maintenant utiliser le théorème 27 pour montrer un résultat très utile dans les applications. L'espace mesurable $(E, \mathcal{E})$ est toujours fixé.

Définition 31 Soit μ et ν deux mesures sur $(E,\mathcal{E})$. On dit que ν est absolument continue par rapport à μ si tout ensemble μ-négligeable est aussi ν-négligeable.

Théorème 32 *Soit μ et ν deux mesures σ-finies sur $(E,\mathcal{E})$. La mesure ν est absolument continue par rapport à μ si et seulement si elle est de la forme $\nu = f \bullet \mu$ pour une fonction mesurable f à valeurs dans $\mathbb{R}_+$.*

Ce théorème est appelé THEOREME DE RADON-NIKODYM. La condition suffisante est évidente: si en effet $A \in \mathcal{E}$ vérifie $\mu(A) = 0$, la fonction $f1_A$ est μ-presque partout nulle, et comme $\nu(A) = \int (f1_A)d\mu$ on a aussi $\nu(A) = 0$. Pour la réciproque, nous commençons par un lemme:

Lemme 33 *Si μ et ν sont deux mesures σ-finies telles que $\nu(A) \leq \mu(A)$ pour tout A, il existe une fonction f mesurable, à valeurs dans $[0,1]$, telle que $\nu = f \bullet \mu$.*

Preuve. a) Supposons d'abord que μ soit une mesure finie. On note $L^2 = L^2(E,\mathcal{E},\mu)$, avec sa norme $||.||_2$. Remarquons que si g est mesurable positive, on a $\int g d\nu \leq \int g d\mu$ (c'est vrai par hypothèse pour les indicatrices, donc par linéarité pour les fonctions étagées, donc par limite monotone pour les fonctions mesurables positives). Si donc $g \in L^2$, on a $\int |g| d\nu \leq \int |g| d\mu \leq \sqrt{\mu(E)}||g||_2$ (appliquer (20)). Par suite $\psi(g) = \int g d\nu$ est une application, clairement linéaire, de L^2 dans $\mathbb{R}$, et $|\psi(g)| \leq \sqrt{\mu(E)}||g||_2$: par suite ψ est un élément du dual de L^2, et d'après le théorème 27 il existe $f \in L^2$ tel que $\int g d\nu = \psi(g) = \int f g d\mu$: en particulier $\nu(A) = \int_A f d\mu$ pour $A \in \mathcal{E}$; d'après le lemme 30 on peut choisir $f \geq 0$ et on a $\nu = f \bullet \mu$. Enfin $\int_A (1-f)d\mu = \mu(A) - \nu(A) \geq 0$ pour tout $A \in \mathcal{E}$, et le lemme 30-(c) entraine $f \leq 1$ μ-p.p., de sorte qu'on peut choisir f à valeurs dans $[0,1]$.

b) Passons au cas général. Il existe une partition mesurable $(E_n)_{n\geq 1}$ de E telle que $\mu(E_n) < \infty$ pour tout n. Notons μ_n et ν_n les restrictions de μ et ν à E_n (rappelons par exemple que $\mu_n(A) = \mu(A \cap E_n)$). On a évidemment $\nu_n(A) \leq \mu_n(A)$ pour tout A, donc (a) implique que $\nu_n = f_n \bullet \mu_n$ pour une fonction f_n à valeurs dans $[0,1]$: il reste à poser $f = \sum_n f_n 1_{E_n}$ pour obtenir le résultat. $\Box$

Preuve du théorème 32. Soit $\eta = \mu + \nu$, qui est aussi une mesure σ-finie. On a $\mu(A) \leq \eta(A)$ et $\nu(A) \leq \eta(A)$ pour tout $A \in \mathcal{E}$, donc il existe deux fonctions g et h à valeurs dans $[0,1]$ telles que $\mu = g\bullet\eta$ et $\nu = h\bullet\eta$, en vertu du lemme ci-dessus. Nous allons montrer que la fonction f qui vaut h/g sur l'ensemble $B = \{g > 0\}$ et 0 sur B^c répond à la question.

D'abord, $\mu(B^c) = \int g1_{B^c}d\eta = 0$, puisque $g1_{B^c} = 0$, donc $\nu(B^c) = 0$ puisque ν est absolument continue par rapport à μ. Donc si $A \in \mathcal{E}$, la proposition 3-15 implique:

$$\nu(A) \;=\; \nu(A \cap B^c) \;=\; \int (h1_{A\cap B^c})d\eta \;=\; \int (fg1_A)d\eta \;=\; \int (f1_A)d\mu,$$

et le résultat s'ensuit. $\Box$

5 La dualité des espaces L^p

La question du dual de L^2 a été réglée au théorème 27, et ici nous allons décrire celui de L^p pour les autres valeurs finies de p. Encore une fois, l'espace mesuré $(E,\mathcal{E},\mu)$ est fixé.

Si $p,q \in [1,\infty]$ vérifient $\frac{1}{p}+\frac{1}{q}=1$, et si $g \in L^q$, en vertu de l'inégalité de Hölder on peut poser pour $f \in L^p$:

$$\psi_g(f) \;=\; \int (fg)d\mu, \tag{26}$$

ce qui définit une application linéaire continue sur L^p, donc un élément du dual $(L^p)'$ dont la norme vérifie $||\psi_g||'_p \leq ||g||_q$. En fait, on a bien mieux, du moins si $p<\infty$:

Théorème 34 *Soit $p \in [1,\infty[$ et $q \in]1,\infty]$ tels que $\frac{1}{p}+\frac{1}{q}=1$, et supposons μ σ-finie. On peut identifier le dual de $(L^p,||.||_p)$ à l'espace $(L^q,||.||_q)$, en associant à toute $g \in L^q$ l'application ψ_g définie par (26) (et en particulier on a $||\psi_g||'_p = ||g||_q$).*

Preuve. a) Comme μ est σ-finie, il existe une partition mesurable $(E_n)_{n\geq 1}$ de E telle que $a_n=\mu(E_n)<\infty$. La fonction $h=\sum_n \frac{1}{n^2(1+a_n)}1_{E_n}$ est mesurable strictement positive, et $\int h^p d\mu = \sum_n \frac{1}{n^{2p}(1+a_n)^p}\mu(E_n) \leq \sum_{n\geq 1}\frac{1}{n^{2p}}<\infty$. Donc la mesure $\eta = h^p \bullet \mu$ est une mesure finie.

b) Soit maintenant ψ un élément du dual de L^p, de norme $||\psi||'_p = a$. Comme $h \in L^p$, on a a fortiori $h1_A \in L^p$ pour $A \in \mathcal{E}$, donc $\psi(h1_A)$ est bien définie, et il vient

$$|\psi(h1_A)| \;\leq\; a||h1_A||_p \;=\; a\left(\int h^p 1_A d\mu\right)^{1/p} \;=\; a\eta(A)^{1/p}. \tag{27}$$

Pour tout $A \in \mathcal{E}$ on note $\mathcal{J}_A$ la classe des partitions finies $\mathcal{E}$-mesurables de A. Si $\mathcal{A}=(A_i)_{1\leq i\leq n} \in \mathcal{J}_A$, on pose

$$\gamma_+(A,\mathcal{A}) \;=\; \sum_{i=1}^n \psi(h1_{A_i})^+, \qquad \gamma_-(A,\mathcal{A}) \;=\; \sum_{i=1}^n \psi(h1_{A_i})^-$$

$$\nu_+(A) \;=\; \sup(\gamma_+(A,\mathcal{A}) : \mathcal{A} \in \mathcal{J}_A), \qquad \nu_-(A) \;=\; \sup(\gamma_-(A,\mathcal{A}) : \mathcal{A} \in \mathcal{J}_A).$$

