J. Melleray & T. Blossier

Université Lyon I
Semestre d'automne 2015-2016

Avertissement. Ces notes contiennent un résumé de ce qui a été fait en cours. Comme en cours, beaucoup de théorèmes sont admis ; nous avons passé beaucoup de temps à traiter les exercices des feuilles de TD, dont les corrections n'apparaissent pas ici. Ces notes sont donc plutôt à considérer comme une liste des théorèmes/techniques à connaître, et seront à compléter par un cours plus complet pour ceux et celles qui désirent des démonstrations détaillées, et surtout à accompagner par beaucoup d'exercices pour acquérir la pratique nécessaire.

Table des matières

Chapitre 1

Une introduction intuitive à la mesure de Lebesgue et aux grands théorèmes de théorie de la mesure.

1.1 Mesures ; le presque partout

Définition 1.1. Une mesure sur un ensemble X est une application $\mu\colon \mathcal{P}(X) \to [0,+\infty]$ avec les propriétés suivantes :

1. $\mu(\emptyset) = 0$.
2. Si $(A_i)_{i\in\mathbb{N}}$ est une suite de parties de X deux à deux disjointes, c'est-à-dire si $A_i \cap A_j = \emptyset$ pour tout $i \neq j$, alors $\mu(\bigcup_{i\in\mathbb{N}} A_i) = \displaystyle\sum_{i=0}^{+\infty} \mu(A_i)$.

Remarque 1.2. Ci-dessus, $\mathcal{P}(X)$ désigne l'ensemble des parties de X. La définition donnée plus haut est très exigeante : en général, on ne demande pas que *toutes* les parties de X aient une mesure. Ceci amène à introduire une notion un peu plus précise (en particulier, dire quelles propriétés on attend de l'ensemble des parties qui admettent une mesure, qu'on appelle les parties *mesurables*). Dans la première partie du cours, on va prétendre que toutes les parties sont mesurables ; pour que les théorèmes présentés dans cette première partie soient vraiment corrects, on va ajouter en gris les hypothèses nécessaires de mesurabilité - que vous pouvez ignorer en première lecture, et sur lesquelles vous pourrez revenir après avoir lu la deuxième partie du cours.

Définition 1.3. Comme on autorise que des parties aient une mesure infinie (intuitivement, la mesure généralise les notions de longueur/d'aire/de volume, et on « voit »bien que $\mathbb{R}$, par exemple, est de longueur infinie), il faut donner quelques précisions :

- Pour tout $t \in [0,+\infty]$, $+\infty + t = t + \infty = +\infty$.
- Pour tout t *fini*, $+\infty - t = +\infty$.
- $(+\infty) - (+\infty)$ n'est pas défini.

Dans l'énoncé des propriétés de l'intégrale, on va aussi devoir écrire des multiplications, avec les conventions suivantes :

- $0 \cdot (+\infty) = (+\infty) \cdot 0 = 0$.
- Si $t > 0$, $t \cdot (+\infty) = (+\infty) \cdot t = +\infty$.

Proposition 1.4. *Soit μ une mesure sur un ensemble X.*

- *Si $A_0, \ldots, A_n$ sont des parties deux à deux disjointes et mesurables, alors*

$$\mu(\bigcup_{i=0}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i) \ .$$

- *Si $A \subseteq B$ et A et B sont mesurables alors $\mu(A) \leq \mu(B)$.*

Preuve:

Pour voir que la première propriété est vraie, il suffit de définir $A_{n+1} = A_{n+2} = \ldots = \emptyset$ et d'appliquer la définition d'une mesure :

$$\mu(\bigcup_{i=0}^{n} A_i) = \mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i=0}^{+\infty} \mu(A_i) = \sum_{i=0}^{n} \mu(A_i) + \sum_{i=n+1}^{+\infty} \mu(A_i) = \sum_{i=0}^{n} \mu(A_i) + 0 = \sum_{i=0}^{n} \mu(A_i) \ .$$

Pour la deuxième propriété on utilise le fait que $B = A \cup (B \setminus A)$ et que $B \setminus A$ est mesurable pour écrire :

$$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A) \ .$$

Attention : si $A \subseteq B$ et A et B sont mesurables, on a envie d'écrire $\mu(A) = \mu(B) - \mu(B \setminus A)$. Ceci peut ne pas avoir de sens, si $\mu(B \setminus A)$ est infini ! Avant d'écrire une telle égalité, il faut vérifier que $\mu(B \setminus A) < +\infty$.

Pour l'instant, on va se concentrer sur une mesure, sur un ensemble très particulier : la mesure de Lebesgue sur $\mathbb{R}$.

Définition 1.5. Il existe une unique mesure λ sur $\mathbb{R}$ (définie sur la tribu borélienne de $\mathbb{R}$), appelée *mesure de Lebesgue* et telle que pour tout segment $[a, b]$ on ait $\lambda([a, b]) = b - a$.

Remarque 1.6. Par convention, à chaque fois qu'on écrira $[a, b]$, on supposera que a, b sont des réels et $a \leq b$.

Intuitivement, $b - a$ correspond à la longueur du segment $[a, b]$; la mesure de Lebesgue permet d'associer une notion de longueur à des parties plus compliquées de $\mathbb{R}$ (au moins, à toutes les parties boréliennes).

Notons quelques conséquences immédiates de cette définition.

Proposition 1.7.
- *Pour tout $x \in \mathbb{R}$ on a $\lambda(\{x\}) = 0$.*
- *Si (x_n) est une suite de nombres réels, alors $\lambda(\{x_n\}_{n\in\mathbb{N}}) = 0$.*
- *Pour tous réels a, b avec $a \leq b$ on a $\lambda([a, b]) = \lambda(]a, b[) = \lambda([a, b[) = \lambda(]a, b]) = b - a$.*

Une idée fondamentale en théorie de la mesure est qu'on peut, la plupart du temps, *négliger* les parties de mesure nulle : elle n'ont pas d'influence sur la valeur d'une intégrale, par exemple.

Définition 1.8. Une partie borélienne $A \subseteq \mathbb{R}$ est dite *négligeable* si $\lambda(A) = 0$. On dit aussi que *presque tout x* appartient à A, ou encore que $x \in A$ est *vrai presque partout.*

En particulier, si $f, g \colon \mathbb{R} \to \mathbb{R}$ sont des fonctions boréliennes, on dit que f et g sont égales *presque partout* si $\lambda(\{x \colon f(x) \neq g(x)\}) = 0$ (c'est-à-dire, si l'ensemble de tous les x tels que $f(x) \neq g(x)$ est négligeable, ou encore si $f(x) = g(x)$ pour presque tout x)

Exemple. La fonction f définie sur $\mathbb{R}$ par $f(x) = \begin{cases} 0 & \text{si } x \notin \mathbb{N} \\ x & \text{sinon} \end{cases}$ est borélienne et nulle presque partout.

1.2 L'intégrale de Lebesgue

L'idée de base de la construction de l'intégrale de Lebesgue est la suivante : si $f \colon \mathbb{R} \to \mathbb{R}$ est borélienne et ne prend qu'un nombre fini de valeurs $\alpha_1, \ldots, \alpha_n$, alors on voudrait avoir $\displaystyle\int_{\mathbb{R}} f d\lambda = \sum_{i=1}^{n} \alpha_i \lambda(\{x \colon f(x) = \alpha_i\})$. Par exemple, si f est une fonction en escalier (qui est continue par morceaux et donc borélienne), cette définition nous redonne la valeur habituelle de l'intégrale de f.

Dans le cadre de la théorie de la mesure, on commence par utiliser un procédé de passage à la limite (qu'on ne détaillera pas) pour étendre la définition ci-dessus à l'intégrale de toutes les fonctions positives et mesurables ; une fois qu'on sait intégrer toutes les fonctions à valeurs positives, on peut définir facilement l'intégrale de n'importe quelle fonction mesurable. En effet, étant donnée une fonction f, on peut définir sa partie positive f^+ et sa partie négative f^- en posant

$$f^+(x) = \begin{cases} f(x) & \text{si } f(x) \geq 0 \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad f^-(x) = \begin{cases} -f(x) & \text{si } f(x) \leq 0 \\ 0 & \text{sinon} \end{cases} \ .$$

Alors, f^+ et f^- sont deux fonctions mesurables et à valeurs positives, et par définition $f = f^+ - f^-$. Si on sait intégrer les fonctions mesurables et à valeurs positives, il nous reste donc simplement à poser $\int_{\mathbb{R}} f d\lambda = \int_{\mathbb{R}} f^+ d\lambda - \int_{\mathbb{R}} f^- d\lambda$ (et ce processus s'applique pour toute mesure, pas seulement pour la mesure de Lebesgue).

On va admettre qu'on peut construire une notion d'intégrale ayant les propriétés suivantes :

1. Si f est la fonction caractéristique d'une partie A de $\mathbb{R}$ borélienne, alors $\int_{\mathbb{R}} f d\lambda = \lambda(A)$.
2. Pour toutes fonctions boréliennes $f, g \colon \mathbb{R} \to [0, +\infty]$ et toute constante $c \geq 0$, on a :

$$\int_{\mathbb{R}} cf d\lambda = c \int_{\mathbb{R}} f d\lambda \text{ et } \int_{\mathbb{R}} (f+g) d\lambda = \int_{\mathbb{R}} f d\lambda + \int_{\mathbb{R}} g d\lambda \ .$$

   (En particulier, l'intégrale de la fonction nulle vaut 0.)
3. Pour toutes fonctions boréliennes $f, g \colon \mathbb{R} \to [0, +\infty]$,

$$\text{si } f = g \text{ presque partout alors } \int_{\mathbb{R}} f d\lambda = \int_{\mathbb{R}} g d\lambda.$$

4. Pour toutes fonctions boréliennes $f, g \colon \mathbb{R} \to [0, +\infty]$,

$$\text{si } f \leq g \text{ (presque partout) alors } 0 \leq \int_{\mathbb{R}} f d\lambda \leq \int_{\mathbb{R}} g d\lambda.$$

Remarque 1.9. Il est pratique d'autoriser le fait que les fonctions prennent la valeur $+\infty$. Si $\lambda(\{x \colon f(x) = +\infty\}) > 0$, et $f \colon \mathbb{R} \to [0, +\infty]$ est borélienne, alors on a toujours $\int_{\mathbb{R}} f d\lambda = +\infty$. En particulier, si $f \colon \mathbb{R} \to [0, +\infty]$ est d'intégrale finie alors $f(x) < +\infty$ presque partout, et on peut se ramener à considérer l'intégrale d'une fonction à valeurs finies en intégrant f sur l'ensemble $\{x \colon f(x) < +\infty\}$, ce que permet la définition suivante.

Définition 1.10. Étant donnée une partie A de $\mathbb{R}$ borélienne, et $f \colon \mathbb{R} \to [0, +\infty]$ borélienne, on définit

$$\int_A f d\lambda = \int_{\mathbb{R}} f \mathbf{1}_A d(\lambda) \ .$$

(En particulier, pour toute constante $c \geq 0$, $\int_A c d\lambda = c\lambda(A)$.)

Remarque 1.11. $\mathbf{1}_A$ désigne la *fonction caractéristique* de A, c'est-à-dire que $\mathbf{1}_A(x)$ vaut 1 si $x \in A$ et 0 sinon. C'est une fonction borélienne si, et seulement si, A est borélien. Ainsi, $f\mathbf{1}_A$ est la fonction qui est égale à f sur A, et à 0 ailleurs.

