

COURS SM 411

Alain Yger

18 septembre 2007

Table des matières

1	L'intégration	1
1.1	Domaines quarrables de la droite, du plan, de l'espace, plus généralement de \mathbb{R}^n	1
1.2	Comment intégrer (et quoi intégrer ?) sur un sous-ensemble quarrable borné de \mathbb{R}^n ?	4
1.3	Le cas des fonctions continues sur un sous-ensemble quarrable borné de \mathbb{R}^n	6
1.4	Le "théorème fondamental de l'analyse"	8
1.5	Le point de vue "discret" ; où l'on entrevoit le théorème de Fubini	8
1.6	Les "briques" quarrables de base en dimension 1,2,3	9
1.6.1	Le cas de la droite réelle	9
1.6.2	Le cas du plan réel	10
1.6.3	Le cas de \mathbb{R}^3	11
1.7	Les "prototypes" d'ensembles quarrables bornés (dans $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$)	12
1.7.1	Le cas de la droite réelle	12
1.7.2	Le cas du plan réel	13
1.7.3	Le cas de l'espace	14
1.8	Comment ramener le calcul d'intégrales multiples au calcul d'intégrales simples ?	16
1.8.1	Extension de la notion d'intégrale aux fonctions mesurables éventuellement non bornées	16
1.8.2	Intégration des fonctions mesurables sur un domaine quarrable éventuellement non borné	16
1.8.3	Le théorème de Fubini	17
1.9	Le changement de variables dans les intégrales multiples	18
1.10	Chemins paramétrés, arcs géométriques	22
1.11	Intégrale curviligne d'une 1-forme le long d'un chemin paramétré C^1 par morceaux	24
1.12	La formule de Green-Riemann dans le plan	25
1.13	Formes exactes et fermées	30
1.14	Nappes paramétrées	32
1.15	La formule de Stokes pour les nappes paramétrées	34
1.16	La formule de Green-Ostrogradski	37
2	Initiation à la théorie des probabilités	41
2.1	Epreuve et ensemble d'évènements	41
2.2	Tribus et probabilités	43
2.3	Notions de probabilité induite et conditionnelle ; indépendance	47
2.4	Variations aléatoires.	50

2.4.1	Variables aléatoires discrètes	50
2.4.2	Variables aléatoires réelles ou vecteurs de variables aléatoires réelles	53
2.5	Variables aléatoires réelles à densité	54
2.6	Fonction de répartition d'une variable aléatoire réelle	55
2.7	Espérance et variance d'une variable aléatoire	55
2.7.1	Le cas des variables aléatoires à valeurs dans un sous-ensemble fini ou dénombrable de \mathbf{R}^n	55
2.7.2	Le cas général des variables aléatoires réelles ou des vecteurs de variables aléatoires à densité	58
2.7.3	Les inégalités de Markov et de Bienaymé-Tchebychev	60
2.8	Indépendance de variables aléatoires réelles	61
2.8.1	Indépendance de deux variables aléatoires réelles	61
2.8.2	Régression linéaire	62
2.8.3	Indépendance mutuelle d'une famille de variables indépendantes	63
2.9	Les théorèmes limite de la théorie des probabilités	63
2.9.1	La notion de convergence en probabilité et la loi faible des grands nombres	63
2.9.2	La notion de convergence en loi et le théorème de la limite centrale	64
2.10	La convergence presque sûre et la loi forte des grands nombres	66

Chapitre 1

L'intégration

1.1 Domaines quarrables de la droite, du plan, de l'espace, plus généralement de \mathbb{R}^n

La droite réelle (pensée comme univers des temps ou univers des fréquences), le plan réel (pensé comme plan des images ou plan des images duales, c'est-à-dire des images diffractées), l'espace \mathbb{R}^3 , l'espace temps \mathbb{R}^4 , sont les exemples les plus importants de modèles sur lesquels vivent les phénomènes physiques (qui eux, sont matérialisés par des fonctions à valeurs réelles ou complexes, voire à valeurs vectorielles).

Il conviendra ultérieurement de leur adjoindre les courbes géométriques lisses du plan ou de l'espace ainsi que les surfaces lisses (comme la sphère de \mathbb{R}^3 par exemple, le cône du futur $t = \sqrt{x^2 + y^2 + z^2}$ de l'espace temps \mathbb{R}^4 privé de son point singulier qu'est l'origine), univers sur lesquels le physicien est souvent amené également à travailler.

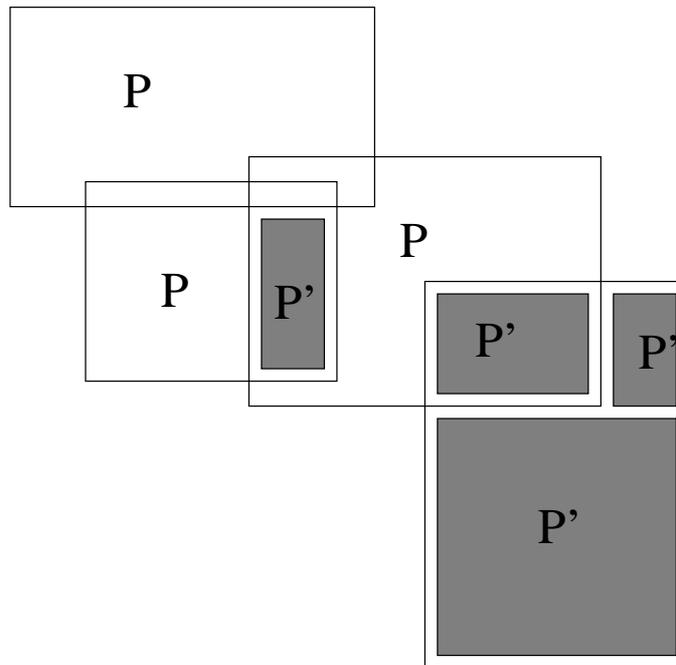
Dans l'univers \mathbb{R}^n , on sait définir le volume n -dimensionnel d'un "pavé"

$$P := [a_1, b_1] \times \cdots \times [a_n, b_n]$$

(avec $-\infty < a_j < b_j < +\infty$, $j = 1, \dots, n$; ce volume vaut par définition

$$\text{vol}_n(P) := (b_1 - a_1) \times \cdots \times (b_n - a_n).$$

On sait donc aussi calculer le volume d'une union finie R de pavés en découpant cette union comme une mosaïque de pavés P' dont les intérieurs $]a_1, b_1[\times \cdots \times]a_n, b_n[$ sont disjoints; le volume de l'union est dans ce cas la somme des volumes des pavés P' (voir figure 1.1).

FIG. 1.1 – Le volume d'une union R de pavés

La physique fait apparaître bien souvent des structures (que l'on qualifie de fractales) s'auto-reproduisant à toutes les échelles, comme par exemple le flocon de neige ; si l'on imagine un tel flocon planaire ou une figure fractale telle l'ensemble de Mandelbrojt représenté en clair sur la figure 1.2, il est difficile d'emblée de décider quelle est la surface de ce domaine !

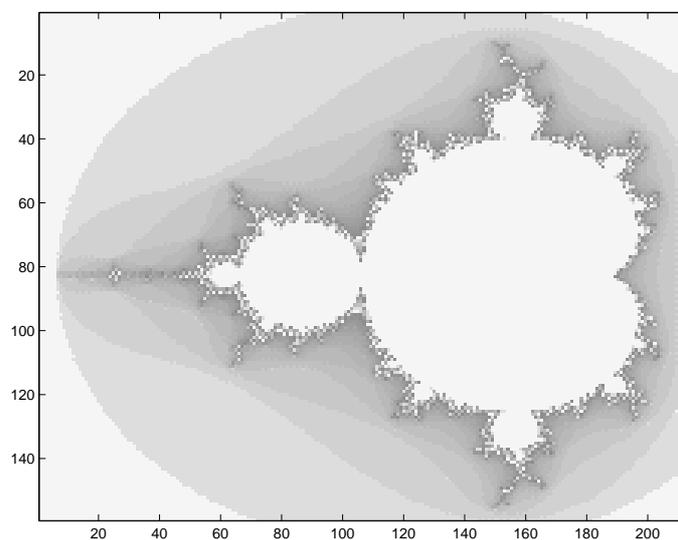


FIG. 1.2 – Une structure planaire “fractale”

Ce qu'il est *a priori* facile de faire est de tenter de mesurer un ensemble borné B de \mathbb{R}^n “de l'extérieur” en définissant le *volume supérieur* de B comme la borne inférieure (c'est-à-dire le plus grand minorant) de l'ensemble

$$\{\text{vol}_n(R) ; R \text{ union de pavés contenant } B\} ;$$

1.1 Domaines quarrables de la droite, du plan, de l'espace, plus généralement de \mathbb{R}^n

on note ce volume supérieur $\text{vol}_n^*(B)$ (voir la figure 1.3).