Si $\varepsilon_i = 1$ lorsque $\psi(h1_{A_i})>0$ et $\varepsilon_i=0$ sinon, on a aussi $\gamma_+(A,\mathcal{A}) = \sum_{i=1}^n \varepsilon_i \psi(h1_{A_i}) = \psi(\sum_{i=1}^n (h\varepsilon_i 1_{A_i}))$, donc $\gamma_+(A,\mathcal{A}) \leq a||h\sum_{i=1}^n \varepsilon_i 1_{A_i}||_p \leq a||h1_A||_p$, donc

$$\nu_+(A) \;\leq\; a\eta(A)^{1/p}, \tag{28}$$

et de même pour ν_-. Enfin, on a $\gamma_+(A,\mathcal{A}) - \gamma_-(A,\mathcal{A}) = \sum_{i=1}^n \psi(h1_{A_i}) = \psi(h1_A)$, donc $\gamma_+(A,\mathcal{A}) = \gamma_-(A,\mathcal{A}) + \psi(h1_A)$ et on en déduit

$$\psi(A) \;=\; \nu_+(A) - \nu_-(A). \tag{29}$$

c) Montrons maintenant que ν_+ est une mesure (nécessairement finie à cause de (28)). D'abord $\nu_+(\emptyset)=0$ est évident. Ensuite, soit B,C deux ensembles mesurables disjoints; la réunion d'une partition dans $\mathcal{J}_B$ et d'une partition dans $\mathcal{J}_C$ étant une partition dans $\mathcal{J}_{B\cup C}$, on a clairement $\nu_+(B\cup C) \geq \nu_+(B)+\nu_+(C)$. A l'inverse, si $\mathcal{A}=(A_i)_{1\leq i\leq n} \in \mathcal{J}_{B\cup C}$, les $(B_i = A_i\cap B)_{1\leq i\leq n}$ et $(C_i = A_i \cap C)_{1\leq i\leq n}$ sont dans $\mathcal{J}_B$ et $\mathcal{J}_C$ respectivement. Comme $(x+y)^+ \leq x^+ + y^+$, il vient:

$$\gamma_+(B\cup C,\mathcal{A}) = \sum_{i=1}^n (\psi(h1_{B_i})+\psi(h1_{C_i}))^+ \leq \sum_{i=1}^n (\psi(h1_{B_i})^+ + \psi(h1_{C_i})^+) \leq \nu_+(B)+\nu_-(C)$$

et donc $\nu_+(B \cup C) \leq \nu_+(B) + \nu_+(C)$: on en déduit que ν_+ est additive.

Pour montrer la σ-additivité, soit $(B_n)_{n\geq 1}$ une suite d'ensembles mesurables deux-à-deux disjoints. On pose $C_n = \cup_{i=1}^n B_i$, qui croît vers $C = \cup_n B_n$, et soit $C'_n = C \backslash C_n$. Par additivité, $\nu_+(C_n) = \sum_{i=1}^n \nu_+(B_i)$ et $\nu_+(C) = \nu_+(C_n) + \nu_+(C'_n)$. Mais $\eta(C'_n) \to 0$ parce que η est une mesure finie, donc (28) implique que $\nu_+(C'_n) \to 0$ (c'est ici qu'intervient l'hypothèse $p < \infty$): on a donc $\nu_+(C) = \sum_n \nu_+(B_n)$, et ν_+ est une mesure. On vérifierait de même que ν_- est une mesure.

d) D'après (28) les mesures finies ν_+ et ν_- sont absolument continues par rapport à η, donc aussi par rapport à μ. D'après le théorème 32 il existe des fonctions ℓ_+ et ℓ_-, μ-intégrables et à valeurs dans $\mathbb{R}_+$, telles que $\nu_+ = \ell_+ \bullet \mu$ et $\nu_- = \ell_- \bullet \mu$. On pose $g = \frac{1}{h}(\ell_+ - \ell_-)$, et on va montrer que $g \in L^q$, que $||g||_q \leq a$ et que $\psi = \psi_g$: comme on a vu avant l'énoncé du théorème que $||\psi_g||'_p \leq ||g||_q$, on en déduira que $||\psi_g||'_p = ||g||_q$, et la preuve sera achevée.

e) (29) montre que $\psi(h1_A) = \int (\ell_+ - \ell_-)1_A d\mu = \int gh1_A d\mu$. En d'autres termes, on a

$$\psi(f) \ = \ \int gf d\mu \tag{30}$$

pour toute fonction f de la forme $f = h1_A$. Par linéarité, on a (30) pour f de la forme $f = hk$ avec k finie étagée: noter que dans ce cas on a $|gf| \leq K(\ell_+ + \ell_-)$ pour une certaine constante K, tandis que ℓ_+ et ℓ_- sont μ-intégrables, donc $\int fg d\mu$ existe et est fini. Supposons maintenant k mesurable avec $|k| \leq K$ pour une constante K. En considérant les parties positive et négative de k, on voit qu'il existe une suite k_n de fonctions étagées mesurables, avec $|k_n| \leq K$, qui converge simplement vers k; d'une part $|hk_n| \leq Kh \in L^p$ et $hk_n \to hk$ simplement, donc $hk_n \to^{L^p} hk$ par (14), donc $\psi(hk_n) \to \psi(hk)$; d'autre part $|ghk_n| \leq K|gh|$ qui est μ-intgrable et $ghk_n \to ghk$ simplement, donc $\int ghk_n d\mu \to \int ghk d\mu$ par le théorème de Lebesgue. (30) étant vraie pour chaque hk_n, elle est vraie aussi pour hk: on a donc montré (30) pour toute fonction mesurable $f = hk$ avec k bornée.

Supposons $p = 1$, donc $q = \infty$, et soit $b > a$. Soit $k = 1_{\{g \geq b\}} - 1_{\{g \leq -b\}}$. (30) implique $\psi(hk) = \int |g| h 1_{\{|g| \geq b\}} d\mu \geq b \int h 1_{\{|g| \geq b\}} d\mu$; on a aussi $||hk||_1 = \int h 1_{\{|g| \geq b\}} d\mu$, et comme $|\psi(hk)| \leq a ||hk||_1$ on arrive à une contradiction, sauf si $\mu(\{|g| \geq b\}) = 0$: par suite on a $|g| \leq b$ μ-p.p. pour tout $b > a$, ce qui entraine que $g \in L^\infty$ et $||g||_\infty \leq a$.

Supposons $p > 1$, donc $q < \infty$. Soit f_n la fonction de même signe que g, et dont la valeur absolue vaut $|g|^{q-1} 1_{\{|g| \leq nh\}}$. f_n/h étant bornée, (30) implique $\psi(f_n) = \int g f_n d\mu = \int |g|^q 1_{\{|g| \leq nh\}} d\mu$; par ailleurs $\int |f_n|^p d\mu = \int |g|^q 1_{\{|g| \leq nh\}} d\mu = \psi(f_n)$ puisque $p(q-1) = q$. Comme $|\psi(f_n)| \leq a ||f_n||_p$ on en déduit que $|\psi(f_n)| \leq a |\psi(f_n)|^{1/p}$, d'où $|\psi(f_n)| \leq a^q$. En d'autres termes, $\int |f_n|^p d\mu = \int |g|^q 1_{\{|g \leq nh\}} d\mu \leq a^q$. Comme $\{|g| \leq nh\}$ croît vers E (car $h > 0$), le théorème de limite monotone entraîne que $\int |g|^q d\mu \leq a^q$: par suite $g \in L^q$, et $||g||_q \leq a$.