Proposition 1.12. *5. Si A est borélien et $\lambda(A) = 0$ alors $\int_A f d\lambda = 0$ pour toute fonction $f \colon \mathbb{R} \to [0, +\infty]$ borélienne.*

6. Si A_1, A_2 sont des parties de $\mathbb{R}$ boréliennes et disjointes, alors pour toute fonction $f \colon \mathbb{R} \to [0, +\infty]$ borélienne, on a

$$\int_{A_1 \cup A_2} f d\lambda = \int_{A_1} f d\lambda + \int_{A_2} f d\lambda \ .$$

Exercice 1.13. Déduire des propriétés énoncées plus haut que si A_1, A_2 sont des parties de $\mathbb{R}$ boréliennes et telles que $\lambda(A_1 \cap A_2) = 0$, alors pour toute fonction $f \colon \mathbb{R} \to [0, +\infty]$ borélienne, on a

$$\int_{A_1 \cup A_2} f d\lambda = \int_{A_1} f d\lambda + \int_{A_2} f d\lambda \ .$$

Remarque 1.14. Pour intégrer une fonction à valeurs positives sur une partie A, il suffit qu'elle soit au moins définie sur A : si $f \colon X \subseteq \mathbb{R} \to \mathbb{R}$ est une fonction borélienne et A une partie borélienne de X, on définit $\int_A f d\lambda = \int_A g d\lambda$ où g est un (n'importe quel) prolongement borélien de f à $\mathbb{R}$ (par exemple g définie par $g(x) = f(x)$ pour $x \in X$ et $g(x) = 0$ pour $x \in \mathbb{R} \setminus X$).

Attention : on peut définir l'intégrale $\int_{\mathbb{R}} f d\lambda$ pour *toute* fonction borélienne f à valeurs positives ; *mais cette intégrale peut valoir* $+\infty$. Une idée intuitive est que, dans le cas d'une fonction à valeurs positives, l'intégrale de f coïncide avec « l'aire sous la courbe représentative de f » ; et cette aire peut éventuellement être infinie : par exemple, si f est la fonction constante égale à 1 sur $\mathbb{R}$, alors le domaine limité par le graphe de f et l'axe des abscisses est un « rectangle »dont un côté est de longueur 1 est l'autre de longueur infinie, et on voit bien que l'aire d'un tel domaine est infinie. Mais si f change de signe, alors la notion d'« aire sous la courbe »devient plus compliquée à interpréter.

Définition 1.15. Soit $f\colon X\subseteq\mathbb{R}\to\mathbb{R}$ une fonction borélienne. On dit que f est *intégrable* sur une partie A borélienne de X si $\int_A |f|d\lambda < +\infty$.

Dans ce cas,f^+ et f^- sont également intégrables sur A (par la propriété (4)) et on définit $\int_A fd\lambda = \int_A f^+d\lambda - \int_A f^-d\lambda$.

Attention, dans le cadre de la théorie de Lebesgue, on définit $\int_A fd\lambda$ uniquement quand f est *intégrable* sur A (c.à.d. $\int_A |f|d\lambda < +\infty$) et on ne considère pas les intégrales généralisées semi-convergentes qui apparaissent dans la théorie de l'intégrale de Riemann (voir la première feuille de TD).

On retrouve les propriétés (1), (2), (3), (4), (5), (6) ci-dessus qui sont cette fois vraies pour des fonctions boréliennes intégrables.

Proposition 1.16. *Pour toutes fonctions boréliennes $f,g\colon X\subseteq\mathbb{R}\to\mathbb{R}$ intégrables sur une partie borélienne $A\subseteq X$ on a*

1. $\int_A 1d\lambda = \lambda(A)$ *;*
2. *pour toutes constantes réelles α et β, la fonction $\alpha f+\beta g$ est également intégrable sur A et*
$$\int_A(\alpha f+\beta g)d\lambda = \alpha\int_A fd\lambda + \beta\int_A gd\lambda \text{ (linéarité de l'intégrale)};$$
3. *si $f=g$ presque partout, alors $\int_A fd\lambda = \int_A gd\lambda$;*
4. *si $f\geq 0$ presque partout, alors $\int_A fd\lambda \geq 0$* (positivité de l'intégrale) *;*
si $f\leq g$ presque partout, alors $\int_A fd\lambda \leq \int_A gd\lambda$ (monotonie de l'intégrale) *;*
5. *si $\lambda(A)=0$ alors $\int_A fd\lambda = 0$;*
6. *si A_1, A_2 sont des parties boréliennes de A telles que $\lambda(A_1\cap A_2)=0$, alors*
$$\int_{A_1\cup A_2} fd\lambda = \int_{A_1} fd\lambda + \int_{A_2} fd\lambda\ .$$

On a de plus

7. L'inégalité triangulaire
$$\left|\int_A (f+g)d\lambda\right| \leq \int_A |f|d\lambda + \int_A |g|d\lambda\ .$$

1.3 Théorèmes d'échanges limite-intégrale

La théorie de Lebesgue nous permet d'intégrer plus de fonctions que la théorie de Riemann. Cela simplifie l'énoncé des théorèmes d'échange limite/intégrale, qui sont plus puissants dans ce contexte.

Théorème 1.17 (Théorème de convergence monotone)**.** *Soit A une partie de $\mathbb{R}$ borélienne et $f_i\colon A\to[0,+\infty[$ une suite de fonctions boréliennes telle pour presque tout $x\in A$ la suite $(f_i(x))$ soit une suite croissante. Alors il existe une fonction borélienne $f\colon A\to[0,+\infty]$ telle que $f_i(x)$ converge vers $f(x)$ presque partout, et on a*
$$\lim_{i\to+\infty}\int_A f_id\lambda = \int_A fd\lambda\ .$$

Remarque 1.18. Comme, pour presque tout x, la suite $(f_i(x))$ est croissante, il est immédiat que cette suite admet une limite dans $[0,+\infty]$. Ainsi, l'existence de la fonction f est une conséquence immédiate de l'hypothèse de monotonie. Ce qui n'est pas trivial, c'est de pouvoir échanger limite et intégrale !

Attention : l'hypothèse du théorème *n'est pas* que les fonctions f_i sont des fonctions croissantes ; c'est que, pour presque tout x *fixé*, la suite $(f_i(x))$ est une suite croissante.

Remarque 1.19. Comme on ne travaille dans cette partie qu'avec la mesure de Lebesgue, on n'a écrit ce théorème que pour la mesure de Lebesgue ; mais il serait vrai aussi pour toute autre mesure (en remplaçant « borélien »par « mesurable ») ; il en va de même de la définition de l'intégrale qu'on a vue plus haut, et du théorème de convergence dominée qu'on va maintenant présenter.

L'avantage de ce théorème est qu'il s'applique même si la suite des intégrales tend vers $+\infty$; mais l'hypothèse demandant que les fonctions soient à valeurs positives, et que la suite $(f_i(x))$ soit croissante pour presque tout x, n'est souvent pas vérifiée dans les exemples. Néanmoins, à partir de ce théorème on peut déduire un autre théorème fondamental.

Théorème 1.20 (Théorème de convergence dominée)**.** *Soit A une partie de $\mathbb{R}$ borélienne, et $f_i \colon A \to \mathbb{R}$ une suite de fonctions boréliennes et satisfaisant les hypothèses suivantes :*

1. *Pour presque tout x, la suite $f_i(x)$ est convergente vers une limite qu'on appelle $f(x)$ (alors f est borélienne).*
2. *Il existe une fonction $g \colon A \to \mathbb{R}$ borélienne à valeurs positives, telle que pour presque tout x on ait $|f_i(x)| \leq g(x)$ et $\int_A g d\lambda < +\infty$.*

Alors f est intégrable, et on a

$$\int_A f d\lambda = \lim_{i \to +\infty} \int_A f_i d\lambda \ .$$

Remarque 1.21. Dans l'énoncé du théorème ci-dessus, il est fondamental que *la fonction g ne dépend pas de i.*

En appliquant les théorèmes de convergence monotone et de convergence dominée sur les sommes partielles d'une suite de fonctions, on obtient le corollaire suivant.

Corollaire 1.22 (Échanges série/intégrale)**.** Soit A une partie de $\mathbb{R}$ borélienne.

1. Si $(f_n)_{n \geq 0}$ est une suite de fonctions positives boréliennes définies sur A, alors

$$\int_A \sum_{n=0}^{+\infty} f_n \ d\lambda = \sum_{n=0}^{+\infty} \int_A f_n d\lambda \qquad \text{(convergence monotone).}$$

2. Si $(f_n)_{n \geq 0}$ est une suite de fonctions boréliennes définies sur A telle que

$$\sum_{n=0}^{+\infty} \int_A |f_n| d\lambda < +\infty$$

alors la série $\displaystyle\sum_{n=0}^{+\infty} f_n(x)$ converge absolument pour presque tout $x \in A$ vers une fonction f intégrable, et on a

$$\int_A f d\lambda = \sum_{n=0}^{+\infty} \int_A f_n d\lambda \qquad \text{(convergence dominée).}$$

Remarque 1.23. La théorie de l'intégrale de Cauchy-Riemann est basée sur la stratégie suivante : d'abord, on définit l'intégrale des fonctions en escalier (qui coïncide avec leur intégrale pour la mesure de Lebesgue). Ensuite, on utilise le fait que toute fonction continue par morceaux f définie sur un segment I est une limite uniforme de fonctions en escalier f_i, et on définit $\displaystyle\int_I f(x)dx = \lim_{i \to +\infty} \int_I f_i(x)dx$.

Comme toute suite de fonctions en escalier convergeant uniformément vers une fonction continue par morceaux sur un segment est bornée par une constante M, et que les fonctions constantes sont intégrables sur les segments, on peut appliquer le théorème de convergence dominée pour conclure que $\displaystyle\int_I f d\lambda = \lim \int_I f_i d\lambda$. Ainsi, on déduit que, pour une fonction continue par morceaux, son intégrale au sens de Riemann et son intégrale au sens de Lebesgue coïncident.

En utilisant le théorème de convergence monotone, on en déduit que l'intégrale généralisée au sens de Riemann d'une application continue positive sur un intervalle ouvert est égale à son intégrale au sens de Lebesgue. Enfin, une fonction continue qui a une intégrale généralisée absolument convergente au sens de Riemann, est également intégrable au sens de Lebesgue et à nouveau par le théorème de convergence dominée, ces intégrales coïncident. Plus précisément :

Théorème 1.24 (Comparaison avec l'intégrale de Riemann)**.** *Les intégrales au sens de Riemann et au sens de Lebesgue coïncident (en particulier) dans les situations suivantes :*

1. *Si $-\infty < a < b < +\infty$ et $f : [a,b] \to \mathbb{R}$ est une fonction continue par morceaux, alors f est intégrable au sens de Riemann et au sens de Lebesgue, et*
$$\int_a^b f(x)dx = \int_{[a,b]} f d\lambda.$$
2. *Si $-\infty \leq a < b \leq +\infty$ et $f :]a,b[\to [0,+\infty[$ est une fonction continue par morceaux, alors*
$$\int_a^b f(x)dx = \int_{]a,b[} f d\lambda.$$
(Rappelons que ces intégrales peuvent valoir $+\infty$.)
3. *Si $-\infty \leq a < b \leq +\infty$ et $f :]a,b[\to \mathbb{R}$ est une fonction continue par morceaux dont l'intégrale généralisée au sens de Riemann est absolument convergente, c'est-à-dire telle que $\displaystyle\int_a^b |f(x)|dx < +\infty$, alors f est intégrable au sens de Lebesgue et*
$$\int_a^b f(x)dx = \int_{]a,b[} f d\lambda.$$

1.4 Continuité et dérivabilité des intégrales à paramètres

Dans cette section, on fixe une partie X borélienne de $\mathbb{R}$, et I un intervalle. Étant donnée une fonction $f\colon X \times I \to \mathbb{R}$ telle que pour tout $t \in I$ $x \mapsto f(x,t)$ soit intégrable sur X, on peut définir la fonction $F\colon t \mapsto \int_X f(x,t)d\lambda(x)$. On s'intéresse à des hypothèses sur f permettant de conclure que F est continue, ou dérivable.

Théorème 1.25 (Théorème de continuité des intégrales à paramètres)**.** *Soit $f\colon X \times I \to \mathbb{R}$ une fonction telle que*

1. *Pour presque tout $x \in X$, la fonction $t \mapsto f(x,t)$ est continue sur I.*
2. *Il existe une fonction $g\colon X \to [0,+\infty[$ intégrable et telle que pour tout $t \in I$ on ait $|f(x,t)| \leq g(x)$ pour presque tout $x \in X$.*
3. *Pour tout t la fonction $x \mapsto f(x,t)$ est borélienne.*

Alors la fonction $F\colon t \mapsto \displaystyle\int_X f(x,t)d\lambda(x)$ est bien définie et continue sur I.