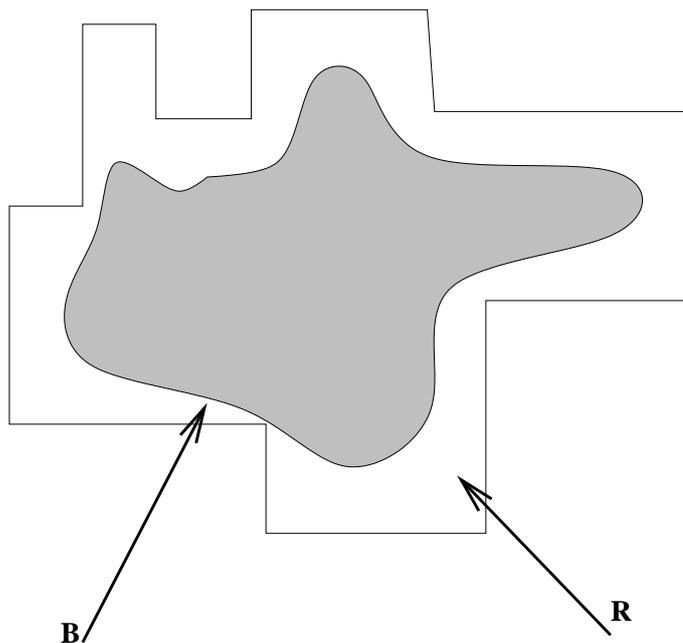


FIG. 1.3 – Le calcul du volume extérieur d'un sous-ensemble borné de \mathbb{R}^n

Si maintenant A est un sous-ensemble borné de \mathbb{R}^n , on dit que A est *quarrable* (ou aussi *mesurable*) si, pour tout $\epsilon > 0$, on peut trouver une union finie de pavés R_ϵ telle que le volume supérieur de l'ensemble

$$\begin{aligned} \Delta(A, R_\epsilon) &:= \{(x_1, \dots, x_n) \in A \text{ tels que } (x_1, \dots, x_n) \notin R_\epsilon\} \\ &\quad \cup \{(x_1, \dots, x_n) \in R_\epsilon \text{ tels que } (x_1, \dots, x_n) \notin A\} \\ &= (A \setminus R_\epsilon) \cup (R_\epsilon \setminus A) \end{aligned}$$

soit au plus ϵ ; sur la figure 1.4, on a représenté un tel ensemble $\Delta(A, R_\epsilon)$. L'ensemble A est donc quarrable si l'on peut trouver des unions finies de rectangles qui "collent" au mieux à A , excepté sur un ensemble de volume supérieur arbitrairement petit (c'est-à-dire intuitivement occupant un espace arbitrairement petit dans l'univers ambiant \mathbb{R}^n).

Si A est un sous-ensemble quarrable borné de \mathbb{R}^n , on appelle *volume n -dimensionnel* de A son volume supérieur; si P_0 est un grand pavé contenant A , on voit que ce volume supérieur est aussi égal au nombre $\text{vol}_n(P_0) - \text{vol}_n^*(P_0 \setminus A)$, ce qui signifie essentiellement que l'on obtient le même nombre en approchant A "de l'extérieur" ou "de l'intérieur"; c'est là précisément ce qui caractérise les sous-ensembles bornés quarrables de \mathbb{R}^n et permet d'en définir sans aucune ambiguïté le volume n -dimensionnel.

Si A est maintenant un sous-ensemble quelconque de \mathbb{R}^n (on oublie le fait qu'il soit borné), on dit que A est quarrable si toutes les intersections $A \cap \{x \in \mathbb{R}^n; \|x\| \leq N\}$, $N = 1, 2, \dots$, le sont; le volume n -dimensionnel de A (qui peut valoir dans ce cas $+\infty$) est alors défini par

$$\text{vol}_n(A) := \lim_{N \rightarrow +\infty} \text{vol}_n(A \cap \{x; \|x\| \leq N\}) \in [0, +\infty]$$

(cette limite existe si l'on tolère $+\infty$ comme limite éventuelle car il s'agit d'une suite croissante de nombres positifs ou nuls).

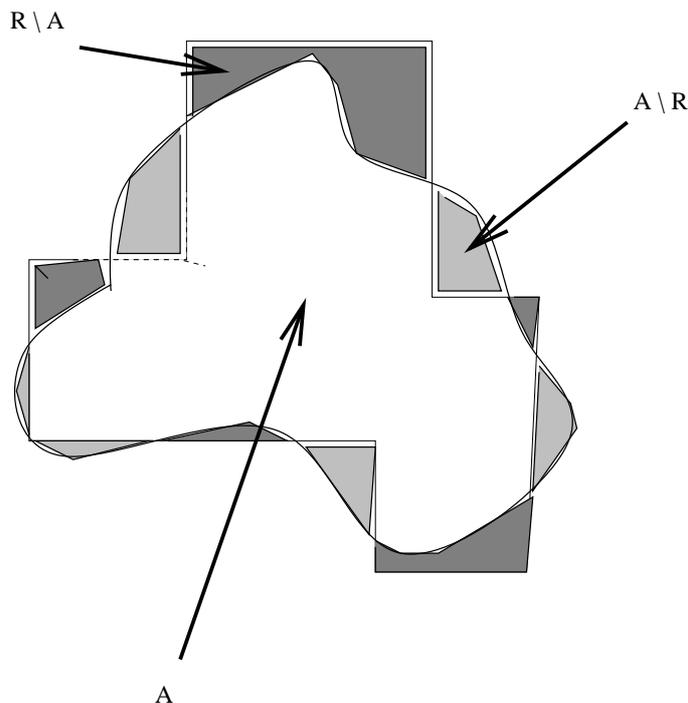


FIG. 1.4 – La différence symétrique $\Delta(A, R)$ de A et d'une union finie de pavés R

1.2 Comment intégrer (et quoi intégrer ?) sur un sous-ensemble quarrable borné de \mathbb{R}^n ?

Soit A un sous-ensemble quarrable borné de \mathbb{R}^n et f une fonction bornée de A dans \mathbb{R} , à valeurs dans $[m, M[$.

On fait une hypothèse supplémentaire sur f , qui n'est pas automatique : dès que α et β sont deux nombres réels tels que $m \leq \alpha < \beta < M$, le sous-ensemble $A \cap f^{-1}([\alpha, \beta[)$ est supposé être un sous-ensemble quarrable de A . Une telle fonction f est dite *mesurable* sur A , à valeurs dans $[m, M[$.

Pour définir l'intégrale sur A d'une fonction mesurable à valeurs dans $[m, M[$, on procède comme suit. On examine le graphe de f au dessus de A et l'on découpe l'intervalle $[m, M[$ en utilisant une subdivision

$$y_0 = m < y_1 < \dots < y_{M-1} < y_M = M ;$$

les sous-ensembles

$$A_j := \{(x_1, \dots, x_n) \in A; y_j \leq f(x_1, \dots, x_n) < y_{j+1}\}, \quad j = 0, \dots, M-1,$$

sont des sous-ensembles quarrables de A partitionnant A ; chacun d'eux a donc un volume n -dimensionnel fini $\text{vol}_n(A_j)$ et l'on peut associer à la subdivision $\mathcal{P} := \{y_0, y_1, \dots, y_m\}$ la quantité

$$I_*(f; \mathcal{P}) := \sum_{j=0}^{M-1} \text{vol}_n(A_j) \times y_j ;$$

ensuite, on définit *l'intégrale de f sur A* comme la borne supérieure de l'ensemble des nombres $I_*(f; \mathcal{P})$ pour toutes les subdivisions de $[m, M[$ possibles; on pourrait d'ailleurs tout aussi bien définir cette intégrale comme la borne inférieure de

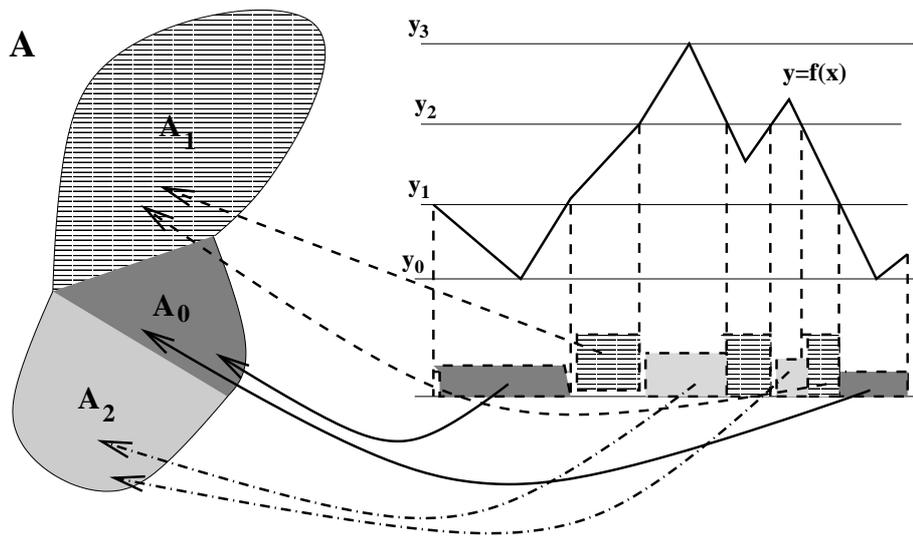


FIG. 1.5 – Découpage en tranches du graphe et intégration d’une fonction mesurable bornée sur un sous-ensemble quarrable borné de \mathbb{R}^n

l’ensemble des nombres

$$I^*(f; \mathcal{P}) := \sum_{j=0}^{M-1} \text{vol}_n(A_j) \times y_{j+1}$$

pour toutes les subdivisions possibles. Sur la figure 1.5, on a schématisé la démarche en représentant le découpage en “tranches” du graphe de f , le partitionnement de A qui lui correspond, et en indiquant les correspondances.

Le fait que les deux approches (par en dessous et par en dessus) coïncident est une conséquence du fait que la fonction f est mesurable. On note l’intégrale de f sur A sous la forme

$$\int \dots \int_A f(x) dx_1 \dots dx_n = \sup_{\mathcal{P}} (I_*(f; \mathcal{P})) = \inf_{\mathcal{P}} (I^*(f; \mathcal{P})),$$

le nombre de symboles intégraux dans $\int \dots \int$ étant égal à la dimension n de l’espace dans lequel on travaille. Si $n = 1$, on parle d’*intégrale* de f sur $A \subset \mathbb{R}$, si $n = 2$, d’*intégrale double* de f sur $A \subset \mathbb{R}^2$, si $n = 3$, d’*intégrale triple* de f sur $A \subset \mathbb{R}^3$, etc...

Cette approche de la notion d’intégrale, fondée sur une exploration par tranches horizontales du graphe des fonctions, est essentiellement celle qu’a proposé en 1904 le mathématicien français Henri Lebesgue (1875-1941).