On a donc montré dans tous les cas que $g \in L^q$ et que $||g||_q \leq a$, tandis que (30) implique $\psi(f) = \psi_g(f)$ si f est mesurable et f/h est bornée. Soit enfin $f \in L^p$, et $f_n = f 1_{\{|f| \leq nh\}}$. On a $f_n \to f$ simplement et $|f_n| \leq |f|$, donc d'après (14) on a $f_n \to^{L^p} f$, par suite $\psi(f_n) \to \psi(f)$ et $\psi_g(f_n) \to \psi_g(f)$. Comme $\psi(f_n) = \psi_g(f_n)$ d'après ce qui précède, on en déduit que $\psi(f) = \psi_g(f)$, et la preuve est enfin achevée. $\square$

Remarque: Le résultat est faux pour $p = \infty$: on a vu que L^1 peut être identifié à une partie de $(L^\infty)'$, via (26), mais ce dernier espace est strictement plus grand que L^1. La description du dual de L^∞ est complexe et dépasse les objectifs de ce cours.

CHAPITRE 6

La transformée de Fourier

1 Définition et propriétés élémentaires

Dans (presque) tout ce chapitre l'espace de base est $\mathbb{R}^d$, muni de la tribu borélienne $\mathcal{R}^d$. On note encore λ_d la mesure de Lebesgue sur $\mathbb{R}^d$, et on rappelle que l'intégrale (quand elle existe) d'une fonction borélienne f sur $\mathbb{R}^d$ est notée $\int f d\lambda_d = \int f(x_1, \ldots, x_d)dx_1 \ldots dx_d = \int f(x)dx$. Rappelons aussi que pour intégrer une fonction à valeurs complexes, on peut intégrer séparérament la partie réelle et la partie imaginaire.

La théorie des transformées de Fourier présente plusieurs aspects complémentaires:

1a) La transformée de Fourier des mesures finies sur $\mathbb{R}^d$.

1b) La transformée de Fourier des fonctions (réelles ou complexes) sur $\mathbb{R}^d$, qui sont intégrables par rapport à la mesure de Lebesgue: quitte à considérer séparément la partie réelle et la partie imaginaire, on se ramène aux fonctions réelles; quitte à écrire une fonction réelle comme différence de deux fonctions positives, on se ramène aux fonctions positives (intégrables): la transformée de Fourier de $f \geq 0$ sera alors simplement la transformée de Fourier de la mesure $\mu = f \bullet \lambda_d$: cet aspect se réduit donc essentiellement à 1a.

2) La transformée de Fourier des fonctions complexes de carré intégrable par rapport à λ_d: nous ne ferons que survoler cet aspect.

3) La théorie des fonctions caractéristiques pour les probabilités: c'est d'une certaine manière un cas particulier de 1, dont nous ne développerons aucunement les aspects spécifiques ici.

Définition 1 a) La *transformée de Fourier* de la mesure μ de masse totale finie sur $(\mathbb{R}^d, \mathcal{R}^d)$ est la fonction de $\mathbb{R}^d$ dans $\mathbb{C}$ définie par

$$\hat{\mu}(u) = \int e^{-2i\pi\langle u,x\rangle}\mu(dx), \qquad (1)$$

où $\langle u, x\rangle$ désigne le produit scalaire usuel sur $\mathbb{R}^d$ (si $u = (u_j)$ et $x = (x_j)$, on a $\langle u, x\rangle = \sum_{j=1}^d u_j x_j$).

b) Si f est une fonction à valeurs complexes, intégrable par rapport à la mesure de

Lebesgue, sa *transformée de Fourier* est la fonction de $I\!R^d$ dans $\mathbb{C}$ définie par

$$\hat{f}(u) \;=\; \int e^{-2i\pi\langle u,x\rangle} f(x)dx; \tag{2}$$

on écrit aussi parfois $\mathcal{F}f$ au lieu de $\hat{f}$.

Noter que $|e^{i\langle u,x\rangle}| = 1$, de sorte que dans (1) et (2) les intégrales sont bien définies. Si f est une fonction positive, Lebesgue-intégrable, on a $\hat{f} = \hat{\mu}$ si $\mu = f \bullet \lambda_d$.

Proposition 2 *a) La transformée de Fourier d'une mesure finie (resp. d'une fonction Lebesgue-intégrable) est une fonction continue.*

b) Les applications $\mu \mapsto \hat{\mu}$ et $f \mapsto \hat{f}$ sont linéaires, et on a

$$|\hat{\mu}(u)| \;\leq\; \mu(I\!R^d), \qquad |\hat{f}(u)| \;\leq\; \int |f(x)|dx. \tag{3}$$

c) La transformée de Fourier du produit de convolution de deux mesures finies (resp. d'une mesure finie et d'une fonction intégrable, resp. de deux fonctions intégrables) est le produit des deux transformées de Fourier.

Preuve. (b) est évident (pour (3) on utilise $|e^{-2i\pi\langle u,x\rangle}| = 1$, et 2-(36)). Pour (a) et (c), il suffit par linéarité de considérer le cas des mesures.

Soit μ une mesure finie. Posons aussi $\psi_u(x) = e^{-2i\pi\langle u,x\rangle}$. Pour chaque $x \in I\!R^d$ la fonction $u \mapsto \psi_u(x)$ est continue, et $|\psi_u(x)| \leq 1$: la proposition 3-14 entraine alors immédiatement (a).

Soit $\mu = \mu_1 \star \mu_2$, où μ_1 et μ_2 sont deux mesures finies. On sait que μ est aussi une mesure finie (cf. l'exemple 2 avant la proposition 4-25), et 4-(34) et 4-(22) impliquent

$$\hat{\mu}(u) \;=\; \int e^{-2i\pi\langle u,x+y\rangle}\mu_1(dx)\mu_2(dy) \;=\; \left(\int e^{-2i\pi\langle u,x\rangle}\mu_1(dx)\right)\left(\int e^{-2i\pi\langle u,y\rangle}\mu_2(dy)\right),$$

de sorte que $\hat{\mu}(u) = \hat{\mu_1}(u)\hat{\mu_2}(u)$.

Lorsque μ est une mesure finie et f est une fonction intégrable, quitte à prendre les parties positives et négatives des parties réelle et imaginaire de f, et à utiliser la linéarité de la transformée de Fourier et du produit de convolution, on peut supposer que $f \geq 0$, et on sait alors que $\mu \star f$ est la densité de la mesure $\mu \star (f \bullet \lambda_d)$; d'après ce qu'on vient de voir, la transformée de Fourier de $\mu \star f$ est alors le produit $\hat{\mu}\hat{f}$. Le résultat concernant le produit de convolution de deux fonctions se montre de la même manière. □

Par exemple, la transformée de Fourier de la mesure de Dirac ε_a en $a \in I\!R^d$ est

$$\hat{\varepsilon}_a(u) \;=\; e^{-2i\pi\langle u,a\rangle} \qquad (\text{en particulier,} \quad \hat{\varepsilon}_0(u) = 1\ ). \tag{4}$$

Cela est cohérent avec l'assertion (c) ci-dessus et le fait que $\varepsilon_0 \star \mu = \mu$ et $\varepsilon_0 \star f = f$. Des changements de variables élémentaires dans (2) permettent de montrer les propriétés suivantes, où f est une fonction complexe Lebesgue-intégrable et où $\overline{a}$ désigne le complexe conjugué de a:

$$g(x) = f(-x) \qquad \Rightarrow \qquad \hat{g}(u) = \hat{f}(-u) = \overline{\hat{\overline{f}}(u)}. \tag{5}$$

$$g(x) = f(x/a), \quad a \in \mathbb{R}\backslash\{0\} \qquad \Rightarrow \qquad \hat{g}(u) = a^d \hat{f}(au). \tag{6}$$