Remarque 1.26. Dans la deuxième condition ci-dessus, *la fonction g ne dépend pas de t*. Dans les applications, il n'est pas toujours facile de trouver une fonction g satisfaisant les conditions du théorème ! La première condition est, elle, normalement facile à vérifier.

Preuve:

Déjà, notons que, pour tout t fixé, la fonction $x \mapsto f(x,t)$ est borélienne et telle que $|f(x,t)| \leq g(x)$ pour presque tout $x \in X$; comme g est intégrable, on en déduit que $\int_X |f(x,t)|d\lambda(x) \leq \int_X g(x)d\lambda(x) < +\infty$, donc la fonction $x \mapsto f(x,t)$ est intégrable, et F est donc bien définie.

Fixons maintenant $t \in I$. Pour montrer que F est continue en t, on doit vérifier que, pour toute suite (t_n) d'éléments de I qui converge vers t, la suite $F(t_n)$ converge vers $F(t)$. Fixons donc une telle suite (t_n). Par définition, on a $F(t_n) = \displaystyle\int_X f(x,t_n)d\lambda(x)$.

La première hypothèse ci-dessus nous permet d'affirmer que la suite de fonctions $f_n\colon x \mapsto f(x,t_n)$ est une suite de fonctions boréliennes et telle que $f_n(x)$ converge vers $f(x,t)$ pour presque tout $x \in X$; et on a aussi $|f_n(x)| \leq g(x)$ pour presque tout x. Comme g est intégrable, on peut appliquer le théorème de convergence dominée pour conclure que
$$\lim_{n\to+\infty} \int_X f_n(x)d\lambda(x) = \int_X f(x,t)d\lambda(x) \ .$$
On vient de montrer que $F(t_n) = \int_X f_n(x)d\lambda(x)$ converge vers $F(t)$, ce qui prouve que F est bien continue en t.

Remarque 1.27. En regardant la preuve, on voit que si, dans les hypothèses du théorème, on avait remplacé « pour presque tout $x \in X$, la fonction $t \mapsto f(x,t)$ est continue sur I »par « pour presque tout $x \in X$, la fonction $t \mapsto f(x,t)$ est continue en un certain $t_0 \in I$ »(en laissant les autres hypothèses inchangées), alors on aurait pu conclure que F est continue en t_0.

Par ailleurs, notons que si X et I sont des segments, et f est une fonction continue, alors f est bornée, c'est-à-dire qu'il existe M tel que $|f(x,t)| \leq M$ pour tout $(x,t) \in X \times I$. Comme les fonctions constantes sont intégrables sur les segments, on voit qu'on peut poser $g(x) = M$ et appliquer le théorème de continuité : si X et I sont des segments, et f est une fonction continue des deux variables (x,t), alors F est continue. Ce corollaire du théorème de continuité peut se démontrer simplement sans utiliser le théorème de convergence dominée.

Théorème 1.28 (Théorème de dérivabilité des intégrales à paramètres)**.** *Soit $f\colon X \to I \to \mathbb{R}$ une fonction telle que*

1. *Pour tout $t \in I$, la fonction $x \mapsto f(x,t)$ est intégrable sur X.*
2. *Pour presque tout $x \in X$, la fonction $t \mapsto f(x,t)$ est dérivable sur I (on note sa dérivée $\frac{\partial f}{\partial t}(x,t)$).*
3. *Il existe une fonction $g\colon X \to [0,+\infty[$ intégrable et telle que pour tout $t \in I$ et tout $x \in X$ on ait $\left|\frac{\partial f}{\partial t}(x,t)\right| \leq g(x)$.*
4. *Pour tout $t \in I$ la fonction $x \mapsto \frac{\partial f}{\partial t}(x,t)$ est borélienne.*

Alors la fonction $F\colon t \mapsto \int_X f(x,t)d\lambda(x)$ est bien définie sur I, dérivable, et on a pour tout $t \in I$ que

$$F'(t) = \int_X \frac{\partial f}{\partial t}(x,t)d\lambda(x) \ .$$

Remarque 1.29. En fait, le théorème reste vrai si l'on remplace la première hypothèse par le fait qu'il existe un $t \in I$ tel que $x \mapsto f(x,t)$ est intégrable sur I (bien sûr, on doit toujours supposer que cette fonction est borélienne pour tout $t \in I$).

Remarque 1.30. Si on ajoute l'hypothèse selon laquelle la fonction $t \mapsto \frac{\partial f}{\partial t}(x,t)$ est continue sur I pour presque tout $x \in X$, alors le théorème de continuité des intégrales à paramètres nous permet de conclure que F' est continue sur I, c'est-à-dire que F est de classe $\mathcal{C}^1$ sur I.

Preuve:

Fixons $t \in I$. Pour vérifier que F est dérivable en t, on doit étudier le taux d'accroissement de F en t ; fixons donc une suite (t_n) d'éléments de I différents de t, et considérons le taux d'accroissement

$$\frac{F(t_n) - F(t)}{t_n - t} = \int_X \frac{f(x,t_n) - f(x,t)}{t_n - t} d\lambda(x) \ .$$

Comme dans la preuve du théorème de continuité, considérons la suite de fonctions $f_n\colon x \mapsto \frac{f(x,t_n) - f(x,t)}{t_n - t}$. Le fait que $t \mapsto f(x,t)$ soit dérivable en t pour presque tout $x \in X$ nous dit que $f_n(x)$ converge vers $\frac{\partial f}{\partial t}(x,t)$ pour presque tout $x \in X$. De plus, l'inégalité des accroissements finis nous dit que, pour presque tout $x \in X$, on a

$$\left|\frac{f(x,t_n) - f(x,t)}{t_n - t}\right| \leq \sup_{s \in [t_n,t]} \frac{\partial f}{\partial t}(x,s) \leq g(x) \ .$$

On peut donc appliquer le théorème de convergence dominée à la suite f_n pour conclure que

$$\lim_{n \to +\infty} \int_X f_n(x)d\lambda(x) = \int_X \lim_{n \to +\infty} f_n(x)d\lambda(x) \ .$$

Autrement dit, on a montré que $\frac{F(t_n) - F(t)}{t_n - t}$ converge vers $\int_X \frac{\partial f}{\partial t}(x,t)d\lambda(x)$ quand n tend vers $+\infty$, c'est-à-dire que F est dérivable en t et $F'(t) = \int_X \lim_{n \to +\infty} f_n(x)d\lambda(x)$.

Chapitre 2

Intégrale de fonctions de plusieurs variables réelles

2.1 La mesure de Lebesgue sur $\mathbb{R}^n$

De même que la mesure de Lebesgue sur $\mathbb{R}$ généralise la notion de longueur, on peut généraliser la notion d'aire (dans $\mathbb{R}^2$), de volume (dans $\mathbb{R}^3$)...

Théorème 2.1. *Soit $n \geq 1$ un entier. Il existe une unique mesure sur $\mathbb{R}^n$ -ou plutôt, sur la tribu borélienne de $\mathbb{R}^n$), notée λ_n, telle que*

$$\lambda_n([a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]) = \prod_{i=1}^{n} (b_i - a_i) \ .$$

On appelle cette mesure la mesure de Lebesgue sur $\mathbb{R}^n$.

Quand $n = 2$, la formule ci-dessus nous dit que la mesure d'un rectangle (dont les côtés sont parallèles aux axes de coordonnées) est égale au produit des longueurs de ses côtés, c'est-à-dire à son aire ; quand $n = 3$ on retrouve la formule pour le volume d'un pavé, etc.

Exactement de la même façon que pour la définition de l'intégrale de Lebesgue sur $\mathbb{R}$ (ou, plus généralement, de l'intégrale par rapport à une mesure quelconque), on peut définir tout d'abord l'intégrale pour les fonctions boréliennes et à valeurs dans $[0, +\infty]$, et ensuite pour les fonctions $f : \mathbb{R}^n \to \mathbb{R}$ boréliennes et telles que $\int_{\mathbb{R}^n} |f| d\lambda_n < +\infty$. Cette notion d'intégrale a les mêmes propriétés que celles qu'on a énoncées au chapitre précédent : positivité, linéarité, inégalité triangulaire... Elle est définie en posant tout d'abord $\displaystyle\int_{\mathbb{R}^n} \mathbf{1}_A d\lambda_n = \lambda_n(A)$, si A est une partie borélienne de $\mathbb{R}^n$ et $\mathbf{1}_A$ est sa fonction caractéristique ; puis en étendant par linéarité aux combinaisons linéaires de fonctions caractéristiques, puis par un procédé de passage à la limite pour définir l'intégrale de toutes les fonctions boréliennes et à valeurs positives. Comme dans le cas de la mesure de Lebesgue sur $\mathbb{R}$, on dit que f est *intégrable* si f est borélienne et $\displaystyle\int_{\mathbb{R}^n} |f| d\lambda_n$ est finie.

On voit donc que, exactement comme dans le cas de la mesure de Lebesgue sur $\mathbb{R}$, les fonctions boréliennes et à valeurs positives jouent un rôle particulier ; on peut toujours calculer leur intégrale, et celle-ci peut valoir $+\infty$.

2.2 Théorèmes de Fubini-Tonelli et Fubini

La méthode de base pour calculer une intégrale d'une fonction de n variables est de se ramener à n calculs successifs d'intégrales de fonctions de 1 variable.

Théorème 2.2 (Théorème de Fubini-Tonelli)**.** *Soit m, p deux entiers et $f : \mathbb{R}^m \times \mathbb{R}^p \to [0, +\infty]$ une fonction* *borélienne. Alors on a*

$$\begin{aligned}\int_{\mathbb{R}^{m+p}} f d\lambda_{m+p} &= \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^p} f(x,y) d\lambda_p(y) \right) d\lambda_m(x) \\ &= \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^m} f(x,y) d\lambda_m(x) \right) d\lambda_p(y)\end{aligned}$$

En pratique, dans les exercices, m et p vaudront le plus souvent 1 ou 2.

Remarque 2.3. L'énoncé ci-dessus sous-entend que les fonctions apparaissant dans les intégrales (par exemple, $x \mapsto \int_{\mathbb{R}^p} f(x,y) d\lambda_p(y)$) sont boréliennes...

Remarque 2.4. Comme dans le cas de $\mathbb{R}$, intégrer une fonction de n variables sur une partie borélienne A de $\mathbb{R}^n$ revient à intégrer la fonction $f \cdot \mathbf{1}_A$ sur $\mathbb{R}^n$, donc l'énoncé ci-dessus peut être utilisé pour calculer les intégrales de fonctions boréliennes définies sur des sous-parties boréliennes de $\mathbb{R}^n$.

Exemple. Calculons l'aire du disque unité $D = \{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 \leq 1\}$.

Par définition, l'aire d'une partie D est l'intégrale de la fonction caractéristique de D sur $\mathbb{R}^2$; on a donc

$$\begin{aligned}\text{aire}(D) &= \int_{\mathbb{R}^2} \mathbf{1}_D(x,y) d\lambda_2(x,y) \\ &= \int_{x=-1}^{1} \left(\int_{y=-\sqrt{1-x^2}}^{\sqrt{1-x^2}} d\lambda(y) \right) d\lambda(x) \\ &= \int_{-1}^{1} 2\sqrt{1-x^2} dx \\ &= 2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2(t) dt \\ &= \pi\end{aligned}$$

On voit que, dans le calcul ci-dessus, on n'a pas pris en compte les points en dehors de D, puisque $\mathbf{1}_D$ vaut 0 en ces points : en pratique, on détermine dans quel domaine varie x (ici, $[-1,1]$) puis, *à x fixé*, dans quel domaine varie y, et on calcule les intégrales correspondantes.

Bien sûr, on aurait pu aussi faire le calcul « dans l'autre sens », en intégrant d'abord par rapport à x puis par rapport à y ; ici ça ne changerait rien (la fonction intégrée et le domaine sont symétriques en x, y) mais dans d'autres cas un des calculs peut être beaucoup plus facile que l'autre.