Si f est une fonction complexe bornée sur A , on l’écrit $f = \text{Re } f + i \text{Im } f$ et elle est mesurable sur A dès que ses parties réelles et imaginaires le sont ; l’intégrale de f sur A est alors

$$\int \dots \int_A f dx_1 \dots dx_n := \int \dots \int_A \text{Re } f dx_1 \dots dx_n + i \int \dots \int_A \text{Im } f dx_1 \dots dx_n;$$

la même idée permet de définir l’intégrale sur A des fonctions mesurables bornées à valeurs dans \mathbb{R}^q (toutes les composantes f_k , $k = 1, \dots, q$ sont mesurables) : c’est le vecteur dont les composantes sont les intégrales sur A des f_k .

1.3 Le cas des fonctions continues sur un sous-ensemble quarrable borné de \mathbb{R}^n

Une fonction définie sur un sous-ensemble A de \mathbb{R}^n est dite *continue* en un point a de A si

$$\lim_{\substack{x \rightarrow a \\ x \in A}} f(x) = f(a).$$

Si $n = 1$ et $A = [a, b]$, on connaît au moins deux manières de calculer l'intégrale d'une fonction continue f de A dans \mathbb{R} :

- la première est “mathématique” et repose sur le fait que l'on dispose explicitement d'une primitive de f au voisinage de $[a, b]$, c'est-à-dire d'une fonction F définie et dérivable au voisinage de $[a, b]$, avec $F'(t) = f(t)$ sur $[a, b]$. Notons que ceci est bien rare en général, ne se produit que si f appartient à une classe de fonctions simples (fractions polynomiales, rationnelles, trigonométriques, expressions simples impliquant le logarithme, etc) et exige souvent l'utilisation de tables de primitives ! Alors, on peut écrire

$$\int_{[a,b]} f(t)dt = F(b) - F(a),$$

nous y reviendrons ;

- la seconde est celle du numéricien, d'usage beaucoup plus fréquent, mais c'est une méthode approchée : elle consiste à découper $[a, b]$ avec un pas de $(b-a)/N$ et à remarquer que

$$\int_{[a,b]} f(t)dt = \lim_{N \rightarrow \infty} \left(\frac{(b-a)}{N} \sum_{j=0}^{N-1} f(\xi_{N,j}) \right),$$

où $\xi_{N,j} \in [a + j(b-a)/N, +(j+1)(b-a)/N]$ pour $j = 0, \dots, N-1$.

Laissons pour l'instant de côté ce que devient la première méthode dans le cas général où $n > 1$ et où $[a, b]$ est remplacé par un sous-ensemble quarrable borné $A \subset \mathbb{R}^n$.

La seconde méthode fonctionne bien lorsque A est un sous-ensemble borné de \mathbb{R}^n , fermé (c'est-à-dire tel que toute limite d'une suite de points de A est dans A) et f une fonction continue à valeurs réelles de A dans \mathbb{R} . Un tel ensemble A est dit *compact* (et l'on admettra qu'il est quarrable) et toute fonction continue sur A est automatiquement bornée sur cet ensemble. On peut donc définir l'intégrale de f sur A et la calculer de manière sensiblement différente que ce que nous avons fait au paragraphe 1.2. Faisons le par exemple si $n = 2$. Pour $\epsilon > 0$ arbitrairement petit, on plaque contre A une “grille” de maille ayant pour cotés $\epsilon \times \epsilon$ et l'on dresse la liste de toutes les mailles $P_{\epsilon,1}, \dots, P_{\epsilon,N(\epsilon)}$ rencontrant A (on laisse tomber les autres), comme sur la figure 1.6. On montre que

$$\iint_A f(x_1, x_2)dx_1dx_2 = \lim_{\epsilon \rightarrow 0} \left(\epsilon^n \sum_{j=1}^{N(\epsilon)} f(M_{\epsilon,j}) \right),$$

où $M_{\epsilon,j}$ désigne, pour tout j entre 1 et $N(\epsilon)$, un point arbitraire de $P_{\epsilon,j} \cap A$. Ceci fournit dans ce cas particulier une autre approche pour calculer l'intégrale d'une fonction continue sur un sous-ensemble compact de \mathbb{R}^n . Ceci rejoint une approche

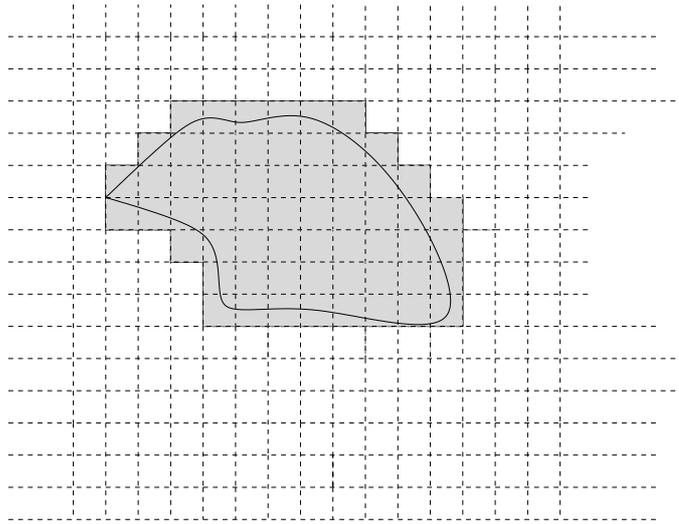


FIG. 1.6 – L'intégration d'une fonction continue sur un compact de \mathbb{R}^n

proposée dès 1854 par le mathématicien allemand Bernhard Riemann (1826-1866) basée, elle, non pas sur le découpage du graphe comme celle de Lebesgue, mais sur le découpage suivant des grilles de plus en plus fines de l'ensemble quarrable A sur lequel f est définie. Si (\mathbf{P}) est une collection de pavés P d'intérieurs disjoints dont l'union recouvre A , on peut remarquer que, si f est une fonction réelle mesurable bornée sur A ,

$$\begin{aligned} \sum_{P \in (\mathbf{P}), P \cap A \neq \emptyset} \text{vol}_n(P) \inf_{P \cap A} (f) &\leq \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &\leq \sum_{P \in (\mathbf{P}), P \cap A \neq \emptyset} \text{vol}_n(P) \sup_{P \cap A} (f) ; \end{aligned}$$

si l'on note

$$\begin{aligned} J_*((\mathbf{P}), f) &:= \sum_{P \in (\mathbf{P}), P \cap A \neq \emptyset} \text{vol}_n(P) \inf_{P \cap A} (f) \\ J^*((\mathbf{P}), f) &:= \sum_{P \in (\mathbf{P}), P \cap A \neq \emptyset} \text{vol}_n(P) \sup_{P \cap A} (f) , \end{aligned}$$

la classe \mathcal{R} de fonctions bornées sur A telles que

$$\sup_{\mathcal{P}} (J_*((\mathbf{P}), f)) = \inf_{\mathcal{P}} (J^*((\mathbf{P}), f))$$

est une classe de fonctions plus restreinte que celle des fonctions mesurables sur A , mais englobant tout de même, si A est compact, la classe des fonctions continues sur A .

Si A est un sous-ensemble quarrable de \mathbb{R}^n et si f est dans la classe \mathcal{R} , l'intégrale de f sur A s'obtient en calculant soit $J_*((\mathbf{P}), f)$, soit $J^*((\mathbf{P}), f)$, soit une expression entre les deux, comme

$$\sum_{P, P \cap A \neq \emptyset} \text{vol}_n(P) f(M_P)$$

où M_P est un point de $P \cap A$ si $P \cap A \neq \emptyset$, puis en faisant tendre le pas de la collection (\mathbf{P}) , c'est-à-dire le diamètre maximal des pavés qui la composent, vers

0. Ceci constitue une autre approche “physique” au calcul approché de l’intégrale d’une fonction réelle bornée sur un sous-ensemble quarrable borné de \mathbf{R}^n (pourvu que cette fonction soit dans la classe \mathcal{R} , on dit aussi soit *intégrable au sens de Riemann* si on la prolonge par 0 hors de A). Dans ce cours (et parce que nous aurons à proposer une initiation à la théorie des probabilités, elle très inspirée du point de vue de Lebesgue), nous privilégierons le point de vue de Lebesgue, permettant d’intégrer une classe de fonctions plus large, à celui de Riemann.

1.4 Le “théorème fondamental de l’analyse”

Souvent la formule très simple

$$\int_{[a,b]} f(t)dt = F(b) - F(a)$$

si f est une fonction continue sur $[a, b] \subset \mathbf{R}$ et F une primitive de f sur $[a, b]$ est qualifiée de “*théorème fondamental de l’analyse*”. Cette formule admet une seconde lecture, plus intéressante si l’on se place du point de vue du physicien : plutôt que de considérer la fonction f , on va s’intéresser plus loin dans le cours à ce qu’on appellera la *forme différentielle* $f(t)dt$, dérivant du *potentiel* F puisque $dF = f(t)dt$ vu que F est une primitive de f . Notre formule se lit aussi

$$\int_{[a,b]} dF = F(b) - F(a);$$

mais l’on peut imaginer l’expression $F(b) - F(a)$ comme l’intégrale sur le *bord* de $[a, b]$, l’extrémité b étant “chargée” avec $+1$, l’origine a étant “chargée” avec -1 (comme un dipôle). Vue comme cela, c’est-à-dire comme

$$\int_{[a,b]} dF = \int_{\text{bord de } [a,b]} F,$$

notre formule devient une version en dimension 1 de ce que l’on appellera la *formule de Stokes* (de Green-Riemann en dimension 2, de Green-Ostrogradski en dimension 3), formule à laquelle nous parviendrons au terme de cette première partie du cours. C’est cette formule de Stokes qui nous permettra d’exprimer dans certains cas

$$\int \dots \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n$$

comme nous venons d’exprimer $\int_{[a,b]} f(t)dt$ sous la forme $F(b) - F(a)$ à l’aide d’une primitive de f sur $[a, b]$ (ou, pour parler en termes plus “physiques”, d’un potentiel dont dérive la forme $f(t)dt$ sur $[a, b]$).

1.5 Le point de vue “discret” ; où l’on entrevoit le théorème de Fubini

Le traitement d’image donne une vision concrète de ce qu’est l’intégrale multiple, en même temps qu’il prépare à une initiation aux probabilités.