Exemple: les séries de Fourier. On sait qu'une série de Fourier est une série de terme général $a_n e^{2in\pi u}$ indicée par $n \in \mathbb{Z}$. Lorsque les a_n sont réels et que $\sum_{n\in\mathbb{Z}} |a_n| < \infty$, la somme d'une telle série apparait donc comme la transformée de Fourier de la mesure suivante sur $\mathbb{R}$:

$$\mu \;=\; \sum_{n\in\mathbb{Z}} a_n \varepsilon_{-n}.$$

2 Injectivité et formule d'inversion

Nous nous proposons de démontrer dans ce paragraphe le résultat fondamental selon lequel deux mesures admettant la même transformée de Fourier sont égales, ainsi que quelques corollaires qui seront énoncés plus loin. Nous allons commencer par un certain nombre de résultats auxiliaires. D'abord, soit la fonction

$$g(x) \;=\; \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{7}$$

Lemme 3 *La fonction g est la densité d'une probabilité sur $\mathbb{R}^d$, et sa transformée de Fourier est $\hat{g}(u) = e^{-2\pi^2 u^2}$.*

Preuve. a) La fonction g est positive, et borélienne puisque continue. Pour montrer que c'est la densité d'une probabilité il suffit donc de prouver que $I = \int g(x)dx$ vaut 1. D'après la proposition 4-20 on a

$$I^2 \;=\; \int_{\mathbb{R}^2} g(x)g(y)dxdy \;=\; \frac{1}{2\pi}\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dxdy.$$

Passons en coordonnées polaires: si $D = \mathbb{R}^2\backslash\{0\}$ et $\Delta =]0,\infty[\times[0,2\pi[$, à tout point $(\rho,\theta) \in \Delta$ on associe un point et un seul $(x,y) = h(\rho,\theta)$ de Δ de sorte que $x = \rho\cos\theta$ et $y = \rho\sin\theta$. h est clairement un C^1-difféomorphisme de Δ dans D, dont le jacobien vaut $Dh(\rho,\theta) = \rho$. Donc en appliquant le théorème 4-21 avec h, Δ et D et la fonction $f(x,y) = e^{-(x^2+y^2)/2}$, et en remarquant que $f\circ h(\rho,\theta) = e^{-\rho^2/2}$, on obtient (puisque l'ensemble $\mathbb{R}^2\backslash D = \{0\}$ est de λ_2-mesure nulle):

$$I^2 \;=\; \frac{1}{2\pi}\int_D f(x,y)dxdy \;=\; \frac{1}{2\pi}\int_\Delta f\circ h(\rho,\theta)\rho d\rho d\theta \;=\; \frac{1}{2\pi}\int_{[0,2\pi[} d\theta \left(\int_{]0,\infty[} e^{-\rho^2/2}\rho d\rho\right)$$

(la dernière égalité vient du théorème de Fubini, la fonction qu'on intègre étant mesurable et positive). En faisant le changement de variable $z = \rho^2/2$ on voit que $\int_0^\infty e^{-\rho^2/2}\rho d\rho = \int_0^\infty e^{-z}dz = 1$, de sorte que $I^2 = \frac{1}{2\pi}\int_0^{2\pi} d\theta = 1$: donc $I = 1$.

b) On a $\hat{g}(u) = \frac{1}{\sqrt{2\pi}}\int f_u(x)dx$, avec $f_u(x) = e^{-2i\pi ux - x^2/2}$. La fonction $u \mapsto f_u(x)$ est clairement dérivable, de dérivée $F_u(x) = -2i\pi x f_u(x)$. Par ailleurs on a $|F_u(x)| \leq 2\pi|x|e^{-x^2/2}$, et la fonction $x \mapsto 2\pi|x|e^{-x^2/2}$ est Lebesgue-intégrable: on peut donc appliquer le théorème de dérivation sous le signe intégral (proposition 3-14), d'après lequel $\hat{g}$ est dérivable, de dérivée donnée par

$$\hat{g}'(u) \;=\; -i\sqrt{2\pi}\int xe^{-2i\pi ux-x^2/2}dx.$$

En faisant une intégration par parties avec $xe^{-x^2/2}$ (dont une primitive est $-e^{-x^2/2}$) et $e^{-2i\pi ux}$ (dont la dérivée en x est $-2i\pi ue^{-2i\pi ux}$), on obtient

$$\hat{g}'(u) \;=\; -i\sqrt{2\pi}e^{-2i\pi ux-x^2/2}|_{-\infty}^{+\infty} - u(2\pi)^{3/2}\int e^{-2i\pi ux-x^2/2}dx \;=\; -4\pi^2u\hat{g}(u).$$

La solution générale de l'équation différentielle à variables séparables $f'(u)=-4\pi^2uf(u)$ étant $f(u)=Ce^{-2\pi^2u^2}$, et comme on a $\hat{g}(0)=\int g(x)dx=1$ d'après (a), on voit que nécessairement $\hat{g}(u)=e^{-2\pi^2u^2}$. $\square$

Ensuite, pour tout $\sigma>0$ on considère la fonction

$$g_\sigma(u) = \frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2} \;=\; \frac{1}{\sigma}g(x/\sigma) \tag{8}$$

(donc $g=g_1$). Il est facile par un changement de variable de vérifier que g_σ est encore la desnité d'une probabilité sur $\mathbb{R}$, et d'après (6) sa transformée de Fourier est

$$\hat{g}_\sigma(u) \;=\; e^{-2\pi^2\sigma^2u^2}. \tag{9}$$

Enfin pour $\sigma>0$ on définit la fonction suivante sur $\mathbb{R}^d$, en utilisant la notation $x=(x_1,\ldots,x_d)$:

$$g_{d,\sigma}(x) \;=\; \prod_{j=1}^d g_\sigma(x_j) \;=\; \frac{1}{(\sigma\sqrt{2\pi})^d}e^{-|x|^2/2\sigma^2}. \tag{10}$$

D'après la proposition 4-20 et (9) sa transformée de Fourier est

$$\hat{g}_{d,\sigma}(u) \;=\; \prod_{j=1}^d \hat{g}_\sigma(u_j) \;=\; e^{-2\pi^2\sigma^2|u|^2}. \tag{11}$$

Lemme 4 *Soit μ une mesure finie sur $(\mathbb{R}^d,\mathcal{R}^d)$. On a:*

a) $(g_{d,\sigma}\star\mu)(x) \;=\; \int_{\mathbb{R}^d}\hat{\mu}(u)e^{2i\pi\langle u,x\rangle-2\pi^2\sigma^2|u|^2}du.$

b) Pour toute fonction continue bornée h sur $\mathbb{R}^d$, l'intégrale $\int hd\mu$ est la limite de $\int_{\mathbb{R}^d}(g_{d,\sigma}\star\mu)(x)h(x)dx$ lorsque $\sigma\to0$.