Dans le cas où f n'est pas à valeurs positives, on a d'abord besoin de s'assurer que $|f|$ est intégrable, ce qui peut être fait en appliquant le théorème de Fubini-Tonelli, ou en utilisant le fait qu'une fonction *continue* sur un *compact* est toujours intégrable (mais dans ce cas il faut bien vérifier que la fonction f est continue, et surtout que le domaine d'intégration est compact...).

Théorème 2.5 (Théorème de Fubini)**.** *Soit $f \colon \mathbb{R}^{m+p} \to \mathbb{R}$ une fonction borélienne telle que $|f|$ est intégrable. Alors on a :*

$$\begin{aligned}\int_{\mathbb{R}^{m+p}} f d\lambda_{m+p} &= \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^p} f(x,y) d\lambda_p(y) \right) d\lambda_m(x) \\ &= \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^m} f(x,y) d\lambda_m(x) \right) d\lambda_p(y)\end{aligned}$$

Remarque 2.6. Comme dans le cas du théorème de Fubini-Tonelli, il est sous-entendu que les fonctions apparaissant dans l'intégrable sont boréliennes (en fait, elles ne sont définies que presque partout, mais ça n'affecte pas la valeur de l'intégrale ; on reverra cela plus en détail quand on énoncera le théorème de Fubini pour des mesures plus générales)

Pour essayer de se convaincre que le théorème de Fubini est « raisonnable », donnons-en une preuve dans le cas où on intègre une fonction *continue* f sur un rectangle $[a,b]\times[c,d]$. Notons d'abord que sous ces hypothèses $|f|$ est bien intégrable puisqu'elle est bornée et que le domaine d'intégration est de mesure finie. On veut montrer que :

$$\int_a^b \left(\int_c^d f(x,y)dy\right) dx = \int_c^d \left(\int_a^b f(x,y)dx\right) dy \ .$$

Pour cela, on introduit deux fonctions $F, G\colon [c,d]\to\mathbb{R}$ en posant

$$F(t) = \int_a^b \left(\int_c^t f(x,y)dy\right) dx \quad \text{et} \quad G(t) = \int_c^t \left(\int_a^b f(x,y)dx\right) dt \ .$$

Commençons par étudier F ; elle est de la forme $F(t) = \displaystyle\int_a^b h(x,t)dt$, avec $h(x,t) = \displaystyle\int_c^t f(x,y)dy$. A x fixé, la fonction $t\mapsto h(x,t) = \displaystyle\int_c^t f(x,y)dy$ est dérivable, de dérivée $f(x,t)$; de plus cette dérivée est bornée. On peut donc appliquer le théorème de dérivation des intégrales à paramètres pour conclure que F est dérivable sur $[c,d]$, de dérivée $F'(t) = \displaystyle\int_a^b f(x,t)dx$.

Passons à G : le théorème de continuité des intégrales à paramètres nous permet de voir que $g\colon y\mapsto \displaystyle\int_a^b f(x,y)dx$ est une fonction continue sur $[c,d]$. Puisque $G(t) = \displaystyle\int_C^t g(y)dy$, on voit que G est dérivable et que $G'(t) = g(t) = \displaystyle\int_a^b f(x,t)dx$.

On voit donc que F et G sont dérivables, et que $F' = G'$ sur $[c,d]$; de plus on a $F(c) = G(c) = 0$. On en conclut que F et G sont égales sur $[c,d]$, en particulier $F(d) = G(d)$, ce qu'on voulait démontrer.

2.3 Théorème de changement de variables

En pratique, pour calculer une intégrale multiple, on est souvent amené à faire un changement de variables pour se ramener à un domaine plus simple sur lequel appliquer le théorème de Fubini. Toutes les fonctions ne sont pas acceptables pour ce théorème : les fonctions que l'on peut utiliser pour faire un changement de variable sont les *difféomorphismes de classe* $\mathcal{C}^1$.

Définition 2.7. Étant donné U un ouvert de $\mathbb{R}^n$, et $\varphi\colon U\to\mathbb{R}^n$ une fonction de classe $\mathcal{C}^1$, on note $J\varphi(x)$ la matrice jacobienne de φ en un point $x\in U$; rappelons qu'il s'agit de la matrice dont le coefficient sur la i-ième ligne et la j-ième colonne vaut $\dfrac{\partial\varphi_i}{\partial x_j}(x)$.

Définition 2.8. Soit U, V deux ouverts de $\mathbb{R}^n$. Une fonction $\varphi\colon U\to V$ est un difféomorphisme de classe $\mathcal{C}^1$ si :

1. φ est une bijection de U sur V.
2. φ est de classe $\mathcal{C}^1$ sur U, c'est-à-dire que chaque dérivée $\dfrac{\partial\varphi_i}{\partial x_j}$ existe et est continue sur U.
3. La bijection réciproque φ^{-1} est également de classe $\mathcal{C}^1$.

Remarque 2.9. Comme $\varphi^{-1}\circ\varphi(x) = x$ pour tout $x\in U$ par définition, on obtient en appliquant la règle de la chaîne que $J\varphi^{-1}(\varphi(x))J\varphi(x) = I_n$ (la matrice identité) pour tout $x\in U$; en particulier $J\varphi(x)$ doit être inversible pour tout $x\in U$ si φ est un difféomorphisme (ce qui revient à dire que son déterminant est non nul).

Un exemple fondamental de difféomorphisme de classe $\mathcal{C}^1$ de $\mathbb{R}^n$ dans $\mathbb{R}^n$ est donné par une application linéaire et inversible ; rappelons que si φ est linéaire, sa matrice jacobienne est simplement la matrice de φ dans la base canonique de $\mathbb{R}^n$.

Souvent, pour vérifier qu'une application est un difféomorphisme de classe $\mathcal{C}^1$, on utilise la caractérisation suivante.

Théorème 2.10 (Théorème d'inversion globale)**.** *Soit U un ouvert de $\mathbb{R}^n$, et $\varphi\colon U \to \mathbb{R}^n$ une application de classe $\mathcal{C}^1$. Si φ est injective sur U et $\det(J\varphi(x)) \neq 0$ pour tout $x \in U$, alors $\varphi(U)$ est un ouvert de $\mathbb{R}^n$ et φ est un difféomorphisme de classe $\mathcal{C}^1$ de U sur $\varphi(U)$.*

Remarque 2.11. Toute la difficulté du théorème précédent est dans la démonstration que $\varphi(U)$ est ouvert. Si on suppose que $\varphi(U)$ est ouvert dans l'énoncé ci-dessus, alors le fait que φ soit un difféomorphisme vient simplement du fait qu'on sait calculer « explicitement »l'inverse d'une matrice inversible (via le lien entre inverse et comatrice).

Nous pouvons maintenant énoncer le théorème de changement de variables.

Théorème 2.12 (Théorème de changement de variables)**.** *Soit U, V deux ouverts de $\mathbb{R}^n$, et $\varphi\colon U \to V$ un difféomorphisme de classe $\mathcal{C}^1$. Rappelons qu'on note λ_n la mesure de Lebesgue sur $\mathbb{R}^n$. Alors on a :*

1. *Pour toute partie B borélienne de U, $\lambda_n(\varphi(B)) = \int_B |\det(J\varphi(x))| d\lambda_n(x)$.*
2. *Si $f\colon V \to [0, +\infty]$ est borélienne, alors*
$$\int_V f(x) d\lambda_n(x) = \int_U f \circ \varphi(y) |\det(J\varphi(y))| d\lambda_n(y) \ .$$
3. *Si $f\colon V \to \mathbb{R}$ est intégrable, alors $y \mapsto f \circ \varphi(y)|\det(J\varphi(y))|$ est intégrable sur U et on a*
$$\int_V f(x) d\lambda_n(x) = \int_U f \circ \varphi(y) |\det(J\varphi(y))| d\lambda_n(y) \ .$$

La définition de l'intégrale fait que (3) est une conséquence immédiate de (2) ; et (1) est un cas particulier de (2) appliqué à la fonction caractéristique de $\varphi(B)$. La définition fait aussi qu'il n'est pas difficile de déduire (2) une fois qu'on a démontré (1).

Ce théorème est difficile à démontrer, et on ne va pas essayer de donner une idée de la preuve dans le cas où φ n'est pas une application linéaire. Discutons un peu la preuve de (1) dans le cas où φ est une application linéaire. Étudions quelques cas particuliers :

1. La matrice M de φ dans la base canonique est la matrice diagonale $\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & t & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix}$. Alors on a $\varphi([0,1]^n) = [0,1] \times [0,1] \times \ldots [0,t] \times [0,1] \times \ \ldots [0,1]$ si $t > 0$, et $[0,1] \times [0,1] \times \ldots [t,0] \times [0,1] \times \ \ldots [0,1]$ si $t < 0$. Dans les deux cas, on a par définition de la mesure de Lebesgue que
$$\lambda_n(\varphi([0,1]^n) = |t| = |\det(M)|\lambda_n([0,1]^n) \ .$$
2. La matrice M est la matrice par blocs $\begin{pmatrix} N & \\ & I_{n-2} \end{pmatrix}$, où $N = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Alors on a $\varphi([0,1]^n) = D \times [0,1]^{n-2}$, où D est le parallélogramme de sommets $(0,0)$, $(0,1)$, (1.1) et $(1,2)$ (faites un dessin !). Par conséquent $\lambda_n(\varphi([0,1]^n) = \text{aire}(D) = 1$, et $\det(M) = \det(N) = 1$. Donc dans ce cas on a bien comme espéré que $\lambda_n(\varphi([0,1]^n) = |\det(M)|\lambda_n(\varphi([0,1]^n)$.
3. La matrice M est une *matrice de permutation*, c'est-à-dire que φ est une application linéaire permutant les vecteurs de base. Alors la déterminant de M vaut ± 1 (ce dont on peut se convaincre par récurrence, en développant le déterminant de M par rapport à la première ligne par exemple, et en utilisant que M a exactement un 1 sur chaque ligne et chaque colonne, et des 0 ailleurs ; ou bien en revenant à la formule définissant le déterminant), et $\varphi([0,1]^n) = [0,1]^n$. On a donc encore $\lambda_n(\varphi([0,1]^n) = |\det(M)|\lambda_n(\varphi([0,1]^n)$

Pour l'instant, on a le résultat désiré pour des matrices de trois types, et $B = [0,1]^n$; par pavage, on peut se convaincre qu'on a alors le résultat désiré pour une matrice d'un de ces trois types et B n'importe quelle partie borélienne de $\mathbb{R}$. Ensuite, un théorème d'algèbre linéaire nous dit que l'on peut écrire toute matrice inversible

M sous la forme $M = M_1 M_2 \dots M_p$, où chaque matrice M_i est d'un des trois types ci-dessus. On peut donc écrire, si φ est une bijection linéaire de matrice M et B une partie borélienne de $\mathbb{R}^n$:

$$\begin{aligned}
\lambda_n(\varphi(B)) &= \lambda_n(M_1 \dots M_p(B)) \\
&= |\det(M_p)| \lambda_n(M_1 \dots M_{p-1}(B)) \\
&= |\det(M_p)| \dots |\det(M_1)| \lambda_n(B) \\
&= |\det(M)| \lambda_n(B).
\end{aligned}$$

Exemple (changement de variables en polaires)**.** On considère l'application $\phi : U =]0, +\infty[\times]0, 2\pi[\to \mathbb{R}^2$ définie par $\phi(r, \theta) = (r\cos\theta, r\sin\theta)$.

Alors, la matrice jacobienne de ϕ est $\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$, de déterminant r.

De plus, ϕ est injective et $\phi(U) = \mathbb{R}^2 \setminus [0, +\infty[= V$.

Ainsi, ϕ est un $\mathcal{C}^1$-difféomorphisme de U sur V. Comme $\lambda_2(\mathbb{R}^2 \setminus V) = 0$, c'est-à-dire $\mathbb{R}^2 \setminus V$ est négligeable, il n'est pas gênant que ϕ ne soit pas un difféomorphisme de U sur $\mathbb{R}^2$ tout entier.