Soit par exemple une image 4×4 donc à 16 pixels, chaque “pixel” ayant une brillance (que l’on peut coder dans un code de couleurs graphiques) valant 1, 2 ou 3. Si le

nombre de pixels à brilliance 1 est n_1 , celui à brilliance 2 valant n_2 , celui à brilliance 3 valant n_3 , l'intégrale de l'image est

$$\iint I = 1 \times n_1 + 2 \times n_2 + 3 \times n_3.$$

1	2	2	3
2	3	1	2
1	2	2	3
3	1	3	1

FIG. 1.7 – Intégration discrète

On voit aussi sur l'exemple que le calcul de l'intégrale de I peut se faire par colonne ou par ligne : sur chaque colonne j , $j = 1, \dots, 4$, on calcule la somme S_j des brilliances des pixels de la colonne, puis on ajoute les S_j pour trouver $\iint I$. On peut faire pareil avec les lignes, ce qui soutend la formule

$$\iint_A f(x, y) dx dy = \int \left(\int_{A_x} f(x, y) dy \right) dx = \int \left(\int_{A^y} f(x, y) dx \right) dy,$$

où A_x est la tranche verticale $A_x := \{y; (x, y) \in A\}$ et A^y la tranche horizontale $A^y := \{x; (x, y) \in A\}$. Ce résultat (dit théorème de Fubini) sera l'une des clefs des calculs pratiques d'intégrales multiples sur des domaines simples (car il permet de descendre la dimension, de passer des intégrales doubles aux intégrales simples, des intégrales triples aux intégrales doubles, etc...)

1.6 Les “briques” quarrables de base en dimension 1,2,3

1.6.1 Le cas de la droite réelle

Parmi les sous-ensembles quarrables bornés de \mathbf{R} , les plus élémentaires sont les intervalles du type (a, b) , avec $-\infty < a < b < +\infty$, les parenthèses signifiant que

les extrémités sont incluses ou exclues, au choix. Si les extrémités sont exclues, on dit que A (ici donc $]a, b[$) est un *intervalle ouvert*, ce qui signifie que, pour tout $t \in A$, il existe un intervalle $]t - \epsilon_t, t + \epsilon_t[$ inclus dans A et contenant t . Si les deux extrémités sont incluses, l'intervalle (ici $A = [a, b]$) est dit *intervalle fermé* (c'est le complémentaire de A , soit $] - \infty, a[\cup]b, +\infty[$ qui dans ce cas a la propriété d'être ouvert, c'est-à-dire d'être un voisinage de chacun de ses points).

Dans tous les cas, le volume 1-dimensionnel de (a, b) (on parle de *longueur* de (a, b)) vaut

$$\text{vol}_1((a, b)) = b - a,$$

(exclure ou inclure les extrémités ne modifiant rien).

1.6.2 Le cas du plan réel

Parmi les sous-ensembles quarrables bornés de \mathbb{R}^2 , les plus élémentaires sont les *parallélogrammes*; un tel parallélogramme (voir la figure 1.8) se construit à partir de deux vecteurs $\vec{v}_1 = (x_1, y_1)$, $\vec{v}_2 = (x_2, y_2)$ linéairement indépendants.

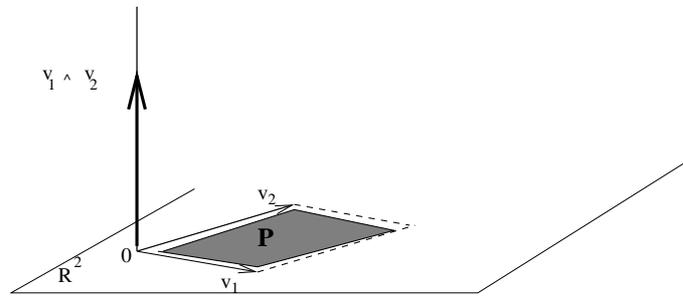


FIG. 1.8 – Parallélogrammes et produit extérieur de deux vecteurs indépendants du plan

Le volume 2-dimensionnel (on dit dans ce cas la *surface*) du parallélogramme P ainsi construit reste inchangé suivant que l'on inclut ou exclut certains sommets ou certaines arêtes; le parallélogramme P construit à partir de $\vec{v}_1 = (x_1, y_1)$ et $\vec{v}_2 = (x_2, y_2)$ a pour volume 2-dimensionnel ou surface

$$\text{vol}_2(P) = \left| \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \right| = |x_1 y_2 - x_2 y_1|$$

comme le montre un découpage élémentaire en triangles (que l'on pourra s'entraîner à faire).

Ce calcul de surface est lié à la notion de *produit extérieur* du plan \mathbb{R}^2 . Pour comprendre cette notion de produit extérieur, il faut imaginer \mathbb{R}^2 comme le plan $\{z = 0\}$ de \mathbb{R}^3 (comme sur la figure 1.8); le produit extérieur $\vec{i} \wedge \vec{j}$ de $\vec{i} = (1, 0) \simeq (1, 0, 0)$ et $\vec{j} = (0, 1) \simeq (0, 1, 0)$ est par définition le vecteur $\vec{k} = (0, 0, 1)$ de \mathbb{R}^3 , tandis que $\vec{j} \wedge \vec{i} = -\vec{k}$, $\vec{i} \wedge \vec{i} = \vec{j} \wedge \vec{j} = \vec{0}$, puis cette opération est prolongée par linéarité à une paire de vecteurs quelconques du plan, ce qui donne

$$(x_1 \vec{i} + y_1 \vec{j}) \wedge (x_2 \vec{i} + y_2 \vec{j}) = (x_1 y_2 - x_2 y_1) \vec{k}.$$

La surface du parallélogramme construit à partir de deux vecteurs linéairement indépendants \vec{v}_1 et \vec{v}_2 de \mathbb{R}^2 est donc la norme du vecteur $\vec{v}_1 \wedge \vec{v}_2$. Par *norme*

d'un vecteur $\vec{v} = (x, y)$ (resp $\vec{v} = (x, y, z)$) de \mathbb{R}^2 (resp. \mathbb{R}^3), on entend la *norme euclidienne*

$$\|\vec{v}\| := \sqrt{x^2 + y^2} \quad (\text{resp. } \|v\| := \sqrt{x^2 + y^2 + z^2}),$$

norme dérivant du produit scalaire usuel dans \mathbb{R}^2 (resp. \mathbb{R}^3) :

$$\begin{aligned} \langle (x_1, y_1), (x_2, y_2) \rangle &= x_1x_2 + y_1y_2 \\ (\text{resp. } \langle (x_1, y_1, z_1), (x_2, y_2, z_2) \rangle &= x_1x_2 + y_1y_2 + z_1z_2); \end{aligned}$$

cette norme vérifie, comme toute norme, l'inégalité triangulaire

$$\|\vec{v}_1 + \vec{v}_2\| \leq \|\vec{v}_1\| + \|\vec{v}_2\|$$

ainsi que $\|\lambda v\| = |\lambda| \|\vec{v}\|$ pour tout vecteur \vec{v} et tout scalaire $\lambda \in \mathbb{R}$; de plus, si \vec{v}_1 et \vec{v}_2 sont orthogonaux, on a, avec cette norme euclidienne, le théorème de Pythagore

$$\|\vec{v}_1\|^2 + \|\vec{v}_2\|^2 = \|\vec{v}_1 + \vec{v}_2\|^2.$$

On a aussi la formule importante suivante, avatar de l'identité $\cos^2 \theta + \sin^2 \theta = 1$, à savoir

$$|\langle \vec{v}_1, \vec{v}_2 \rangle|^2 + \|\vec{v}_1 \wedge \vec{v}_2\|^2 = \|\vec{v}_1\|^2 \times \|\vec{v}_2\|^2.$$

Si $\vec{v}_1 = (x_1, y_1)$ et $\vec{v}_2 = (x_2, y_2)$ sont deux vecteurs de \mathbb{R}^2 , la quantité

$$\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}$$

représente la *surface orientée* du parallélogramme P construit sur les deux vecteurs \vec{v}_1 et \vec{v}_2 , pris dans cet ordre (l'ordre des vecteurs est ici important!); cette surface orientée est exactement la surface de P si le déterminant des deux vecteurs \vec{v}_1, \vec{v}_2 est strictement positif (on dit alors que la base $\{\vec{v}_1, \vec{v}_2\}$ est *directe*); c'est l'opposé de la surface de P dans l'autre cas (on dit que la base $\{\vec{v}_1, \vec{v}_2\}$ est *indirecte*).

1.6.3 Le cas de \mathbb{R}^3

Parmi les les sous-ensembles quarrables bornés de \mathbb{R}^3 , les plus élémentaires sont les *parallélépipèdes*; un tel parallélépipède (voir la figure 1.9) se construit à partir de trois vecteurs $\vec{v}_1 = (x_1, y_1, z_1)$, $\vec{v}_2 = (x_2, y_2, z_2)$, $\vec{v}_3 = (x_3, y_3, z_3)$ linéairement indépendants.

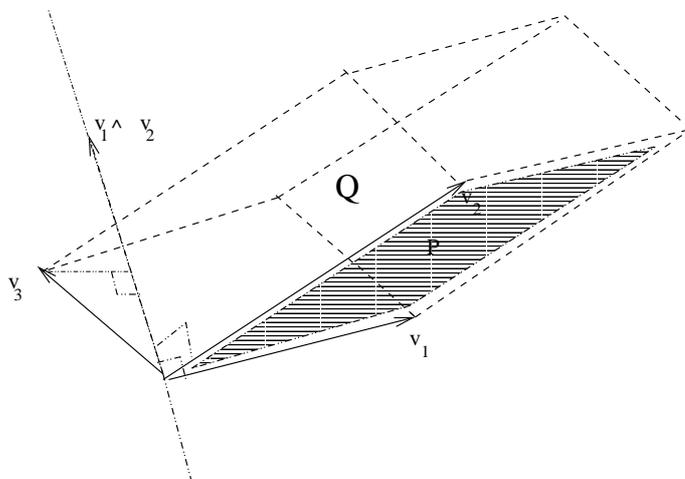


FIG. 1.9 – volume du parallélépipède

Le volume 3-dimensionnel (on dit cette fois *volume*) du parallélépipède Q ainsi construit reste inchangé suivant que l'on inclut ou exclut certains sommets, certaines faces, ou certaines arêtes ; le parallélépipède Q construit à partir de $\vec{v}_1 = (x_1, y_1, z_1)$, $\vec{v}_2 = (x_2, y_2, z_2)$ et $\vec{v}_3 = (x_3, y_3, z_3)$ a pour volume

$$\text{vol}_3(Q) = \left| \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} \right|$$

comme le montre un découpage élémentaire en pyramides (que l'on pourra s'entraîner à faire) ; on en verra d'ailleurs une autre explication plus loin.