Preuve. a) Remarquons que $g_{d,\sigma}(x)=\frac{1}{(\sigma\sqrt{2\pi})^d}\hat{g}_{d,1/2\pi\sigma}(-x)$ par (10) et (11). Donc d'après 4-(36) et (2) et le théorème de Fubini, il vient

$$\begin{aligned}(g_{d,\sigma}\star\mu)(x) \;&=\; \textstyle\int g_{d,\sigma}(x-y)\mu(dy) \;=\; \frac{1}{(\sigma\sqrt{2\pi})^d}\int \hat{g}_{d,1/2\pi\sigma}(y-x)\mu(dy)\\ &=\; \textstyle\frac{1}{(\sigma\sqrt{2\pi})^d}\int\mu(dy)\left(\int g_{d,1/2\pi\sigma}(z)e^{i\langle y-x,z\rangle}dz\right)\\ &=\; \textstyle\int e^{-i\langle x,z\rangle-2\pi^2\sigma^2|z|^2/2}dz\left(\int e^{i\langle y,z\rangle}\mu(dy)\right),\end{aligned}$$

d'où le résultat.

b) Soit $I_\sigma = \int (g_{d,\sigma} \star \mu)(x)h(x)dx$. On a la suite d'égalités:

$$\begin{aligned} I_\sigma &= \int h(x)dx \left(\int g_{d,\sigma}(x-y)\mu(dy)\right) = \int \mu(dy) \left(\int h(x)g_{d,\sigma}(x-y)dx\right) \quad \text{(par Fubini)} \\ &= \int \mu(dy) \left(\int h(y+z)g_{d,\sigma}(z)dz\right) \quad \text{(changement de variable } z = x-y) \\ &= \int \mu(dy) \left(\int h(y+z)\tfrac{1}{\sigma^d} g_{d,1}(z/\sigma)dz\right) \quad \text{(puisque } g_{d,\sigma}(z) = \tfrac{1}{\sigma^d} g_{d,1}(z/\sigma)) \\ &= \int \mu(dy) \left(\int h(y+u\sigma)g_{d,1}(u)du\right) \quad \text{(changement de variable } u = z/\sigma). \end{aligned}$$

Posons alors $k_\sigma(y) = \int h(y+u\sigma)g_{d,1}(u)du$, et soit C une constante telle que $|h(x)| \leq C$ pour tout x. On a $|h(u+u\sigma)g_{d,1}(u)| \leq Cg_{d,1}(u)$, et d'après 4-(22) et le fait que g est d'intégrale 1 par rapport à la mesure de Lebesgue, on a $\int_{I\!R^d} g_{d,1}(u)du = 1$ également, de sorte que $|k_\sigma(y)| \leq C$. Comme h est continue, on a $h(y+u\sigma) \to h(y)$ quand $\sigma \to 0$. On peut alors appliquer une première fois le théorème de Lebesgue pour obtenir que $k_\sigma(y)$ converge quend $\sigma \to 0$ vers $\int h(y)g_{d,1}(u)du = h(y)$. En appliquant une seconde fois le même théorème, on obtient que $\int k_\sigma(y)\mu(dy) \to \int h(y)\mu(dy)$, et le résultat est prouvé. □

Nous arrivons maintenant au théorème fondamental *d'injectivité* de la transformée de Fourier:

Théorème 5 *a) La transformée de Fourier $\hat\mu$ caractérise la mesure finie μ (i.e. deux mesures finies ayant même transformée de Fourier sont égales).*

b) La transformée de Fourier $\hat f$ caractérise la fonction complexe Lebesgue-intégrable f à un ensemble λ_d-négligeable près (i.e. deux fonctions intégrables ayant même transformée de Fourier sont égales λ_d-presque partout).

Preuve. a) Il suffit d'appliquer le lemme 4: si on connait $\hat\mu$, on connait aussi $g_{d,\sigma}\star\mu$ d'après le lemme 4-(a), donc aussi $\int h d\mu$ pour toute fonction continue bornée h d'après le lemme 4-(b): il reste à montrer que si μ et μ' sont deux mesures finies telles que $\int hd\mu = \int hd\mu'$ pour toute fonction continue bornée h, on a $\mu = \mu'$. Pour tout rectangle $A = \prod_{j=1}^d]-\infty, a_j[$ il est facile de construire une suite $(h_n)_{n\geq 1}$ de fonctions continues telles que $0 \leq h_n \leq 1$ et que $\lim_n h_n = 1_A$. D'après le théorème de Lebesgue on a $\mu(A) = \lim_n \int h_n d\mu$, et de même pour μ'. Par suite $\mu(A) = \mu'(A)$ pour tout rectangle comme ci-dessus, et on sait que cela entraine $\mu = \mu'$.

b) Si on remplace μ par une fonction positive Lebesgue-intégrable f, le lemme précédent reste encore valide (puisque cela revient à prendre pour μ la mesure $f \bullet \lambda_d$). Par linéarité on remarque alors que le lemme reste aussi valide pour μ remplacé par une fonction complexe intégrable f.

Deux fonctions complexes f et f', Lebesgue-intégrable, ayant même transformée de Fourier vérifient donc $\int_A f(x)dx = \int_A f'(x)dx$ pour tout rectangle $A = \prod_{j=1}^d]-\infty, a_j]$, par le même argument que ci-dessus: le lemme 5-30-(b) (appliqué séparément pour les parties réelles et imaginaires de f et f') permet alors de conclure. □

On peut être plus précis: en combinant les deux assertions du lemme 4 on voit que si h est une fonction continue bornée, on a:

$$\int h d\mu \; = \; \lim_{\sigma \downarrow 0} \int h(x)dx \left(\int \hat{\mu}(u) e^{2i\pi\langle u,x\rangle - 2\pi^2\sigma^2|u|^2} du \right), \tag{12}$$

ce qui est une *formule d'inversion* des transformée de Fourier des mesures finies. Pour les fonctions, on peut faire mieux:

Théorème 6 *a) Si μ est une mesure finie dont la transformée de Fourier $\hat{\mu}$ est Lebesgue-intégrable, elle admet une densité continue et bornée g par rapport à la mesure de Lebesgue, donnée par la formule*

$$g(x) \; = \; \int e^{2i\pi\langle u,x\rangle} \hat{\mu}(u) du. \tag{13}$$

b) Si f est une fonction complexe Lebesgue-intégrable, dont la transformée de Fourier est également Lebesgue-intégrable, on a

$$f(x) \; = \; \int e^{2i\pi\langle u,x\rangle} \hat{f}(u) du \qquad \text{pour } \lambda_d\text{-presque tout } x. \tag{14}$$

Vu le théorème 5(b), dans (b) ci-dessus on ne peut pas faire mieux que l'égalité λ_d-p.p.; d'ailleurs, le membre de droite de (14) est continu borné, ce qui n'est pas nécessairement le cas de f.

Preuve. a) Soit g définie par (13). L'intégrand du membre de droite est continu en x et majoré en module par la fonction intégrable $|\hat{\mu}|$, donc g est bornée, et continue grâce à la proposition 3-14. Par ailleurs, si h est continue à support compact dans $I\!R^d$, on peut échanger limite et intégrales dans le membre de droite de (12) (théorème de Lebesgue). On obtient alors $\int h d\mu = \int h(x)g(x)dx$ pour toute fonction h continue à support compact.

Soit maintenant $\mathcal{C}$ la classe des rectangles $A = \prod_{j=1}^d]a_j, b_j]$ avec $-\infty < a_j < b_j < \infty$. Cette classe est stable par intersection, contient une suite $(E_n)_{n\geq 1}$ croissant vers $I\!R^d$, et engendre la tribu $\mathcal{R}^d$. De plus si $A \in \mathcal{C}$ il est facile de construire des fonctions h_n, h, continues à support compact, telles que $h_n \to 1_A$ et $0 \leq h_n \leq h$. On déduit alors de $\int h_n d\mu = \int h_n(x)g(x)dx$ et du théorème de Lebesgue que $\mu(A) = \int_A g(x)dx$.

Le lemme 5-30-(b) appliqué aux fonctions 0 et g' = partie imaginaire de g (qui vérifie $\int_A g'(x)dx = 0$ pour tout $A \in \mathcal{C}$ d'après ce qui précède) implique $g' = 0 \;\; \lambda_d$-p.p., et la continuité de g (donc de g') entraine qu'en fait $g' = 0$, de sorte que g est à valeurs réelles.