Par exemple, calculons

$$I = \int_D (x + y)^2 dxdy, \text{ où } D = \{(x, y) \colon x^2 + y^2 < 1\}.$$

En utilisant le théorème de changement de variables avec les coordonnées polaires (et le théorème de Fubini), on obtient $\phi^{-1}(D) =]0, 1[\times]0, 2\pi[$ et

$$\begin{aligned}
I &= \int_{\phi^{-1}(D)} (r\cos\theta + r\sin\theta)^2 r dr d\theta \\
&= \int_0^1 \left(\int_0^{2\pi} r^3 (\cos^2\theta + \sin^2\theta + 2\cos\theta\sin\theta) d\theta \right) dr \\
&= \int_0^1 r^3 \left(\int_0^{2\pi} (1 + \sin 2\theta) d\theta \right) dr \\
&= \int_0^1 2\pi r^3 dr \\
&= \frac{\pi}{2}.
\end{aligned}$$

Chapitre 3

Tribus et mesures

On va maintenant prendre en compte les questions de *mesurabilité* qui ont été éludées précédemment. L'idée est que l'on n'a pas vraiment besoin de savoir mesurer *toutes* les parties d'un ensemble X ; en fait, c'est même souvent impossible. La définition ci-dessous regroupe les propriétés que doit avoir l'ensemble des parties que l'on sait mesurer.

Définition 3.1. Soit X un ensemble, et $\mathcal{A}$ une famille de parties de X. On dit que $\mathcal{A}$ est une *tribu* si $\mathcal{A}$ satisfait les conditions suivantes :

1. $\emptyset \in \mathcal{A}$ et $X \in \mathcal{A}$.
2. Si $A \in \mathcal{A}$ alors $X \setminus A \in \mathcal{A}$.
3. Si $(A_i)_{i\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$ alors $\bigcup_i A_i \in \mathcal{A}$.

On dit alors que $(X, \mathcal{A})$ est un *espace mesurable*

Par exemple, la famille $\mathcal{P}(X)$ formée par toutes les parties de X est une tribu. La famille $\{\emptyset, X\}$ n'ayant que $\emptyset$ et X comme éléments en est une aussi. Mais il y a bien d'autres exemples ; décrivons-en un.

Exemple. Soit X un ensemble, et A une partie de X. Alors

$$\mathcal{A} = \{B \subseteq X \colon B \subseteq A \text{ ou } B \subseteq X \setminus A\}$$

est une tribu.

Notons que, si $\mathcal{A}$ est une tribu, alors $\mathcal{A}$ est stable par union finie : si $A_0, \dots, A_n \in \mathcal{A}$, alors la suite obtenue en posant $A_i = \emptyset$ pour $i \geq n+1$ est une suite d'éléments de $\mathcal{A}$, et

$$\bigcup_{i=0}^{n} A_i = \bigcup_{i=0}^{+\infty} A_i \in \mathcal{A}.$$

On voit aussi que, si (A_i) est une suite de parties de $\mathcal{A}$, alors $\bigcap_{i\in\mathbb{N}} A_i \in \mathcal{A}$: en effet,

$$\bigcap_{i=0}^{+\infty} A_i = X \setminus \bigcup_{i=0}^{+\infty} (X \setminus A_i) \ .$$

Enfin, on voit que si A, B appartiennent à une tribu $\mathcal{A}$, alors $A \setminus B = A \cap (X \setminus B)$ y appartient aussi.

On peut produire beaucoup de tribus grâce à la propriété suivante.

Proposition 3.2. *Soit X un ensemble, et $(\mathcal{A}_i)_{i\in I}$ une famille de tribus sur X. Alors $\mathcal{A} = \bigcap \mathcal{A}_i$ est une tribu.*

En particulier, si $\mathcal{E}$ est un ensemble de parties de X, il existe toujours une plus petite tribu contenant $\mathcal{E}$: l'intersection de toutes les tribus contenant $\mathcal{E}$. On l'appelle *tribu engendrée* par $\mathcal{E}$ et on la note $\sigma(\mathcal{E})$. Si $\mathcal{A}$ est une tribu, et $\mathcal{E} \subseteq \mathcal{A}$ est tel que $\mathcal{A} = \sigma(\mathcal{E})$, on dit que $\mathcal{A}$ est *engendrée par $\mathcal{E}$*.

Évidemment, un espace mesurable est là pour être mesuré !

Définition 3.3. Soit $(X, \mathcal{A})$ un espace mesurable. Une *mesure* sur X est une application $\mu \colon \mathcal{A} \to [0, +\infty]$ vérifiant les propriétés suivantes :

1. $\mu(\emptyset) = 0$.
2. Si $(A_i)_{i\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$ et $A_i \cap A_j = \emptyset$ pour tout $i \neq j$ alors

$$\mu(\bigcup_{i=0}^{+\infty} A_i) = \sum_{i=0}^{+\infty} \mu(A_i) \ .$$

Dans le cas particulier où $\mu(X) = 1$, on dit que μ est une *mesure de probabilité*.

La seconde propriété ci-dessus est appelée *σ-additivité* de la mesure.

Exemple. On peut définir une mesure sur $(X, \mathcal{P}(X))$ en posant $\mu(A) = 0$ pour tout A, ou encore $\mu(\emptyset) = 0$ et $\mu(A) = +\infty$ dès que A est non vide.

Ces mesures ne sont pas très intéressantes ! Un peu plus utile : étant donné $x \in X$ fixé, on peut considérer la *mesure de Dirac* δ_x définie sur $\mathcal{P}(X)$ par :

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

Sur $\mathcal{P}(\mathbb{N})$, on peut aussi définir la *mesure de comptage* μ, en définissant :

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

Et notre mesure de Lebesgue, dans tout ça ? Rappelons que, pour la définir, on a commencé par poser $\lambda(]a,b[) = b - a$ pour tout intervalle ouvert $]a,b[$. Si l'on sait mesurer les intervalles ouverts, on doit pouvoir mesurer les ensemble qui sont la réunion d'une suite d'intervalles ouverts : ce sont tous les ouverts. Mais si on sait mesurer les ouverts, on doit aussi savoir mesurer leurs complémentaires : les fermés. Et alors on doit savoir mesurer les unions de suites de fermés, les complémentaires d'unions de suites de fermés, les unions de complémentaires d'unions de suites de fermés... On se doute que cette énumération pourrait continuer longtemps, et on est alors content de se rendre compte que ce que l'on est en train de dire c'est que, du moment qu'on sait mesurer les intervalles ouverts, on doit savoir mesurer tous les ensembles appartenant à la tribu engendrée par les intervalles ouverts.

Définition 3.4. Soit X un espace topologique. La *tribu borélienne* sur X est la tribu engendrée par les ouverts, c'est-à-dire la plus petite tribu contenant tous les ouverts, et est notée $\mathcal{B}(X)$.

Définition 3.5. La mesure de Lebesgue sur $\mathbb{R}^n$ est l'unique mesure définie sur la tribu borélienne de $\mathbb{R}^n$ et telle que

$$\lambda(\prod_{i=1}^{n} [a_i, b_i]) = \prod_{i=1}^{n} (b_i - a_i) \ .$$

On admettra à la fois qu'une telle mesure existe, et qu'elle est unique !

Avant de passer à l'intégration, regroupons quelques propriétés utiles des mesures.

Proposition 3.6. *Soit $(X, \mathcal{A}, \mu)$ un espace mesuré.*

1. *Si $A, B \in \mathcal{A}$ et $A \subseteq B$ alors $\mu(B) \geq \mu(A)$ (on dit que μ est* croissante*).*
2. *Si $A, B \in \mathcal{A}$, on a toujours $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$; si $\mu(A \cap B) < +\infty$ alors on peut aussi écrire $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.*
3. *Si $(A_i)_{i\in\mathbb{N}}$ sont telles que $A_i \in \mathcal{A}$ pour tout $i \in \mathbb{N}$, alors on a*

$$\mu\left(\bigcup_{i=0}^{+\infty} A_i\right) \leq \sum_{i=0}^{+\infty} \mu(A_i) \ .$$

4. *Si $(A_i)_{i\in\mathbb{N}}$ sont telles que $A_i \in \mathcal{A}$ et $A_i \subseteq A_{i+1}$ pour tout $i \in \mathbb{N}$; alors*

$$\mu\left(\bigcup_{i=0}^{+\infty} A_i\right) = \lim_{i\to+\infty} \mu(A_i) \ .$$

5. *Si $(A_i)_{i\in\mathbb{N}}$ sont telles que $A_i \in \mathcal{A}$ et $A_{i+1} \subseteq A_i$, et $\mu(A_0)$ est finie, alors*

$$\mu\left(\bigcap_{i=0}^{+\infty} A_i\right) = \lim_{i\to+\infty} \mu(A_i) \ .$$

Attention, pour la dernière propriété, il faut supposer que A_0 (ou un des A_i) soit de mesure finie pour que le résultat soit vrai en général : par exemple, dans $\mathbb{R}$ muni de la mesure de Lebesgue, considérons $A_i = [i, +\infty[$. Alors $\mu(A_i) = +\infty$ pour tout i, mais $\bigcap A_i = \emptyset$ et donc $\mu(\bigcap A_i) = 0$.

Preuve:

Pour (1), il suffit d'utiliser le fait que $B \setminus A \in \mathcal{A}$ et d'écrire

$$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A) \text{ puisque } \mu(B \setminus A) \geq \emptyset \ .$$

Pour démontrer (2), on commence par écrire que $A \cup B$ est la réunion disjointe de $A \setminus B$, $A \cap B$, et $B \setminus A$, pour obtenir

$$\begin{aligned} \mu(A \cup B) &= \mu(A \setminus B) + \mu(A \cap B) + \mu(B \setminus A) \\ &= \mu(A) + \mu(B \setminus A) \end{aligned}$$

La deuxième égalité vient du fait que A est la réunion disjointe de $A \setminus B$ et $A \cap B$. En ajoutant $\mu(A \cap B)$, on obtient

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B \setminus A) + \mu(A \cap B) = \mu(A) + \mu(B) \ .$$

Pour démontrer (3), on peut par exemple commmencer par introduire, pour tout $i \in \mathbb{N}$,

$$B_i = A_i \setminus \bigcup_{j<i} A_j \ .$$

Alors chaque B_i appartient à $\mathcal{A}$, les B_i sont deux à deux disjointes et leur réunion est égale à $\bigcup_i A_i$ (pourquoi ?) ; enfin, $B_i \subseteq A_i$ donc $\mu(B_i) \leq \mu(A_i)$ pour tout i. En utilisant la σ-additivité, on obtient :

$$\mu(\bigcup_{i=0}^{++\infty} A_i) = \mu(\bigcup_{i=0}^{+\infty} B_i) = \sum_{i=0}^{+\infty} \mu(B_i) \leq \sum_{i=0}^{+\infty} \mu(B_i) \ .$$

Pour (4), on procède de la même façon, en introduisant les mêmes B_i que ci-dessus et en utilisant que, comme la suite (A_i) est croissante, on a pour tout k

$$A_k = \bigcup_{i=0}^{k} A_i = \bigcup_{i=0}^{k} B_i \ .$$

Mais alors on a aussi, pour tout k :

$$\mu(A_k) = \sum_{i=0}^{k} \mu(B_i)$$

Et alors, quand k tend vers $+\infty$, on voit que $\mu(A_k)$ tend vers

$$\sum_{k=0}^{+\infty} \mu(B_i) = \mu\left(\bigcup_{i=0}^{+\infty} B_i\right) = \mu\left(\bigcup_{i=0}^{+\infty} A_i\right) \ .$$

Enfin, pour montrer (5), on peut par exemple poser $C_i = A_0 \setminus A_i$, pour tout i ; alors $B_i \in \mathcal{A}$, la suite B_i est croissante, $\mu(B_i) = \mu(A_0) - \mu(A_i)$ pour tout i, et

$$\bigcup_{i=0}^{+\infty} B_i = A_0 \setminus \bigcap_{i=0}^{+\infty} A_i \ .$$

De plus, le résultat précédent nous dit que

$$\lim_{i\to+\infty}\mu(B_i)=\mu(\bigcup_{i=0}^{+\infty}B_i)\ .$$

Autrement dit,

$$\lim_{i\to+\infty}(\mu(A_0)-\mu(A_i))=\mu(A_0)-\mu(\bigcap_{i=0}^{+\infty}A_i)\ .$$

Etant donnée une mesure, on a à disposition une notion de « grosseur » : plus la mesure est grande, plus la partie est grosse. Si la mesure est nulle, alors la partie est même *négligeable* : du point de vue de la mesure, elle n'est pas différente de l'ensemble vide. Bien sûr, différentes mesures donnent lieu à différentes notions de grosseur : la mesure de Dirac δ_x pense qu'une partie est négligeable si elle ne contient pas x, et aussi grosse que l'espace entier si elle contient x ; la mesure de comptage prend en compte le nombre d'éléments d'un ensemble ; et la mesure de Lebesgue généralise la notion de longueur, d'une façon qui est parfois difficile à se représenter.