Ici encore, ce calcul de volume est lié au produit extérieur, cette fois opération naturelle de $\mathbf{R}^3 \times \mathbf{R}^3$ dans \mathbf{R}^3 , défini par les règles :

$$\begin{aligned} \vec{i} \wedge \vec{j} &= -\vec{j} \wedge \vec{i} = \vec{k} \\ \vec{j} \wedge \vec{k} &= -\vec{k} \wedge \vec{j} = \vec{i} \\ \vec{k} \wedge \vec{i} &= -\vec{i} \wedge \vec{k} = \vec{j} \\ \vec{i} \wedge \vec{i} &= \vec{j} \wedge \vec{j} = \vec{k} \wedge \vec{k} = \vec{0}, \end{aligned}$$

et prolongée ensuite par linéarité, ce qui donne, si $\vec{v}_1 = (x_1, y_1, z_1)$ et $\vec{v}_2 = (x_2, y_2, z_2)$,

$$\vec{v}_1 \wedge \vec{v}_2 = (y_1 z_2 - z_1 y_2) \vec{i} + (z_1 y_2 - z_2 y_1) \vec{j} + (x_1 y_2 - x_2 y_1) \vec{k} = \begin{vmatrix} x_1 & x_2 & \vec{i} \\ y_1 & y_2 & \vec{j} \\ z_1 & z_2 & \vec{k} \end{vmatrix}.$$

On remarque d'ailleurs que le volume de Q s'exprime sous la forme

$$\text{vol}_3(Q) = \left| \langle \vec{v}_1 \wedge \vec{v}_2, \vec{v}_3 \rangle \right|; \quad (*)$$

en fait, la norme de $\vec{v}_1 \wedge \vec{v}_2$ représente la surface 2-dimensionnelle du parallélogramme P construit à partir de \vec{v}_1 et \vec{v}_2 , calculée dans le plan vectoriel contenant ces deux vecteurs en prenant la même unité de longueur (c'est-à-dire concrètement le même mètre !) que l'unité choisie pour mesurer les longueurs dans \mathbf{R}^2 ; la formule (*) correspond donc au calcul attendu du volume d'un parallélépipède en multipliant la surface de la base par la longueur de la hauteur (voir encore la figure 1.9).

Le déterminant des trois vecteurs $\vec{v}_1, \vec{v}_2, \vec{v}_3$ est appelé *volume orienté* du parallélépipède Q construit à partir de $\vec{v}_1, \vec{v}_2, \vec{v}_3$ (l'ordre des vecteurs est ici important !). On l'appelle aussi *produit mixte* des trois vecteurs. Ce volume orienté est exactement le volume de Q si le déterminant des trois vecteurs indépendants $\vec{v}_1, \vec{v}_2, \vec{v}_3$ est strictement positif (on dit alors que la base $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ est *directe*) ; c'est l'opposé du volume de Q dans l'autre cas (on dit que la base $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ est *indirecte*).

1.7 Les “prototypes” d'ensembles quarrables bornés (dans $\mathbf{R}, \mathbf{R}^2, \mathbf{R}^3$)

1.7.1 Le cas de la droite réelle

Les ensembles quarrables bornés que nous considérerons (et sur lesquels nous intégrerons les fonctions) sont les unions finies d'intervalles (a_i, b_i) disjoints.

Si A n'est constitué que d'un intervalle, on remarque que A est “d'un seul tenant” ; on peut joindre continuellement deux points de A en restant dans A sans lever le crayon ; on dit alors que A est *connexe*.

1.7.2 Le cas du plan réel

Les ensembles quarrables du plan que nous considérerons sont les domaines bornés dits élémentaires (ou parfois de *Jordan généralisés*) ; ils sont du type représenté sur la figure 1.10.

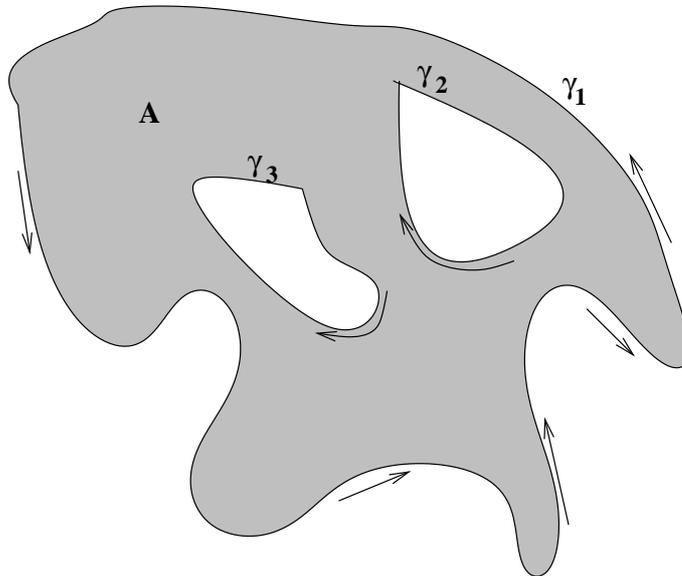


FIG. 1.10 – Un domaine de Jordan “généralisé”

Un tel domaine est un domaine connexe (c’est-à-dire d’un seul tenant) du plan limité par un nombre fini de supports de *lacets paramétrés sans point double*, C^1 par morceaux, ces supports étant supposés disjoints.

Précisons donc le vocabulaire :

- un *lacet paramétré de classe C^1 sans point double* est une application $\gamma : [0, 1] \rightarrow \mathbb{R}^2$ de classe C^1 par morceaux (ce qui signifie qu’il existe une subdivision $0 = t_0 < \dots < t_M = 1$ telle que la restriction de γ à chaque intervalle $[t_j, t_{j+1}]$, $j = 0, \dots, M - 1$, ait ses deux composantes de classe C^1 sur cet intervalle fermé), telle que $\gamma(0) = \gamma(1)$ (pour justifier le qualificatif *lacet*) et que la restriction de γ à $[0, 1[$ soit injective (il n’y a pas de point double autre que l’origine-extrémité du lacet)
- le *support* d’un lacet paramétré de classe C^1 par morceaux et sans point double est par définition l’image par γ de l’intervalle $[0, 1]$.

Un domaine de Jordan généralisé se présente donc comme un “guyère” borné, avec un bord externe et des bords internes (les bords des trous). Le *bord* du domaine A (que l’on peut décider d’inclure ou non dans A , voire d’inclure partiellement, ce qui ne change pas ni la propriété que le domaine A soit quarrable, ni le calcul de sa surface) est dit *bord de A* .

L’ensemble A privé de son bord est dit *intérieur* de A et noté $\text{int}(A)$; c’est un *ouvert* de \mathbb{R}^2 , au sens où quelque soit (x, y) dans $\text{int}(A)$, il existe un disque

$$D((x, y), \epsilon(x, y)) = \{(u, v) \in \mathbb{R}^2 ; \|(u - x, v - y)\| < \epsilon(x, y)\}$$

inclus dans $\text{int}(A)$. L’union de A et de son bord est lui un ensemble fermé ; tout autre ensemble obtenu en ne retirant qu’une partie du bord n’est ni ouvert ni fermé.

On appelle *bord orienté* de A le bord de A , avec un sens de parcours imposé par la règle suivante, dite *règle du bonhomme d'Ampère* : on doit parcourir les supports des lacets de manière à conserver le domaine A à sa gauche tout au long du parcours ; le bord “externe” (ici le support de γ_1 sur la figure) hérite donc de l'orientation trigonométrique (dans le sens contraire des aiguilles d'une montre), tandis que le bord “interne” (ici les supports des lacets γ_2 et γ_3) hérite de l'orientation inverse. Cette notion de bord orienté jouera tout son rôle lorsque nous aurons défini (plus loin) la notion *d'intégrale curviligne* ou, pour les physiciens, de *travail d'un champ de forces* le long d'un chemin paramétré.

Un tel domaine de Jordan généralisé est tel, s'il n'y a pas de trou, que tout lacet fermé enroulé dans le domaine A puisse être déformé de manière continue en un point, ce en restant dans le domaine ; on dit dans ces conditions que A est *simplement connexe* ; un tel domaine de Jordan simplement connexe (c'est-à-dire sans trous) est dit *domaine de Jordan*. Il est clair que dès qu'il y a un trou, le domaine n'est pas simplement connexe : un lacet fermé entourant le trou ne saurait être déformé en un point en restant dans le domaine !

C'est sur les domaines de Jordan généralisés que nous intégrerons naturellement les fonctions mesurables réelles ou complexes bornées de deux variables (x, y) .

1.7.3 Le cas de l'espace

Les domaines quarrables de l'espace \mathbf{R}^3 que nous considérerons sont les domaines bornés connexes limités par surface fermée S sans point double, cette surface se présentant comme union des supports de nappes paramétrées de classe C^1 .