Soit alors les mesures $\nu_+ = g^+ \bullet \lambda_d$ et $\nu_- = g^- \bullet \lambda_d$, qui vérifient $\nu_+(A) < \infty$ et $\nu_-(A) < \infty$ pour $A \in \mathcal{C}$. On a donc en fait $\mu(A) + \nu_-(A) = \nu_+(A)$ pour tout $A \in \mathcal{C}$, et le théorème 4-1 implique $\mu + \nu_- = \nu_+$. Si alors $N \in \mathcal{R}^d$ est λ_d-négligeable, il vient $\nu_+(N) = \nu_-(N) = 0$, donc $\mu(N) = 0$: par suite μ est absolument continue par rapport à λ_d, et d'après le théorème de Radon-Nikodym il existe une fonction k positive Lebesgue-intégrable, telle que $\mu = k \bullet \lambda_d$. Si $E_n =]-n,n]^d$ les fonctions $k1_{E_n}$ et $g1_{E_n}$ sont Lebesgue-intégrables et vérifient $\int_A (k1_{E_n})(x)dx = \int_{A\cap E_n} k(x)dx = \int_{A\cap E_n} g(x)dx = \int_A (g1_{E_n})(x)dx$ pour tout $A \in \mathcal{C}$, donc le lemme 5-30-(b) entraine $k1_{E_n} = g1_{E_n} \;\; \lambda_d$-p.p. pour tout n. On a donc aussi $k = g \;\; \lambda_d$-p.p., ce qui achève la démonstration de (a).

b) Lorsque $f \geq 0$ le résultat découle de (a) appliqué à la mesure $\mu = f \bullet \lambda_d$ (puisqu'alors $\hat{\mu} = \hat{f}$, et que si g est une densité de μ par rapport à λ_d on a $f = g$ λ_d-p.p. d'après le lemme 5-30). On passe au cas général en prenant les parties positives et négatives des parties réelle et imaginaire de f. $\square$

3 Quelques résultats de densité

Nous interrompons un moment l'exposé de la théorie de la transformée de Fourier pour donner les résultats de "densité" qui nous seront nécessaires. Le premier est un résultat général de théorie de la mesure.

Proposition 7 *Soit $(E, \mathcal{E})$ un espace mesurable muni d'une mesure finie μ et $\mathcal{G}$ une algèbre de parties de E, engendrant la tribu $\mathcal{E}$. Pour tout $A \in \mathcal{E}$ il existe une suite $(A_n)_{n \geq 1}$ d'éléments de $\mathcal{G}$ telle que $\mu(A \Delta A_n) \to 0$ quand $n \to \infty$.*

Preuve. Notons $\mathcal{D}$ la classe des $A \in \mathcal{E}$ pour lesquels il existe une suite $A_n \in \mathcal{G}$ telle que $\mu(A \Delta A_n) \to 0$. Soit $A, B \in \mathcal{D}$ avec $A \subset B$, et deux suites $A_n, B_n \in \mathcal{G}$ associées comme ci-dessus. Comme $\mathcal{G}$ est une algèbre on a $C_n = B_n \cap (A_n)^c \in \mathcal{G}$, tandis que $(B \backslash A) \Delta C_n \subset (A \Delta A_n) \cup (B \Delta B_n)$. On a donc

$$\mu((B \backslash A) \Delta C_n) \ \leq \ \mu(A \Delta A_n) + \mu(B \Delta B_n) \ \to \ 0,$$

de sorte que $B \backslash A \in \mathcal{D}$. De même si $A_n \in \mathcal{D}$ est une suite croissante, de limite A, pour tout $m \in I\!N^*$ il existe n tel que $\mu(A \backslash A_n) \leq 1/m$; pour tout $i \leq n$ il existe $C_i \in \mathcal{G}$ tel que $\mu(A_i \Delta C_i) \leq 1/nm$. Si alors $B_m = \cup_{i=1}^n C_i$, on a $B_m \in \mathcal{G}$ et $A \Delta B_m \subset (A \backslash A_n) \cup (\cup_{i=1}^n A_i \Delta C_i)$, donc

$$\mu(A \Delta B_m) \ \leq \ \mu(A \backslash A_n) + \sum_{i=1}^n \mu(A_i \Delta C_i) \ \leq \ \frac{1}{m} + \frac{n}{nm} \ = \ \frac{2}{m},$$

donc $\mu(A \Delta B_m) \to 0$ quand $m \to \infty$. Par suite $\mathcal{D}$ est un λ-système, et le lemme 4-2 implique que $\mathcal{D} = \mathcal{E}$: on a donc le résultat cherché. $\square$

Le résultat suivant est plus qu'il nous faut pour la suite:

Proposition 8 *Soit μ une mesure de Radon sur $I\!R^d$ (= une mesure telle que $\mu(K) < \infty$ pour tout compact K). Si $p \in [1, \infty[$ et si $f \in L^p = L^p(I\!R^d, \mathcal{R}^d, \mu)$, il existe une suite $(f_n)_{n \geq 1}$ de fonctions indéfiniment différentiables à supports compacts qui converge vers f dans L^p.*

Preuve. Quitte à approcher séparément f^+ et f^-, on peut supposer que $f \geq 0$. Si les (g_n) vérifient $0 \leq g_n \leq f$ et croisssent vers f, on a $g_n \to^{L^p} f$ par 5-(14): il suffit donc d'approcher dans L^p chaque fonction g_n par une suite de fonctions C^∞ à supports compacts, donc on peut en fait supposer f étagée. Si $f = \sum_{j=1}^k a_j 1_{A_j}$, par linéarité il suffit d'approcher chaque indicatrice 1_{A_j}: par suite on peut supposer que $f = 1_A$ avec $\mu(A) < \infty$ (puisque $f \in L^p$).

Soit les ensembles $E_n =]-n, n]^d$. Si $\varepsilon > 0$ il existe m tel que $\mu(A \cap (E_m)^c) \leq \varepsilon$ puisque $\mu(A) < \infty$. Par ailleurs notons $\mathcal{G}$ la classe des réunions finies de rectangles deux-à-deux disjoints de la forme $\prod_{j=1}^d]a_j, b_j]$: il est très simple de vérifier que $\mathcal{G}$ est une algèbre, et on sait que la tribu engendrée est $\mathcal{R}^d$. Le lemme précédent appliqué à la restriction de μ à E_m (qui est une mesure finie puisque μ est de Radon) permet de trouver $B \in \mathcal{G}$ tel que $\mu(E_m \cap (A\Delta B)) \leq \varepsilon$, et on peut bien-sûr supposer que $B \subset E_m$. On a $||1_A - 1_B||_p = \mu(A\Delta B)^{1/p}$, et $\mu(A\Delta B) \leq \mu(A \cap (E_m)^c) + \mu(E_m \cap (A\Delta B)) \leq 2\varepsilon$. Comme ε est arbitraire, il suffit donc de montrer le résultat pour chaque B ci-dessus, ce qui revient à supposer que $A \in \mathcal{G}$ et $A \subset E_m$ pour un m. Enfin, par linéarité une nouvelle fois, il suffit de considérer le cas où A est un rectangle borné: il est alors très facile de construire des fonctions indéfiniment différentiables f_n telles que $0 \leq f_n \leq 1_{E_m}$ pour un m fixé, et que $f_n(x) \to 1_A(x)$ pour tout x. En appliquant une nouvelle fois 5-(14) on obtient que $f_n \to^{L^p} 1_A$, et la preuve est achevée. $\square$

Remarque: Ce résultat est faux lorsque $p = \infty$: on ne peut pas approcher une indicatrice d'ensemble par une suite de fonctions continues, au sens de L^∞: en effet, la convergence dans L^∞ est "presque" la convergence uniforme. De la même manière, les quelques résultats qui suivent sont faux pour $p = \infty$.

Voici maintenant quelques applications.