Exemple. On peut définir une mesure de probabilité sur les boréliens de $\mathbb{R}$ en posant

$$\mu(A)=\frac{1}{\sqrt{2\pi}}\int_A e^{-\frac{x^2}{2}}\,d\lambda(x)\ .$$

Cette mesure s'appelle la *mesure gaussienne* et est un cas particulier de *mesure à densité*.

Pour vérifier que c'est une mesure, on commence par voir qu'elle est bien définie, à valeurs positives, et que $\mu(\emptyset)=0$. Pour vérifier la σ-additivité, on prend une suite de boréliens $(A_i)_{i\in\mathbb{N}}$ deux à deux disjoints, et on appelle A leur union. Alors on a $\mathbf{1}_A=\sum_{i=0}^{+\infty}\mathbf{1}_{A_i}$, et l'on peut donc écrire :

$$\begin{aligned}\mu(A) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\mathbf{1}_A(x)d\lambda(x)\\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\sum_{i=0}^{+\infty} e^{-\frac{x^2}{2}}\mathbf{1}_{A_i}(x)d\lambda(x)\\ &= \frac{1}{\sqrt{2\pi}}\sum_{i=0}^{+\infty}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\mathbf{1}_{A_i}(x)d\lambda(x)\\ &= \sum_{i=0}^{+\infty}\mu(A_i)\ .\end{aligned}$$

La passage de la deuxième à la troisième ligne vient du théorème d'échange série-intégrale pour les fonctions à valeurs positives, qui est lui-même une conséquence du théorème de convergence monotone. Tout ce qu'on vient d'esquisser dans cet exemple s'applique en fait dans un cadre plus général, comme on va le voir dans le chapitre suivant.

Chapitre 4

Fonctions mesurables et intégration

On va maintenant décrire brièvement quelles fonctions définies sur un espace mesuré on peut intégrer. L'idée est que, si $(X, \mathcal{A}, \mu)$ est un espace mesuré et f est la fonction caractéristique d'une partie $A \in \mathcal{A}$, alors on voudrait poser $\int_X f d\mu = \mu(A)$. Une fois qu'on a fait ça, on sait comment intégrer toutes les fonctions qui s'écrivent sous la forme $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$: l'intégrale d'une telle fonction devra être $\sum \alpha_i \mu(A_i)$. Et un procédé de passage à la limite nous permettra d'intégrer encore plus de fonctions - les fonctions pour lesquelles ce processus de passage à la limite fonctionne bien sont les *fonctions mesurables.*

4.1 Fonctions mesurables

Définition 4.1. Soit $(X, \mathcal{A})$ un espace mesurable. On dit que $f: X \to \mathbb{R}$ est *mesurable* si, pour tout ouvert O de $\mathbb{R}$, $f^{-1}(O) \in \mathcal{A}$.

Plus généralement, si $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ sont des espaces mesurables, on dit que $f: X \to Y$ est mesurable si $f^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{B}$.

Exemple. Soit $(X, \mathcal{A})$ un espace mesurable, et f la fonction caractéristique d'une partie $A \in \mathcal{A}$, vue comme une fonction de X dans $\mathbb{R}$. Montrer que f est mesurable.

La définition de fonction mesurable à valeurs réelles revient donc à demander que f soit mesurable comme fonction définie sur X muni de la tribu $\mathcal{A}$, et à valeurs dans $\mathbb{R}$ muni de la tribu borélienne. Comme un ouvert est une réunion d'intervalles ouverts, on obtient la même définition en demandant $f^{-1}(I) \in \mathcal{A}$ pour tout intervalle ouvert I.

Un cas particulier est spécialement important : celui où f est une fonction de $\mathbb{R}^n$, muni de sa tribu borélienne, dans $\mathbb{R}$, muni de sa tribu borélienne. Si f est mesurable pour ces tribus, on dit alors que f est *borélienne*. Rappelons qu'une fonction $f: \mathbb{R} \to \mathbb{R}$ est continue si, et seulement si, $f^{-1}(O)$ est ouvert pour tout ouvert O de $\mathbb{R}$. Par conséquent, *toute fonction continue est borélienne* ; la réciproque n'est pas vraie : la fonction caractéristique de $[0, 1]$ (ou de n'importe quelle autre borélien) est borélienne, mais pas du tout continue.

Une bonne raison de s'autoriser à considérer ces fonctions plus générales est qu'on est souvent confronté à des limites simples de suites de fonctions, qui en général ne sont pas continues...

Proposition 4.2. *Soit $(X, \mathcal{A})$ un espace mesurable, et $f_n: X \to \mathbb{R}$ une suite de fonctions mesurables. On suppose que $(f_n(x))$ converge pour tout $x \in X$, vers un réel que l'on note $f(x)$. Alors f est mesurable.*

Démonstration. Soit $I =]a, b[$ un intervalle ouvert. Si $f(x) \in I$, alors $f_n(x)$ appartient à I pour tout n assez grand, par définition de la limite. Réciproquement, si $f_n(x) \in I$ pour tout n suffisamment grand, alors $f(x) \in [a, b]$. Ainsi, on voit que $f(x) \in I$ si, et seulement si, il existe $\varepsilon > 0$ tel que $f_n(x) \in]a + \varepsilon, b - \varepsilon[$ pour tout n suffisamment grand ; menant à l'égalité suivante :

$$f^{-1}(]a, b[) = \bigcup_{N \in \mathbb{N}^*} \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \{x \colon f_m(x) \in \left]a + \frac{1}{N}, b - \frac{1}{N}\right[\} .$$

Chacun des ensembles $\{x \colon f_m(x) \in \left]a + \frac{1}{N}, b - \frac{1}{N}\right[\}$ appartient à $\mathcal{A}$ puisque f_m est mesurable ; une intersection d'une suite d'éléments de $\mathcal{A}$ est encore un élément de $\mathcal{A}$, et de même pour une réunion, ce qui nous donne la conclusion souhaitée. $\square$

La somme, le produit et la composée de fonctions mesurables sont encore des fonctions mesurables.

4.2 Intégrale des fonctions mesurables

Dans cette partie, on fixe un espace mesuré $(X, \mathcal{A}, \mu)$. On commence par le cas particulier des fonctions mesurables à valeurs positives, qui sont exactement les fonctions de la forme

$$f = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i} \ ,$$

où chaque A_i appartient à $\mathcal{A}$ et $\alpha_i \geq 0$. Pour une telle fonction, on pose

$$\int_X f d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i) \ .$$

Il faut faire un peu attention, et vérifier que la définition précédente ne dépend pas de la façon dont on écrit f sous la forme $\sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$...

Ensuite, un procédé de passage à la limite que l'on ne décrira pas ici nous permet de définir l'intégrale d'une fonction mesurable et à valeurs dans $[0, +\infty]$; cette intégrale peut valoir $+\infty$! Puis, en écrivant une fonction mesurable quelconque f sous la forme $f = f^+ - f^-$, on arrive à définir l'intégrale des fonctions mesurables f, à la condition que $\int_X |f| d\mu < +\infty$, exactement comme dans le cas de la mesure de Lebesgue sur $\mathbb{R}$. On dit que ces fonctions sont *intégrables*.

Comme sur $\mathbb{R}$, on définit, étant donnée une partie mesurable $A \subseteq X$, et une fonction mesurable $f \colon X \to \mathbb{R}$,

$$\int_A f d\mu = \int_X f \mathbf{1}_A d\mu.$$

Répétons ici les propriétés de l'intégrale ainsi obtenue :

Proposition 4.3. *Pour toutes fonctions mesurables $f, g \colon X \to \mathbb{R}$ intégrables sur une partie mesurable $A \subseteq X$ on a*

1. $\int_A 1 d\mu = \mu(A)$ *;*
2. *pour toutes constantes réelles α et β, la fonction $\alpha f + \beta g$ est également mesurable et intégrable sur A et*

$$\int_A (\alpha f + \beta g) d\mu = \alpha \int_A f d\mu + \beta \int_A g d\mu \text{ (linéarité de l'intégrale);}$$

3. *si $f = g$ presque partout, alors $\int_A f d\mu = \int_A g d\mu$;*
4. *si $f \geq 0$ presque partout, alors $\int_A f d\mu \geq 0$* (positivité de l'intégrale) *;*
   si $f \leq g$ presque partout, alors $\int_A f d\mu \leq \int_A g d\mu$ (monotonie de l'intégrale) *;*
5. *si $\mu(A) = 0$ alors $\int_A f d\mu = 0$;*
6. *si A_1, A_2 sont des parties mesurables de A telles que $\mu(A_1 \cap A_2) = 0$, alors*

$$\int_{A_1 \cup A_2} f d\mu = \int_{A_1} f d\mu + \int_{A_2} f d\mu \ .$$

On a de plus

7. L'inégalité triangulaire

$$\left| \int_A (f + g) d\mu \right| \leq \int_A |f| d\mu + \int_A |g| d\mu \ .$$

Remarque 4.4. Rappelons que l'intégrale est d'abord définie pour les fonctions mesurables à valeurs positives ($+\infty$ compris) et possède les propriétés ci-dessus pour ces fonctions. Quand on écrit que f est une fonction mesurable de X à valeurs dans l'intervalle fermé $[0, +\infty]$, on entend que $A = \{x \in X \colon f(x) = +\infty\}$ est mesurable, et que pour tout intervalle ouvert I de $\mathbb{R}$, $f^{-1}(I)$ est mesurable. Dans ce cas, son intégrale peut valoir $+\infty$. Si f est intégrable (c.à.d. si $\int_X f d\mu < +\infty$) alors, par monotonie, la partie A où f vaut $+\infty$ est négligeable (c.à.d. $\mu(A) = 0$).

Proposition 4.5 (Inégalité de Tchebychev)**.** *Soit* $f\colon X \to [0,+\infty]$ *une fonction mesurable. Alors, pour tout* $\alpha > 0$*, on a*

$$\mu(\{x\colon f(x) \geq \alpha\}) \leq \frac{1}{\alpha}\int_X f d\mu \ .$$

Preuve:

Notons $B = \{x \in X\colon f(x) \geq \alpha\}$. Alors $\alpha \mathbf{1}_B \leq f$ et par monotonie de l'intégrale, on a

$$\alpha\mu(B) \leq \int_X f d\mu \ .$$

Les propriétés de l'intégrale rappelées dans la proposition 4.3 se vérifient directement à partir de la définition de l'intégrale, à l'exception de la linéarité ou plus précisément de la propriété $\int_A (f+g)d\mu = \int_A f d\mu + \int_A g d\mu$ qui est un corollaire du théorème de convergence monotone. Pour montrer ce dernier, nous allons utiliser la propriété supplémentaire suivante que nous admettrons. Cette propriété se vérifie tout d'abord pour les fonctions simples positives de la forme $\sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$, puis pour toutes les fonctions positives intégrables par le procédé de passage à la limite permettant de définir l'intégrale.