Une *nappe paramétrée de classe C^1* est par définition une application

$$\Phi : (u, v) \rightarrow (\varphi_1(u, v), \varphi_2(u, v), \varphi_3(u, v))$$

d'un domaine de Jordan fermé D de \mathbf{R}^2 , à valeurs dans \mathbf{R}^3 , de classe C^1 dans D (y compris jusqu'au bord), telle que le vecteur

$$\frac{\partial \vec{\Phi}}{\partial u} \wedge \frac{\partial \vec{\Phi}}{\partial v}$$

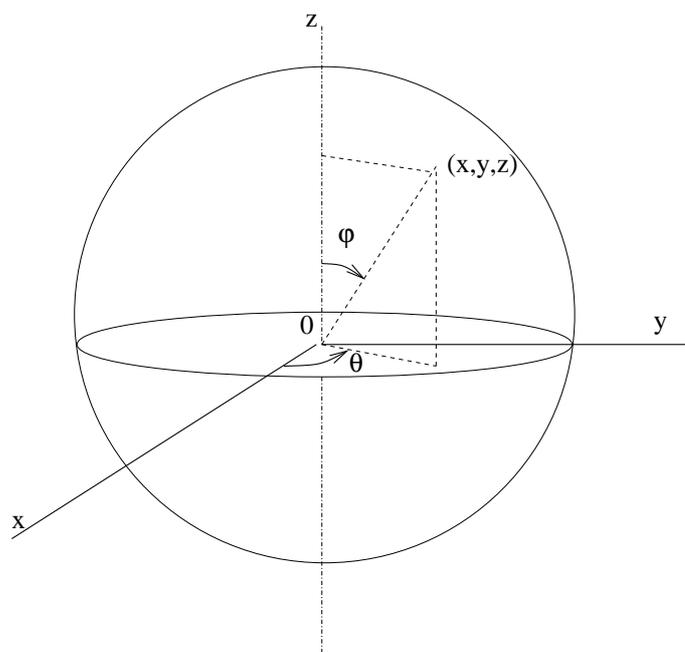
ne s'annule pas dans D ; le support d'une telle nappe est l'image par Φ de D .

Un tel domaine de \mathbf{R}^3 peut avoir des trous car la surface S qui le borde n'est pas nécessairement connexe. Les exemples typiques sont la sphère pleine ou la sphère avec des cavités sphériques, ou le “pneu” plein (ou avec des cavités). Dans le premier cas, le bord externe de A est la sphère (voir la figure 1.11), dans le second cas, la surface S est la chambre à air gonflée, dit *tore*.

La surface de la sphère S_r de centre l'origine et de rayon r se paramètre avec les deux angles d'Euler, la *longitude* θ et la *colatitude* φ (voir la figure 1.11), suivant les formules :

$$\begin{aligned} x &= r \sin \varphi \cos \theta \\ y &= r \sin \varphi \sin \theta \\ z &= r \cos \varphi, \end{aligned}$$

avec $\theta \in [0, 2\pi]$ et $\varphi \in [0, \pi]$.

FIG. 1.11 – La sphère de rayon r paramétrée par les angles d'Euler

L'exemple du tore est donné sur la figure 1.12 ; pour le paramétrer, on se donne deux nombres a et b strictement positifs (avec $a > b$ et l'on paramètre la surface par

$$\begin{aligned}x &= (a + b \sin \varphi) \cos \theta \\y &= (a + b \sin \varphi) \sin \theta \\z &= b \cos \varphi,\end{aligned}$$

avec $\theta \in [0, 2\pi]$ et $\varphi \in [0, 2\pi]$ (θ correspond au repérage en “longitude”, φ à un repérage le long des cercles méridiens verticaux) ; on pourra par exemple s'entraîner à représenter θ et φ pour un point donné du tore sur la figure 1.12.

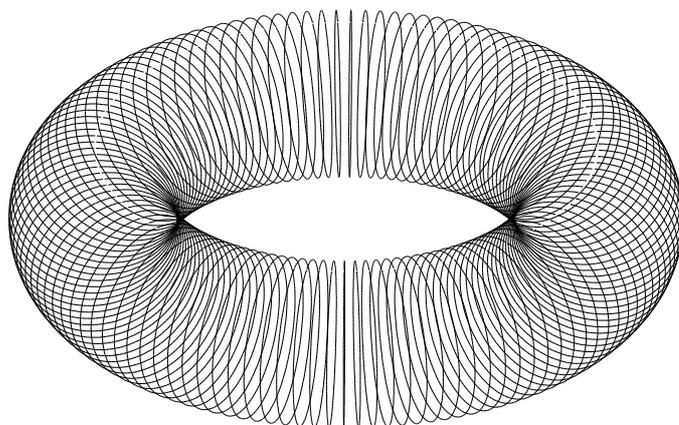


FIG. 1.12 – Le tore paramétré

1.8 Comment ramener le calcul d'intégrales multiples au calcul d'intégrales simples ?

1.8.1 Extension de la notion d'intégrale aux fonctions mesurables éventuellement non bornées

L'intégration des fonctions mesurables positives (mais non nécessairement bornées) sur un domaine quarrable A de \mathbb{R}^n ne pose pas de problème si l'on convient que

$$\int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n \in [0, +\infty].$$

il se peut d'ailleurs fort bien que l'intégrale d'une fonction positive non bornée soit finie comme le montre l'exemple de $f(t) = 1/\sqrt{t}$ sur $[0, 1]$ car

$$\int_{[0,1]} \frac{dt}{\sqrt{t}} = \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^1 \frac{dt}{\sqrt{t}} = \lim_{\epsilon \rightarrow 0} [2\sqrt{t}]_{\epsilon}^1 = 2.$$

Si maintenant f est une fonction mesurable à valeurs complexes sur A , on remarque que f s'écrit

$$f = (\operatorname{Re} f)^+ - (\operatorname{Re} f)^- + i((\operatorname{Im} f)^+ - (\operatorname{Im} f)^-),$$

où les quatre fonctions

$$\begin{aligned} (\operatorname{Re} f)^+ &:= \sup(\operatorname{Re} f, 0) \\ (\operatorname{Re} f)^- &:= \sup(-\operatorname{Re} f, 0) \\ (\operatorname{Im} f)^+ &:= \sup(\operatorname{Im} f, 0) \\ (\operatorname{Im} f)^- &:= \sup(-\operatorname{Im} f, 0) \end{aligned}$$

sont des fonctions positives. On dit que f est intégrable sur A si ces quatre fonctions le sont et l'on note alors

$$\int \cdots \int_A f = \int \cdots \int_A (\operatorname{Re} f)^+ - \int \cdots \int_A (\operatorname{Re} f)^- + i \int \cdots \int_A (\operatorname{Im} f)^+ - i \int \cdots \int_A (\operatorname{Im} f)^-.$$

1.8.2 Intégration des fonctions mesurables sur un domaine quarrable éventuellement non borné

Une fonction définie sur un ensemble quarrable non borné A , mesurable, positive, est dite intégrable sur A si

$$\sup_{N \in \mathbb{N}^*} \left(\int \cdots \int_{A \cap \{\|x\| \leq N\}} f(x_1, \dots, x_n) dx_1 \cdots dx_n \right) < +\infty;$$

l'intégrale de f sur A est alors

$$\int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n := \lim_{N \rightarrow +\infty} \left(\int \cdots \int_{A \cap \{\|x\| \leq N\}} f(x_1, \dots, x_n) dx_1 \cdots dx_n \right).$$

Une fonction définie sur un domaine quarrable A (éventuellement non borné), mesurable, à valeurs complexes, est dite intégrable sur A si les quatre fonctions $(\operatorname{Re} f)^{\pm}$, $(\operatorname{Im} f)^{\pm}$ le sont ; l'intégrale de f sur A est alors donnée par la formule

$$\int \cdots \int_A f = \int \cdots \int_A (\operatorname{Re} f)^+ - \int \cdots \int_A (\operatorname{Re} f)^- + i \int \cdots \int_A (\operatorname{Im} f)^+ - i \int \cdots \int_A (\operatorname{Im} f)^-.$$

Par exemple, la fonction

$$(x, y) \rightarrow \exp(-x^2 - y^2)$$

est intégrable sur \mathbb{R}^2 et l'on a

$$\begin{aligned} \int \int_{\mathbb{R}^2} \exp(-x^2 - y^2) dx dy &= \left(\int_{\mathbb{R}} \exp(-x^2) dx \right) \times \left(\int_{\mathbb{R}} \exp(-y^2) dx \right) \\ &= \left(\int_{\mathbb{R}} \exp(-x^2) dx \right)^2 = \pi \end{aligned}$$

(on le justifiera un peu plus loin dans le cours).

1.8.3 Le théorème de Fubini

L'approche discrète proposée à la section 1.5 suggère le résultat capital suivant, permettant de ramener le calcul des intégrales multiples à celui des intégrales simples :

Théorème 1.1 (théorème de Fubini) *Soit A un sous-ensemble quarrable de $\mathbb{R}^p \times \mathbb{R}^q$ (borné ou non) et f une fonction mesurable sur A , à valeurs complexes, intégrable sur A . Alors, pour calculer l'intégrale de f sur A , on peut utiliser l'une des deux formules suivantes*

$$\int \cdots \int_A f(x_1, \dots, x_p, y_1, \dots, y_q) dx dy = \int_{\{\mathbf{x}; A_{\mathbf{x}} \neq \emptyset\}} \left(\int \cdots \int_{A_{\mathbf{x}}} f(\mathbf{x}_1, \dots, \mathbf{x}_p, y_1, \dots, y_q) dy \right) d\mathbf{x},$$

ou bien

$$\int \cdots \int_A f(x_1, \dots, x_p, y_1, \dots, y_q) dx dy = \int_{\{\mathbf{y}; A^{\mathbf{y}} \neq \emptyset\}} \left(\int \cdots \int_{A^{\mathbf{y}}} f(x_1, \dots, x_p, \mathbf{y}_1, \dots, \mathbf{y}_q) dx \right) d\mathbf{y},$$

où, pour tout $x \in \mathbb{R}^p$,

$$A_{\mathbf{x}} := \{y \in \mathbb{R}^q; (\mathbf{x}, y) \in A\}$$

et

$$A^{\mathbf{y}} := \{x \in \mathbb{R}^p; (x, \mathbf{y}) \in A\}$$

sont les "tranches" horizontales ou verticales de l'ensemble mesurable A .