Lemme 9 *Soit f une fonction de $L^p = L^p(\mathbb{R}^d, \mathcal{R}^d, \lambda_d)$, pour un $p \in [1, \infty[$, et notons $\tau_t f$ la "translatée" de f définie par $\tau_t f(x) = f(x+t)$ (pour $t \in \mathbb{R}^d$). Alors $t \mapsto \tau_t f$ est une fonction continue de $\mathbb{R}^d$ dans L^p.*

Preuve. Par un changement de variable évident, il est clair que $\tau_t f$ est dans L^p et $||\tau_t f||_p = ||f||_p$. Soit $\varepsilon > 0$. La proposition précédente nous donne une fonction continue à support compact g telle que $||f - g||_p \leq \varepsilon$. On a

$$||\tau_t f - \tau_s f||_p \;\leq\; ||\tau_t f - \tau_t g||_p + ||\tau_t g - \tau_s g||_p + ||\tau_s g - \tau_s f||_p.$$

On a $||\tau_t f - \tau_t g||_p = ||\tau_s g - \tau_s g||_p = ||f - g||_p \leq \varepsilon$. Par ailleurs si s est fixé on a $\tau_t g(x) \to \tau_s g(x)$ pour tout x lorsque $t \to s$ puisque g et continue, et $|\tau_t g|$ est majoré par $C 1_K$ pour une certaine constante C et un compact convenable K lorsque t décrit la boule de centre s et de rayon 1: cette fonction étant dans L^p, 5-(14) implique $||\tau_t g - \tau_s g||_p \leq \varepsilon$ si t est assez proche de s. Par suite $||\tau_t f - \tau_s f||_p \leq 3\varepsilon$ pour t assez proche de s, et on a le résultat puisque ε est arbitraire. $\square$

Corollaire 10 *Si f est dans $L^p = L^p(\mathbb{R}^d, \mathcal{R}^d, \lambda_d)$ pour un $p \in [1, \infty[$, les fonctions $g_{d,\sigma} \star f$ convergent vers f dans L^p.*

Preuve. Lorsque $p = 1$ il n'y a pas de problème pour définir le produit de convolution puisque les deux fonctions sont intégrables. Si $p > 1$, la fonction $y \mapsto f(x-y)$ est dans L^p (mais pas forcément dans L^1), et il est facile de vérifier que si $1/p + 1/q = 1$, alors $g_{d,\sigma}$ est dans L^q: d'après Hölder, le produit de ces deux fonctions est dans L^1, de sorte qu'on peut définir le produit de convolution par la formule 4-(37).

Comme $\int g_{d,\sigma}(x)dx = 1$, on a

$$\begin{aligned} ||g_{d,\sigma} \star f - f||_p^p &= \int dx \left| \int g_{d,\sigma}(y)(f(x-y) - f(x))dy \right|^p \\ &\leq \int dx \left(\int g_{d,\sigma}(y)|f(x-y) - f(x)|^p dy \right) \end{aligned}$$

en appliquant Hölder aux fonctions $y \mapsto f(x-y) - f(x)$ et $y \mapsto 1$, pour $1/p + 1/q = 1$ et relativement à la probabilité de densité $g_{d,\sigma}$ par rapport à λ_d. D'après Fubini, il vient alors

$$||g_{d,\sigma} \star f - f||_p^p \leq \int g_{d,\sigma}(y)||\tau_{-y}f - f||_p^p dy = \int g_{d,1}(z)||\tau_{-z\sigma} - f||_p^p dz$$

par le changement de variables $y = z\sigma$. Il suffit alors d'appliquer le lemme précédent, le théorème de Lebesgue et le fait que $||\tau_t f - f||_p \leq 2||f||_p$ pour obtenir que l'expression ci-dessus tend vers 0 si $\sigma \to 0$. $\square$

Terminons par une application aux transformées de Fourier. La transformée de Fourier d'une fonction intégrable n'est pas nécessairement intégrable, mais on a:

Proposition 11 *Si f est Lebesgue-intégrable sur $I\!\!R^d$, alors $\hat{f}(u) \to 0$ quand $|u| \to \infty$.*

Preuve. On pose $h_\sigma = g_{d,\sigma} \star f - f$. (3) implique $|\hat{h}_\sigma| \leq ||h_\sigma||_1$, qui tend vers 0 d'après le corollaire ci-dessus. La proposition 2 entraîne que $\hat{h}_\sigma = (\hat{g}_{d,\sigma} - 1)\hat{f}$, de sorte que (11) implique $\hat{f}(u) = \hat{h}_\sigma(u)/(e^{-2\pi^2\sigma^2|u|^2} - 1)$. Si $\varepsilon > 0$ on choisit alors σ de sorte que $||h_\sigma||_1 \leq \varepsilon$, puis A de sorte que $1 - e^{-2\pi^2\sigma^2 A^2} \geq 1/2$. Si $|u| > A$ on a alors $|\hat{f}(u)| \leq 2\varepsilon$, et comme ε est arbitraire on a le résultat. $\square$

4 La transformée de Fourier dans L^2

Nous allons voir qu'on peut aussi définir la transformée de Fourier des fonctions sur $I\!\!R^d$ qui sont de carré intégrable (et pas nécessairement intégrables). Dans ce cas, la formule (2) peut ne pas avoir de sens, et il faut opérer autrement.

Dans ce paragraphe, nous notons $L^2_{\mathbb{C}}$ l'ensemble des (classes d'équivalence pour l'égalité presque partout des) fonctions complexes sur $I\!\!R^d$, dont le carré du module $|f|^2$ est Lebesgue-intégrable. C'est évidemment un espace vectoriel sur le corps $\mathbb{C}$, sur lequel on définit une norme $||f||_2 = \sqrt{\int |f(x)|^2 dx}$. De manière plus précise, cette norme est associée au produit scalaire - complexe - défini par $\langle f, g \rangle = \int f(x)\overline{g(x)}dx$, et on $||f||_2^2 = \langle f, f \rangle$: tout marche comme dans le cas réel, sauf que la symétrie du produit scalaire est remplacée ici par la propriété $\langle f, g \rangle = \overline{\langle g, f \rangle}$. On démontre exactement comme au chapitre précédent que $L^2_{\mathbb{C}}$ est un *espace de Hilbert* (sur $\mathbb{C}$).

Commençons par un lemme, où on désigne par C_{int} l'ensemble des fonctions complexes sur $I\!\!R^d$ qui sont continues, bornées et Lebesgue-intégrables. Une telle fonction f vérifie $|f|^2 \leq C|f|$ si $C = \sup|f(x)|$, de sorte qu'elle est aussi de carré intégrable.

Lemme 12 *Si $f \in C_{int}$, alors $\hat{f} \in L^2_{\mathbb{C}}$ et $||f||_2 = ||\hat{f}||_2$.*

Preuve. Exactement comme dans la preuve du théorème 6, le lemme 4 est valable avec μ remplacée par la fonction intégrable f, à condition que dans la partie (b) on lise $\int f(x)h(x)dx$ au lieu de $\int hd\mu$. Il vient alors, puisque $|f|^2 = f\overline{f}$ et f est continue bornée:

$$\begin{aligned} ||f||_2^2 \;=\; \int f(x)\overline{f(x)}dx &= \lim_{\sigma\downarrow 0} \int \overline{f(x)}dx \left(\int \hat{f}(u)e^{2i\pi\langle u,x\rangle - 2\pi^2\sigma^2|u|^2}du\right) \\ &= \lim_{\sigma\downarrow 0} \int \hat{f}(u)e^{-2\pi^2\sigma^2|u|^2}du \left(\int \overline{f(x)}e^{2i\pi\langle u,x\rangle}dx\right) \\ &= \lim_{\sigma\downarrow 0} \int \hat{f}(u)e^{-2\pi^2\sigma^2|u|^2}\overline{\hat{f}(u)}du, \end{aligned}$$

où la seconde égalité vient du théorème de Fubini (qu'on peut appliquer puisque $\hat{f}$ est bornée et f est intégrable). L'intégrand de la dernière expression ci-dessus est réel positif et croît vers $|\hat{f}(u)|^2$ lorsque $\sigma \downarrow 0$: le résultat provient alors du théorème de limite monotone. □

Rappelons qu'on note aussi $\mathcal{F}f = \hat{f}$. Ce qui précède signifie qu'on peut considérer $\mathcal{F}$ comme une application du sous-espace C_{int} de $L^2_{\mathbb{C}}$ dans $L^2_{\mathbb{C}}$, qui est clairement linéaire, et que cette application préserve la norme $||.||_2$.