Proposition 4.6 (Mesures à densité)**.** *Soit* $f\colon X \to [0,+\infty]$ *une fonction mesurable. On définit une application* $\nu : \mathcal{A} \to [0,+\infty]$ *par*

$$\nu(A) = \int_A f d\mu \ .$$

Alors, ν *est une mesure sur* X*, appelée* mesure de densité f *par rapport à* μ*.*

Théorème 4.7 (Théorème de convergence monotone)**.** *Soit* $f_n\colon X \to [0,+\infty]$ *une suite de fonctions mesurables telle pour presque tout* $x \in X$ *la suite* $(f_n(x))$ *soit une suite croissante. Alors il existe une fonction mesurable* $f\colon X \to [0,+\infty]$ *telle que* $f_n(x)$ *converge vers* $f(x)$ *presque partout, et on a*

$$\lim_{n\to+\infty}\int_X f_n d\mu = \int_X f d\mu \ .$$

Preuve:

On peut supposer que pour tout $x \in X$, la suite $(f_n(x))$ est une suite croissante (la partie où ce n'est pas le cas étant négligeable). Dans ce cas, pour chaque $x \in X$, la suite $(f_n(x))$ a une limite (finie ou infinie) que l'on note $f(x)$. Cette limite f est alors mesurable (voir la preuve de la proposition 4.2 et la remarque 4.4).

De plus, pour tout $n \in \mathbb{N}$, on a $f_n \leq f_{n+1} \leq f$ et donc, par monotonie de l'intégrale, la suite $\int_X f_n d\mu$ a une limite vérifiant

$$\lim_{n\to+\infty}\int_X f_n d\mu \leq \int_X f d\mu \ .$$

Pour l'inégalité réciproque, considérons $\alpha \in]0,1[$. Pour $n \in \mathbb{N}$, posons $E_n = \{x \in X\colon f_n(x) \geq \alpha f(x)\}$. Alors, par linéarité, chaque E_n est mesurable. De plus, par hypothèse, on a $E_n \subseteq E_{n+1}$ pour tout $n \in \mathbb{N}$, et $\cup_{n\in\mathbb{N}} E_n = X$. Comme $A \mapsto \int_A \alpha f d\mu$ définie une mesure, on en déduit par la proposition 3.6 4 que

$$\int_X \alpha f d\mu = \lim_{n\to+\infty}\int_{E_n} \alpha f d\mu \leq \lim_{n\to+\infty}\int_{E_n} f_n d\mu \leq \lim_{n\to+\infty}\int_X f_n d\mu \ .$$

En faisant tendre α vers 1, on conclut que

$$\int_X \alpha f d\mu \leq \lim_{n\to+\infty}\int_X f_n d\mu \ .$$

Théorème 4.8 (Théorème de convergence dominée)**.** *Soit* $f_n\colon X \to \mathbb{R}$ *une suite de fonctions mesurables satisfaisant les hypothèses suivantes :*

1. *Pour presque tout* $x \in X$*, la suite* $f_n(x)$ *est convergente vers une limite qu'on appelle* $f(x)$*.*
2. *Il existe une fonction* $g\colon X \to [0,+\infty]$ *mesurable et intégrable, telle que pour presque tout* x *on ait*

$$\forall n \in \mathbb{N} \ , \ |f_n(x)| \leq g(x).$$

Alors f est intégrable (ainsi que les f_n), et on a

$$\int_X f d\mu = \lim_{n\to+\infty} \int_X f_n d\mu \ .$$

Rappelons que dans le théorème ci-dessus, il est fondamental que *la fonction g ne dépende pas de n* et que g soit intégrable, c'est-à-dire $\int_X g d\mu < \infty$.

Dans l'énoncé, l'application f est définie presque partout et on la définit par 0 (ou par une fonction mesurable quelconque) sur le complémentaire qui est négligeable. Par la proposition 4.2, la fonction f est mesurable.

Preuve:

Pour tout $n \in \mathbb{N}$, on a $|f_n| \leq g$ presque partout et de même $|f| \leq g$ presque partout par passage à la limite. Comme g est intégrable, par monotonie les f_n et f le sont également.

Pour simplifier, on suppose par la suite la convergence simple et les inégalités vraies partout. On pose pour chaque $n \in \mathbb{N}$,

$$h_n = \inf_{k\geq n} (2g - |f - f_k|).$$

On peut alors vérifier que les fonctions h_n sont mesurables, positives et majorées par $2g$. De plus, pour tout x, la suite $h_n(x)$ est croissante et converge vers $2g$. Par le théorème de convergence monotone, on obtient que $\lim\limits_{n\to+\infty} \int_X h_n d\mu = \int_X 2g d\mu$.

Par définition, on a pour tout n :

$$\int_X h_n d\mu \leq \int_X (2g - |f - f_n|) d\mu \leq \int_X 2g d\mu \ .$$

Ainsi,

$$\left| \int_X f d\mu - \lim_{n\to+\infty} \int_X f_n d\mu \right| = \lim_{n\to+\infty} \left| \int_X (f - f_n) d\mu \right| \leq \lim_{n\to+\infty} \int_X |f - f_n| d\mu = 0.$$

Ces théorèmes ont pour corollaires les théorèmes analogues d'échanges séries/intégrales et de continuité ou dérivabilité des intégrales à paramètres vus dans la première partie pour la mesure de Lebesgue.

Discutons quelques exemples.

1. Si $X = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ et μ est la mesure de comptage sur $\mathbb{N}$, que vaut l'intégrale dans ce cas ? Pour le comprendre, commençons par le cas d'une fonction positive $f\colon \mathbb{N} \to [0, +\infty[$, et même, par un cas simple, où il existe N tel que $f(n) = 0$ pour tout $n \geq N$. Dans ce cas, f ne prend qu'un nombre fini de valeurs $x_1, \ldots, x_k$ et on a par définition de l'intégrale

$$\int_{\mathbb{N}} f d\mu = \sum_{i=1}^{k} x_i \left|\{p \in \mathbb{N} \colon f(p) = x_i\}\right|$$

Dans la somme ci-dessus, x_i apparaît exactement autant de fois qu'il existe d'entiers p pour lesquels $f(p) = x_i$; autrement dit, la somme ci-dessus est simplement la somme de toutes les valeurs de f, c'est-à-dire

$$\int_{\mathbb{N}} f d\mu = \sum_{p=0}^{N} f(p) \ .$$

On vient de voir que, au sens de la mesure de comptage sur $\mathbb{N}$, on a $\int_N f d\mu = \sum_{p=0}^{+\infty} f(p)$, du moment que f est à valeurs positives et nulle pour n suffisamment grand. Maintenant, si jamais $f\colon \mathbb{N} \to [0, +\infty[$ est une fonction quelconque, alors on peut considérer la suite (f_N) définie par

$$f_N(i) = \begin{cases} f(i) & \text{si } i \leq N \\ 0 & \text{sinon} \end{cases}$$

Pour tout i fixé, la suite $(f_N(i))$ est croissante vers $f(i)$; on peut donc appliquer le théorème de convergence monotone pour déduire que

$$\int_{\mathbb{N}} f d\mu = \lim_{N\to+\infty} \int_{\mathbb{N}} f_n d\mu = \lim_{N\to+\infty} \sum_{i=0}^{N} f(i) = \sum_{i=0}^{+\infty} f(i) \ .$$

Ainsi, on a $\int_{\mathbb{N}} f d\mu = \sum_{i=0}^{+\infty} f(i)$ pour toute fonction positive f ; on en déduit qu'une fonction $f\colon \mathbb{N} \to \mathbb{R}$ est intégrable si, et seulement si, la série $\sum_{i=0}^{+\infty} |f(i)|$ converge (ou encore la série $\sum_{i=0}^{+\infty} f(i)$ est absolument convergente) , et que dans ce cas-là, on a encore $\int_{\mathbb{N}} f d\mu = \sum_{i=0}^{+\infty} f(i)$.
Par conséquent, la sommation de séries absolument convergentes peut être vue comme un cas particulier de calcul d'intégrale, relativement à la mesure de comptage sur $\mathbb{N}$.

2. Supposons maintenant que $(X, \mathcal{A}, \mu)$ est un espace mesuré quelconque, $f\colon X \to [0, +\infty]$ est mesurable et ν est la mesure de densité f par rapport à μ. Alors, on sait que, pour toute partie mesurable A, on a
$$\int_X \mathbf{1}_A d\nu = \nu(A) = \int_A f d\mu = \int_X f \mathbf{1}_A d\mu \ .$$
En suivant le même cheminement que ci-dessus (et en admettant le fait que toute fonction mesurable, à valeurs positives, est une limite presque partout de fonctions mesurables, positives et ne prenant qu'un nombre fini de valeurs) on peut alors montrer qu'on a, pour toute fonction mesurable $g\colon X \to [0, +\infty]$:
$$\int_X g d\nu = \int_X f g d\mu \ .$$
Cette formule permet de déduire que les fonctions ν-intégrables sont les fonctions $g\colon X \to \mathbb{R}$ telles que $\int_X f|g| d\mu < +\infty$, et que pour ces fonctions on a aussi $\int_X g d\nu = \int_X f g d\mu$.

4.3 Mesure produit et théorèmes de Fubini

Définition 4.9. Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. On dit que $(X, \mathcal{A}, \mu)$ est *σ-fini* s'il existe une suite de parties $(A_n)_{n \in \mathbb{N}}$ telles que $\mu(A_n) < +\infty$ pour tout n, et $X = \bigcup_n A_n$.

Cette hypothèse est par exemple vérifiée quand $\mu(X) < +\infty$ (donc en particulier quand μ est une mesure de probabilité), quand $X = \mathbb{N}$ muni de la mesure de comptage, ou quand $X = \mathbb{R}^n$ muni de la mesure de Lebesgue.

On voudrait maintenant pouvoir énoncer le théorème de Fubini-Tonelli dans un cadre plus général que celui utilisé au début des notes de cours ; pour cela il nous faut d'abord expliquer comment on peut munir $X \times Y$ d'une structure d'espace mesuré quand X, Y sont tous les deux munis d'une telle structure.

Définition 4.10. Soit $(X, \mathcal{A}, \mu_1)$ et $(Y, \mathcal{B}, \mu_2)$ deux espaces mesurés σ-finis. On note $\mathcal{A} \otimes \mathcal{B}$ la tribu engendrée par les parties de la forme $A \times B$, où $A \in \mathcal{A}$, $B \in \mathcal{B}$; on l'appelle *tribu produit* des tribus $\mathcal{A}$ et $\mathcal{B}$.

Alors il existe une unique mesure ν sur $\mathcal{A} \otimes \mathcal{B}$ vérifiant $\nu(A \times B) = \mu_1(A)\mu_2(B)$ pour tout $A \in \mathcal{A}$ et tout $B \in \mathcal{B}$. Cette mesure est notée $\mu_1 \otimes \mu_2$, et est σ-finie.

On n'essaiera pas de rentrer dans le détail de la construction de cette mesure ; notons que $\mathcal{B}(\mathbb{R}^{n+m}) = \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m)$ et que , si λ_n désigne la mesure de Lebesgue sur $\mathbb{R}^n$, alors on a toujours $\lambda_{n+m} = \lambda_n \otimes \lambda_m$.

La mesure produit $\mu_1 \otimes \mu_2$ étant définie à partir de μ_1 et μ_2, on s'attend à ce qu'il en aille de même de l'intégrale d'une fonction mesurable relativement à $\mu_1 \otimes \mu_2$. Et c'est effectivement le contenu des théorèmes de Fubini.

Théorème 4.11 (Fubini–Tonelli)**.** *Soit $(X_1, \mathcal{T}_1, \mu_1)$ et $(X_2, \mathcal{T}_2, \mu_2)$ deux espaces mesurés σ-finis. Soit $f\colon X_1 \times X_2 \to [0, +\infty[$ une fonction $\mathcal{T}_1 \otimes \mathcal{T}_2$-mesurable. Alors :*

1. *$y \mapsto f(x, y)$ est une fonction mesurable (sur $(X_2, \mathcal{T}_2)$ dans $[0, +\infty]$) pour tout $x \in X_1$, et $x \mapsto \int_{X_2} f(x, y) d\mu_2(y)$ est une fonction mesurable (sur $(X_1, \mathcal{T}_1)$).*
2. *$x \mapsto f(x, y)$ est une fonction mesurable (sur $(X_1, \mathcal{T}_1)$ dans $[0, +\infty]$) pour tout $y \in X_2$, et $y \mapsto \int_{X_1} f(x, y) d\mu_1(x)$ est une fonction mesurable (sur $(X_2, \mathcal{T}_2)$).*

3. *On a*

$$\int_{X_1\times X_2} f(x,y)d\mu_1\otimes\mu_2(x,y) = \int_{X_1}\left(\int_{X_2} f(x,y)d\mu_2(y)\right)d\mu_1(x) = \int_{X_2}\left(\int_{X_1} f(x,y)d\mu_1(x)\right)d\mu_2(y)\ .$$

Comme dans le cas des fonctions définies sur $\mathbb{R}^n$, on en déduit facilement un théorème qui s'applique à toutes les fonctions intégrables (et pour vérifier qu'une fonction est intégrable, on peut commencer par appliquer le théorème de Fubini–Tonelli à $|f|$).