Le résultat est faux si f n'est pas intégrable sur A , quand bien même les deux autres expressions de l'intégrale calculée par "tranches" existeraient ! Par contre, pour une fonction mesurable positive, les deux formules sont vraies, le résultat pouvant valoir $+\infty$. C'est d'ailleurs ainsi que l'on applique le théorème : on vérifie que f est intégrable en calculant par le biais de l'une ou l'autre des formules

$$\int \cdots \int_A |f(x_1, \dots, x_p, y_1, \dots, y_q)| dx dy$$

(noton que la fonction intégrée est ici positive, ce qui justifie le recours à l'une ou l'autre des formules) et, si l'on trouve un nombre fini, on sait que l'on est autorisé à calculer l'intégrale de f (et non plus de $|f|$) par ce procédé d'intégration par tranches.

Exemple 1.1

Il est toujours important, pour appliquer le théorème de Fubini, de représenter graphiquement le domaine d'intégration, ce qui permet de ne pas commettre d'erreur lorsque l'on représente les "tranches" A_x ou A^y . Sur le domaine

$$\Delta_a := \{(x, y) \in \mathbb{R}^2; x \geq 0, y \geq 0, x + y \leq a\},$$

où a est un nombre strictement positif, domaine que nous avons représenté sur la figure 1.13, on a

$$\begin{aligned} \iint_{\Delta_a} e^{-x-y} dx dy &= \int_0^a \left[\int_0^{a-x} e^{-x-y} dy \right] dx = \int_0^a e^{-x} \left[-e^{-y} \right]_0^{a-x} dx \\ &= \int_0^a e^{-x} (1 - e^{x-a}) dx = \left[-e^{-x} \right]_0^a - a e^{-a} \\ &= 1 - (a+1)e^{-a}. \end{aligned}$$

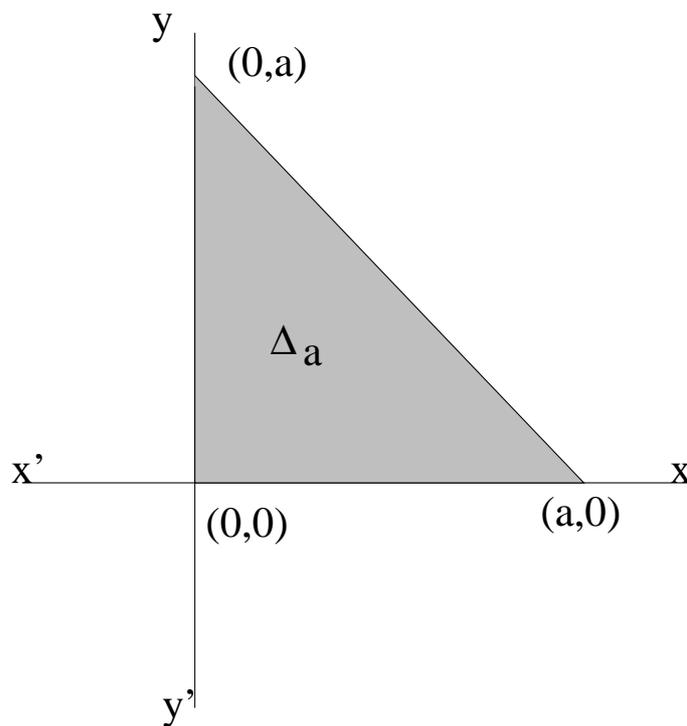


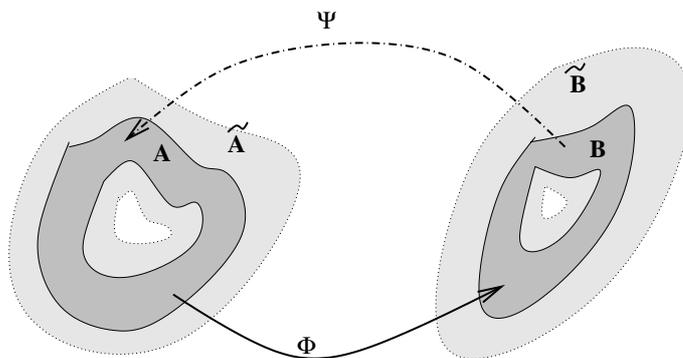
FIG. 1.13 – Le domaine Δ_a

1.9 Le changement de variables dans les intégrales multiples

Outre le théorème de Fubini, le second outil majeur dans le calcul d'intégrales multiples est la *formule de changement de variables*.

Soient A et B deux sous-ensembles quarrables de \mathbb{R}^n , non d'ailleurs nécessairement bornés.

Par définition une *application* Φ de classe C^1 de A dans B est la restriction à A d'une application $\tilde{\Phi}$ définie et de classe C^1 dans un sous-ensemble ouvert \tilde{A} de \mathbb{R}^n contenant A et telle que $\tilde{\Phi}(A) \subset B$ (voir la figure 1.14).

FIG. 1.14 – Application de classe C^1 de A dans B

On rappelle qu'une application $\tilde{\Phi}$ d'un ouvert U de \mathbf{R}^n à valeurs dans \mathbf{R}^n est une application

$$(x_1, \dots, x_n) \rightarrow \begin{pmatrix} \tilde{\Phi}_1(x_1, \dots, x_n) \\ \vdots \\ \tilde{\Phi}_n(x_1, \dots, x_n) \end{pmatrix},$$

où les fonctions composantes $\tilde{\Phi}_k$, $k = 1, \dots, n$, admettent des dérivées partielles $\partial \tilde{\Phi}_k / \partial x_k$, $k = 1, \dots, n$, continues dans U .

On suppose que A et B sont quarrables tous les deux et tels qu'il existe à la fois une application Φ de classe C^1 de A dans B (se prolongeant en $\tilde{\Phi}$ de \tilde{A} dans \mathbf{R}^n) et une application Ψ de classe C^1 de B dans A (se prolongeant en $\tilde{\Psi}$ de \tilde{B} dans \mathbf{R}^n) avec $\Psi \circ \Phi = \text{Id}_A$ et $\Phi \circ \Psi = \text{Id}_B$.

Exemple 1.2 C'est le cas par exemple si $A = [\theta_1, \theta_2] \times [r_1, r_2]$, où $0 \leq \theta_1 < \theta_2 < 2\pi$ et $0 < r_1 < r_2 < +\infty$, et

$$B := \{(x, y) \in \mathbf{R}^2; r_1 < \|(x, y)\| < r_2, \arg(x + iy) \in [\theta_1, \theta_2]\}$$

comme sur la figure 1.15, l'application Φ étant dans ce cas l'application

$$(r, \theta) \rightarrow (r \cos \theta, r \sin \theta).$$

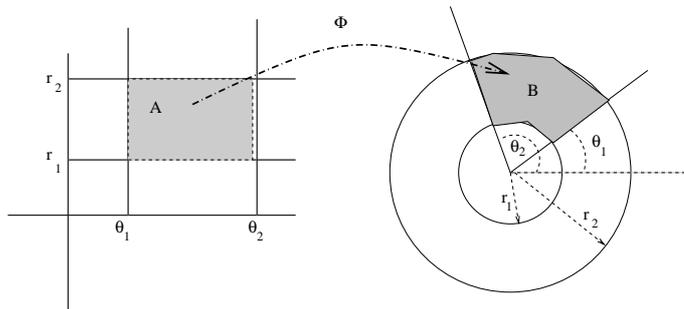


FIG. 1.15 – Un domaine en coordonnées polaires

Supposons que f soit une fonction intégrable dans B . Si $u = (u_1, \dots, u_n)$ est un point de A , l'image par $\tilde{\Phi}$ du pavé $P = [u_1, u_1 + \Delta u_1] \times \dots \times [u_n, u_n + \Delta u_n]$ inclus dans \tilde{A} au voisinage de u est un pavé déformé $\tilde{\Phi}(P)$ de l'ouvert \tilde{B} (voir la figure 1.15 où pour

simplifier nous avons pris $n = 2$), pourvu que les longueurs $\Delta u_1, \dots, \Delta u_n$ des arêtes de P soient assez petites.

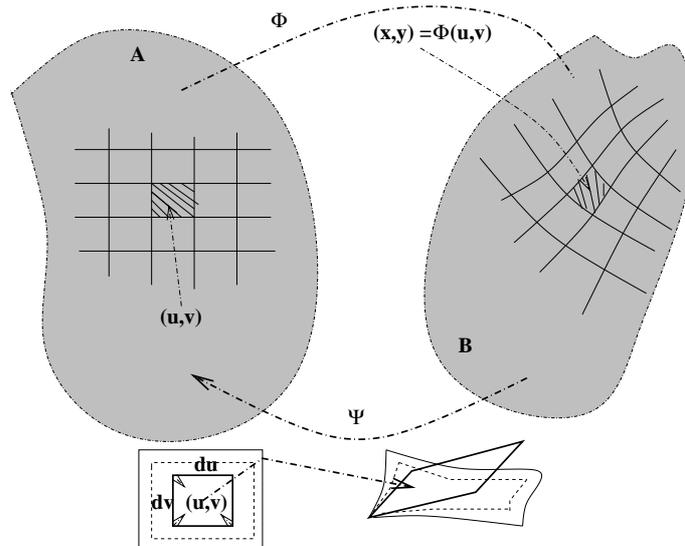


FIG. 1.16 – Déformation de pavés par changement de variables

Comme l'application $\tilde{\Phi}$ est de classe C^1 , on peut l'approcher au voisinage du point (u_1, u_2, \dots, u_n) par une application affine (dite *application affine tangente* à $\tilde{\Phi}$ au point (u_1, \dots, u_n)). Cette application affine tangente est l'application

$$(v_1, \dots, v_n) \rightarrow \tilde{\Phi}(u_1, \dots, u_n) + \begin{pmatrix} \frac{\partial \tilde{\Phi}_1}{\partial u_1}(u) & \cdots & \frac{\partial \tilde{\Phi}_1}{\partial u_n}(u) \\ \vdots & & \vdots \\ \frac{\partial \tilde{\Phi}_n}{\partial u_1}(u) & \cdots & \frac{\partial \tilde{\Phi}_n}{\partial u_n}(u) \end{pmatrix} \bullet \begin{pmatrix} v_1 - u_1 \\ \vdots \\ v_n - u_n \end{pmatrix}. \quad (*)$$