Théorème 13 *L'application $\mathcal{F}$ de C_{int} dans $L^2_{\mathbb{C}}$ définie ci-dessus admet une extension unique, notée encore $\mathcal{F}$, de $L^2_{\mathbb{C}}$ dans lui-même, qui est un isomorphisme d'espaces de Hilbert (= elle est linéaire bijective et préserve la norme), et qui coïncide avec la transformée de Fourier du (2) pour les fonctions de $L^2_{\mathbb{C}}$ qui sont Lebesgue-intégrables. De plus, l'inverse de $\mathcal{F}$ sur $L^2_{\mathbb{C}}$ est donnée par*

$$(\mathcal{F}^{-1}f)(u) \;=\; (\mathcal{F}f)(-u). \tag{15}$$

Si $f \in L^2_{\mathbb{C}}$, la fonction $\mathcal{F}f$ est encore appelée la transformée de Fourier de f, et on l'écrit même parfois sous la forme (2) bien que l'intégrale n'ait pas de sens en général. Noter toutefois que dans ce cas, $\mathcal{F}f$ est la limite dans $\mathbb{L}^2$ des fonctions $u \mapsto \int_{\{x:|x|\leq A\}} e^{-2i\pi\langle u,x\rangle}f(x)dx$ lorsque $A \to \infty$. Remarquer aussi que (15) est l'analogue de (14). Enfin, $\mathcal{F}^{-1}$ est appelée la *transformée de Fourier inverse*.

Preuve. a) l'existence et l'unicité de l'extension vont provenir de ce que C_{int} est dense dans $L^2_{\mathbb{C}}$, ce qui signifie que toute fonction f de $L^2_{\mathbb{C}}$ est limite pour la norme $||.||_2$ d'une suite $(f_n)_{n\geq 1}$ de fonctions de C_{int}: cette propriété découle immédiatement de la proposition 8 appliquée aux parties réelle et imaginaire de f, compte tenu du fait qu'une fonction indéfiniment dérivable à support compact est dans C_{int}.

Soit en effet f et f_n comme ci-dessus. La suite (f_n) est de Cauchy dans $L^2_{\mathbb{C}}$, donc il en est de même de la suite $(\mathcal{F}f_n)$ par le lemme 12, donc cette dernière suite converge vers une limite notée $\mathcal{F}f$. Si (f'_n) est une autre suite de C_{int} telle que $||f'_n - f||_2 \to 0$, on a aussi $||f'_n - f_n||_2 \to 0$, donc $||\mathcal{F}f'_n - \mathcal{F}f_n||_2 \to 0$: en d'autres termes, $\mathcal{F}f$ ne dépend pas de la suite (f_n) choisie, et cela définit une extension de $\mathcal{F}$ à $L^2_{\mathbb{C}}$ qui est évidemment linéaire, et qui préserve la norme. Si $\mathcal{F}'$ était une autre extension, on aurait aussi $||\mathcal{F}f_n - \mathcal{F}'f||_2 = ||f_n - f||_2 \to 0$, de sorte que nécessairement $\mathcal{F}'f = \mathcal{F}f$: donc l'extension est unique.

b) Supposons maintenant que $f \in L^2_{\boldsymbol{C}}$ soit en plus Lebesgue-intégrable. Nous pouvons définir sa transformée de Fourier $\hat{f}$ par (2), et aussi la fonction $\mathcal{F}f$ comme ci-dessus. En examinant la preuve de la proposition 8 on voit facilement qu'on peut trouver une suite (f_n) de fonctions indéfiniment dérivables à support compact, convergeant vers f dans $L^2_{\boldsymbol{C}}$ et dans $L^1_{\boldsymbol{C}}$ simultanément ($L^1_{\boldsymbol{C}}$ désigne évidemment l'espace des fonctions complexes Lebesgue-intégrable, avec la norme $||f||_1 = \int |f(x)|dx$). D'une part la proposition 2 implique que $|\hat{f}_n - \hat{f}| \leq ||f_n - f||_1 \to 0$; d'autre part on a vu ci-dessus que $\hat{f}_n = \mathcal{F}f_n \to \mathcal{F}f$ dans $L^2_{\boldsymbol{C}}$. On en déduit que $\mathcal{F}f = \hat{f}$.

c) Soit G l'image de $L^2_{\boldsymbol{C}}$ par $\mathcal{F}$. Nous allons montrer maintenant que $G = L^2_{\boldsymbol{C}}$: cela achèvera de prouver que $\mathcal{F}$ est un isomorphisme.

D'abord, comme $\mathcal{F}$ est linéaire, G est un espace vectoriel, et on va voir qu'il est fermé: si $f_n \in L^2_{\boldsymbol{C}}$ et si $\mathcal{F}f_n \to g$, on a $||f_n - f_m||_2 = ||\mathcal{F}f_n - \mathcal{F}f_n||_2 \to 0$ qund $n, m \to \infty$, donc la suite (f_n) converge vers une limite f dans $L^2_{\boldsymbol{C}}$; en vertu de ce qui précède, on a donc $g = \mathcal{F}f$, donc $g \in G$ et G est fermé.

Comme G est un sous-espace vectoriel fermé de $L^2_{\boldsymbol{C}}$, pour montrer que $G = L^2_{\boldsymbol{C}}$ il suffit en vertu de la proposition 5-21 de montrer que si $f \in L^2_{\boldsymbol{C}}$ est orthogonal à G, alors $f = 0$. Mais on a vu que $g_{d,\sigma}$ est la transformée de Fourier d'une fonction de C_{int} (cf. (10) et (11)), donc $g_{d,\sigma} \in G$. Il en est de même de ses translatées $\tau_a g_{d,\sigma}$ (car on a $\tau_a(\mathcal{F}h) = \mathcal{F}h'$ si $h'(x) = h(x)e^{-2i\pi\langle a,x\rangle}$). Donc si $f \in L^2_{\boldsymbol{C}}$ est orthogonale à G on a

$$(g_{d,\sigma} \star \overline{f})(x) \;=\; \int g_{d,\sigma}(y-x)\overline{f(y)}dy \;=\; \langle \tau_{-x} g_{d,\sigma}, f\rangle \;=\; 0,$$

où ci-dessus $\langle .,. \rangle$ désigne le produit scalaire dans $L^2_{\boldsymbol{C}}$. Ceci étant vrai pour tout $\sigma > 0$, le corollaire 10 implique que $f = 0$.

d) Il reste à prouver (15). Lorsque $f \in L^2_{\boldsymbol{C}} \cap L^1_{\boldsymbol{C}}$, cette formule n'est autre que (14). Comme $\mathcal{F}$ et $\mathcal{F}^{-1}$ préservent la norme $||.||_2$, et comme $L^2_{\boldsymbol{C}} \cap L^1_{\boldsymbol{C}}$ est dense dans $L^2_{\boldsymbol{C}}$, le résultat est alors évident. $\square$