Théorème 4.12 (Fubini)**.** *Soit* $(X_1, \mathcal{T}_1, \mu_1)$ *et* $(X_2, \mathcal{T}_2, \mu_2)$ *deux espaces mesurés σ-finis. Soit* $f\colon X_1\times X_2\to\mathbb{R}$ *une fonction* intégrable*. Alors :*

1. $y\mapsto f(x,y)$ *est une fonction intégrable (sur X_2) pour presque tout $x\in X_1$, et $x\mapsto\int_{X_2} f(x,y)d\mu_2(y)$ est une fonction intégrable (sur X_1).*
2. $x\mapsto f(x,y)$ *est une fonction intégrable (sur X_1) pour preseque tout $y\in X_2$, et $y\mapsto\int_{X_1} f(x,y)d\mu_1(x)$ est une fonction intégrable (sur X_2)*
3. *On a*

$$\int_{X_1\times X_2} f(x,y)d\mu_1\otimes\mu_2(x,y) = \int_{X_1}\left(\int_{X_2} f(x,y)d\mu_2(y)\right)d\mu_1(x) = \int_{X_2}\left(\int_{X_1} f(x,y)d\mu_1(x)\right)d\mu_2(y)\ .$$

Chapitre 5

Un peu de théorie des probabilités

Pour clore ces notes, on va essayer de faire le lien avec certains notions de théorie des probabilités. Comme en probabilité le symbole X est en général utilisé pour une variable aléatoire, on ne l'utilisera plus pour désigner un ensemble - on lui préférera le symbole Ω. De même, on va remplacer le symbole μ par p.

Définition 5.1. Si $(\Omega, \mathcal{T}, p)$ est un espace mesuré et p est une mesure de probabilité, on dira que $(\Omega, \mathcal{T}, p)$ est un *espace probabilisé*.

Définition 5.2. Soit $(\Omega, \mathcal{T}, p)$ un espace probabilisé. On dit que $A, B \in \mathcal{T}$ sont *indépendants* si $p(A \cap B) = p(A)p(B)$.

Quand on travaille avec un espace de probabilité, il faut penser qu'une partie $A \in \mathcal{T}$ correspond à un "événement" et que $p(A)$ est la probabilité que cet événement se produise.

Définition 5.3. Soit $(\Omega, \mathcal{T}, p)$ un espace probabilisé, et $(\mathcal{T}_k)$ une suite de tribus sur Ω, toutes contenues dans $\mathcal{T}$.

On dit que les tribus $\mathcal{T}_0, \ldots, \mathcal{T}_n$ sont *indépendantes* si, à chaque fois qu'on considère $A_0 \in \mathcal{T}_0, \ldots, A_n \in \mathcal{T}_n$, on a

$$p(A_0 \cap \ldots \cap A_n) = p(A_0)p(A_1)\ldots p(A_n) \ .$$

On dit que la suite $(\mathcal{T}_k)$ est indépendante si $\mathcal{T}_0, \ldots, \mathcal{T}_n$ sont indépendantes pour tout $n \in \mathbb{N}$.

En fait, plutôt que de parler de *tribus* indépendantes, on voudrait parler de *variables aléatoires* indépendantes.

Définition 5.4. Soit $(\Omega, \mathcal{T})$ un espace mesurable. Une *variable aléatoire* est une fonction mesurable $X \colon \Omega \to \mathbb{R}$. La *tribu engendrée par* X est la tribu formée par les parties de la forme $X^{-1}(B)$ pour B un borélien de $\mathbb{R}$ (c'est-à-dire, la tribu image réciproque par X de la tribu borélienne de $\mathbb{R}$). On note cette tribu $\mathcal{T}(X)$.

Notons que, par définition, $\mathcal{T}(X) \subseteq \mathcal{T}$ dès que X est une variable aléatoire sur $(\Omega, \mathcal{T})$. Évidemment, il est plus intéressant de parler de variables aléatoires dans un contexte où une mesure de probabilité est présente.

Définition 5.5. Soit $(\Omega, \mathcal{T}, p)$ un espace probabilisé, et X une variable aléatoire sur $(\Omega, \mathcal{T})$. La *loi* de X est la mesure p_X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par

$$p_X(B) = p(X^{-1}(B)) = p(\{\omega \in \Omega \colon X(\omega) \in B\}) = p(X \in B) \ .$$

Ainsi, la mesure (pour p_X) d'une partie de $\mathbb{R}$ est la probabilité (pour p!) qu'un élément de l'univers Ω soit envoyé dans B par X. On a vu en T.D que p_X est bien une mesure de probabilité (la *mesure image* de p par X). En théorie des probabilités, l'univers Ω est rarement explicité, de même que la variable ω, et on emploie le plus souvent la notation $p(X \in B)$ comme à la fin de la ligne ci-dessus.

Définition 5.6. Soit $(\Omega, \mathcal{T}, p)$ un espace probabilisé, et X une variable aléatoire sur $(\Omega, \mathcal{T})$. La *fonction de répartition* de X est la fonction $F \colon \mathbb{R} \to [0,1]$ définie par

$$F(t) = p_X(]-\infty, t]) \ .$$

On admettra le résultat suivant : si deux variables aléatoires X, Y ont la même fonction de répartition (c'est-à-dire, p_X et p_Y prennent les mêmes valeurs sur les intervalles de la forme $]-\infty, t]$), alors elles ont la même loi (c'est-à-dire, p_X et p_Y prennent les mêmes valeurs sur tous les boréliens ; cela vient du fait que les intervalles $]-\infty, t]$ engendrent la tribu borélienne). On dit alors que X, Y sont *identiquement distribuées* ou *équidistribuées*.

On peut maintenant définir l'indépendance de variables aléatoires.

Définition 5.7. Soit $(\Omega, \mathcal{T}, p)$ un espace probabilisé, et (X_i) une suite de variables aléatoires sur $(\Omega, \mathcal{T})$. On dit que $X_0, \ldots, X_n$ sont *indépendantes* si les tribus $\mathcal{T}(X_0), \ldots, \mathcal{T}(X_n)$ sont indépendantes. On dit que la suite (X_i) est *indépendante* si $X_0, \ldots, X_n$ sont indépendantes pour tout n.

De manière équivalente, une suite de variables aléatoires (X_i) sur $(\Omega, \mathcal{T})$ est indépendante si et seulement si, pour tout n et pour tous boréliens $B_0, \ldots, B_n$, on a

$$p(X_0 \in B_0 \text{ et } X_1 \in B_1 \text{ et } \ldots \text{ et } X_n \in B_n) = p(X_0 \in B_0)p(X_1 \in B_1) \ldots p(X_n \in B_n) .$$

Proposition 5.8. *Si $(\Omega, \mathcal{T}, p)$ est un espace probabilisé, et X, Y sont deux variables aléatoires sur Ω, on peut considérer la fonction $Z \colon \Omega \to \mathbb{R}^2$ définie par $Z(\omega) = (X(\omega), Y(\omega))$. Alors la mesure μ_Z poussée en avant de p par Z (mesure image de p par Z sur $\mathcal{B}(\mathbb{R}^2)$) est la mesure produit $p_X \otimes p_Y$.*

Preuve:

Pour $A, B \in \mathcal{B}(\mathbb{R})$, on a

$$\{\omega \in \Omega \colon X(\omega) \in A \text{ et } Y(\omega) \in B\} = \{\omega \in \Omega \colon X(\omega) \in A\} \cap \{\omega \in \Omega \colon Y(\omega) \in B\} = X^{-1}(A) \cap Y^{-1}(B) .$$

Le premier ensemble à droite de l'égalité ci-dessus appartient à la tribu engendrée par X, le second à la tribu engendrée par Y ; comme ces deux tribus sont indépendantes, on a

$$\begin{aligned} \mu_Z(A \times B) &= p(Z^{-1}(A \times B)) \\ &= p(X^{-1}(A) \cap Y^{-1}(B)) \\ &= p(X^{-1}(A))p(Y^{-1}(B)) \\ &= p(X \in A)p(Y \in B) \\ &= p_X(A)p_Y(B) . \end{aligned}$$

Comme μ_Z est une mesure, on obtient par unicité que μ_Z doit être égale à la mesure produit $p_X \otimes p_Y$.

Concluons ces notes par un calcul de loi d'une somme de deux variables indépendantes.

Théorème 5.9. *Soit $(\Omega, \mathcal{T}, p)$ un espace probabilisé, et X, Y deux variables aléatoires sur $(\Omega, \mathcal{T})$. On suppose que p_X, p_Y sont de densité de probabilité f, g relativement à la mesure de Lebesgue, c'est-à-dire $p_X(B) = \int_B f d\lambda$, $p_Y(B) = \int_B g d\lambda$ pour tout borélien B de $\mathbb{R}$. On suppose également que X et Y sont indépendantes.*

Alors, la loi de $X + Y$ est la mesure sur $\mathbb{R}$ dont la densité par rapport à la mesure de Lebesgue est la fonction h définie par

$$\forall x \in \mathbb{R} \quad h(x) = \int_{\mathbb{R}} f(t)g(x - t)d\lambda(t) .$$

La fonction h ci-dessus est appelée *produit de convolution* de f et g.

Preuve:

Notons μ_h la mesure de densité h par rapport à la mesure de Lebesgue. Il nous suffit de montrer que les fonctions de répartition de μ_h et p_{X+Y} sont égales. Fixons donc $a \in \mathbb{R}$, et calculons

$$\begin{aligned} p_{X+Y}(]-\infty, a]) &= p(X + Y \leq a) \\ &= p_X \otimes p_Y(\{(x, y) \in \mathbb{R}^2 \colon x + y \leq a\}) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{x+y \leq a}(x, y) dp_Y(y) \right) dp_X(x) \\ &= \int_{\mathbb{R}} \left(\int_{-\infty}^{a-x} g(y) d\lambda(y) \right) f(x) d\lambda(x) . \end{aligned}$$

La seconde ligne vient du fait que dire que $X+Y \leq a$ est équivalent à dire que $(X,Y) \in \{(x,y) \in \mathbb{R}^2 \colon x+y \leq a\}$, et qu'on sait que la mesure associée à (X,Y) est la mesure produit $p_X \otimes p_Y$ puisque X et Y sont indépendantes ; la troisième ligne correspond à l'application du théorème de Fubini-Tonelli, qu'on peut utiliser puisqu'on manipule des mesures de probabilité ; et la dernière ligne est une conséquence du fait que p_X, p_Y sont de densité f, g par rapport à la mesure de Lebesgue.
Si maintenant, dans l'intégrale en y ci-dessus, on pose $u = y + x$ pour voir un peu mieux ce qui arrive aux bornes, on obtient finalement

$$p_{X+Y}(]-\infty, a]) = \int_{\mathbb{R}} \int_{-\infty}^{a} g(u-x) f(x) d\lambda(u) d\lambda(x) \ .$$

En utilisant une dernière fois le théorème de Fubini (appliqué cette fois-ci à la mesure de Lebesgue sur $\mathbb{R}^2$), on obtient

$$p_{X+Y}(]-\infty, a]) = \int_{-\infty}^{a} \left(\int_{\mathbb{R}} g(u-x) f(x) d\lambda(x) \right) d\lambda(u)$$

On vient de démontrer que

$$p_{X+Y}(]-\infty, a]) = \int_{-\infty}^{a} h(u) d\lambda(u) = \mu_h(]-\infty, a]) \ .$$

Ainsi, p_{X+Y} et μ_h ont la même fonction de répartition, et sont donc égales, ce qu'on voulait démontrer.

Index