Si (u_1, \dots, u_n) est un point de A , on peut donc supposer que l'image par $\tilde{\Phi}$ du pavé "infinitésimal" $[u_1, u_1 + du_1] \times \cdots \times [u_n, u_n + du_n]$ "colle" de manière infinitésimale à l'image de ce pavé par l'application affine (*); dans le cas $n = 2$, cette image est un parallélogramme (voir en bas de la figure 1.15 où nous avons représenté l'image par $\tilde{\Phi}$ de P et son approximation par un parallélogramme lorsque P devient un pavé infinitésimal du type $[u, u + du] \times [v, v + dv]$), dans le cas $n = 3$, c'est un parallélépipède, etc... On a donc

$$\text{vol}_n(\tilde{\Phi}([u_1, u_1 + du_1] \times \cdots \times [u_n, u_n + du_n])) \simeq \left| \det \begin{pmatrix} \frac{\partial \tilde{\Phi}_1}{\partial u_1}(u) & \cdots & \frac{\partial \tilde{\Phi}_1}{\partial u_n}(u) \\ \vdots & & \vdots \\ \frac{\partial \tilde{\Phi}_n}{\partial u_1}(u) & \cdots & \frac{\partial \tilde{\Phi}_n}{\partial u_n}(u) \end{pmatrix} \right| du_1 \cdots du_n.$$

Il est ainsi évident intuitivement que l'intégrale de f sur B se calcule en "ajoutant" les nombres

$$f(\tilde{\Phi}(u_1, \dots, u_n)) \left| \det \begin{pmatrix} \frac{\partial \tilde{\Phi}_1}{\partial u_1}(u) & \cdots & \frac{\partial \tilde{\Phi}_1}{\partial u_n}(u) \\ \vdots & & \vdots \\ \frac{\partial \tilde{\Phi}_n}{\partial u_1}(u) & \cdots & \frac{\partial \tilde{\Phi}_n}{\partial u_n}(u) \end{pmatrix} \right|$$

pour tous les points (u_1, \dots, u_n) de A .

On peut donc énoncer la règle suivante : si A et B sont deux domaines quarrables de \mathbb{R}^n tels qu'il existe des applications Φ et Ψ de classe C^1 respectivement de A dans

B et de B dans A telles que $\Psi \circ \Phi = \text{Id}_A$ et $\Phi \circ \Psi = \text{Id}_B$, une fonction f mesurable de B dans \mathbb{C} est intégrable sur B si et seulement si la fonction

$$(u_1, \dots, u_n) \rightarrow f(\Phi(u_1, \dots, u_n)) \left| \det \begin{pmatrix} \frac{\partial \Phi_1}{\partial u_1}(u) & \cdots & \frac{\partial \Phi_1}{\partial u_n}(u) \\ \vdots & & \vdots \\ \frac{\partial \Phi_n}{\partial u_1}(u) & \cdots & \frac{\partial \Phi_n}{\partial u_n}(u) \end{pmatrix} \right|$$

est intégrable sur A et l'on a de plus la formule clef :

$$\begin{aligned} & \int \dots \int_B f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int \dots \int_A f(\Phi(u_1, \dots, u_n)) \left| \det \begin{pmatrix} \frac{\partial \Phi_1}{\partial u_1}(u) & \cdots & \frac{\partial \Phi_1}{\partial u_n}(u) \\ \vdots & & \vdots \\ \frac{\partial \Phi_n}{\partial u_1}(u) & \cdots & \frac{\partial \Phi_n}{\partial u_n}(u) \end{pmatrix} \right| du_1 \cdots du_n \end{aligned}$$

(pour simplifier, on a noté encore Φ le prolongement $\tilde{\Phi}$ de Φ à \tilde{A}).

Cette formule est la *formule de changement de variables*. Le déterminant

$$\det \begin{pmatrix} \frac{\partial \Phi_1}{\partial u_1}(u) & \cdots & \frac{\partial \Phi_1}{\partial u_n}(u) \\ \vdots & & \vdots \\ \frac{\partial \Phi_n}{\partial u_1}(u) & \cdots & \frac{\partial \Phi_n}{\partial u_n}(u) \end{pmatrix}$$

dont on prend la valeur absolue (ceci est très important !) dans cette formule est dit *jacobien* du changement de variables Φ .

Exemple 1.3

- Si $n = 2$ et si Φ est l'application

$$(r, \theta) \rightarrow (r \cos \theta, r \sin \theta)$$

correspondant au changement de variables consistant à passer des coordonnées cartésiennes aux coordonnées polaires, on voit que le jacobien de Φ au point courant (r, θ) vaut r . La formule de changement de variables dans ce cas est donc

$$\iint_{\Phi(A)} f(x, y) dx dy = \iint_A f(r \cos \theta, r \sin \theta) r dr d\theta ;$$

il faut juste prendre garde au fait que Φ doit bien réaliser une bijection entre A et son image $\Phi(A)$ (comme dans l'exemple 1.2).

- Si $n = 3$ et si Φ est l'application

$$(r, \theta, \varphi) \rightarrow (r \cos \theta \sin \varphi, r \sin \theta \sin \varphi, r \cos \varphi)$$

correspondant au changement de variables consistant à passer des coordonnées cartésiennes aux coordonnées sphériques ($\theta \in [0, 2\pi[$ et $\varphi \in [0, \pi]$ étant les angles d'Euler repérant le point de la sphère de rayon r comme sur la figure 1.11) on voit que le jacobien de Φ au point courant (r, θ, φ) vaut $-r^2 \sin \varphi$. La formule de changement de variables dans ce cas est donc

$$\iiint_{\Phi(A)} f(x, y, z) dx dy dz = \iiint_A f(r \cos \theta \sin \varphi, r \sin \theta \sin \varphi, r \cos \varphi) r^2 \sin \varphi dr d\theta d\varphi ;$$

il faut encore prendre garde au fait que Φ doit bien réaliser une bijection entre A et son image $\Phi(A)$.

1.10 Chemins paramétrés, arcs géométriques

Un *chemin paramétré continu* γ de \mathbb{R}^n est par définition une application continue d'un intervalle $[a, b]$ de \mathbb{R} , à valeurs dans \mathbb{R}^n ; le *support* de ce chemin paramétré $\gamma : [a, b] \rightarrow \mathbb{R}^n$ est l'image par γ de $[a, b]$. Ce chemin est dit *lacet continu* si $\gamma(a) = \gamma(b)$, *lacet continu simple* si $\gamma(a) = \gamma(b)$ et si de plus la restriction de γ à $[a, b[$ est injective. Dans tous les cas, le point $\gamma(a)$ est dit *origine* du chemin paramétré, le point $\gamma(b)$ *extrémité* de ce chemin.

On appelle *paramétrisation* du chemin paramétré continu $\gamma : [a, b] \rightarrow \mathbb{R}^n$ toute application continue $\tilde{\gamma}$ d'un intervalle $[c, d]$ de \mathbb{R} , à valeurs dans \mathbb{R}^n , telle qu'il existe une bijection strictement croissante φ de $[a, b]$ dans $[c, d]$ avec $\gamma = \tilde{\gamma} \circ \varphi$.

Deux chemins paramétrés continus $\gamma_1 : [a_1, b_1] \rightarrow \mathbb{R}^n$ et $\gamma_2 : [a_2, b_2] \rightarrow \mathbb{R}^n$ définissent le même *arc géométrique continu* si et seulement si il existe une bijection continue φ entre $[a_1, b_1]$ et $[a_2, b_2]$ telle que $\gamma_2 \circ \varphi = \gamma_1$; notons que deux chemins paramétrés continus définissant le même arc géométrique ne constituent pas forcément deux paramétrisations du même chemin continu paramétré (une bijection entre deux intervalles peut être tout aussi bien strictement croissante que strictement décroissante). Si les deux chemins paramétrés $\gamma_1 : [a, b] \rightarrow \mathbb{R}^n$ et $\gamma_2 : [c, d] \rightarrow \mathbb{R}^n$ sont deux paramétrisations du même chemin paramétré continu, on dit que γ_1 et γ_2 définissent le même *arc géométrique continu orienté*.

On se restreindra souvent aux *chemins paramétrés C^1 par morceaux de \mathbb{R}^n* : un tel chemin est par définition une application γ d'un intervalle $[a, b]$ de \mathbb{R} , à valeurs dans \mathbb{R}^n , continue, telle qu'il existe une subdivision $t_0 = a < t_1 < \dots < t_N = b$ de manière à ce que la restriction de γ à $[t_j, t_{j+1}]$ (pour $j = 0, \dots, N-1$) soit une application de classe C^1 sur $[t_j, t_{j+1}]$ (c'est-à-dire se prolonge en une application de classe C^1 dans un intervalle ouvert $]t_j - \epsilon, t_{j+1} + \epsilon[$ un petit peu plus gros). Le *support* d'un tel chemin paramétré C^1 par morceaux $\gamma : [a, b] \rightarrow \mathbb{R}^n$ est par définition l'ensemble $\gamma([a, b])$.

On appelle *paramétrisation* d'un chemin C^1 par morceaux $\gamma : [a, b] \rightarrow \mathbb{R}^n$ toute application C^1 par morceaux $\tilde{\gamma} : [c, d] \rightarrow \mathbb{R}^n$ d'un intervalle $[c, d]$ de \mathbb{R} , à valeurs dans \mathbb{R}^n , telle qu'il existe une application strictement croissante et C^1 par morceaux (donc continue) φ de $[a, b]$ dans $[c, d]$ avec $\gamma = \tilde{\gamma} \circ \varphi$.

Deux chemins paramétrés de classe C^1 par morceaux $\gamma_1 : [a_1, b_1] \rightarrow \mathbb{R}^n$ et $\gamma_2 : [a_2, b_2] \rightarrow \mathbb{R}^n$ définissent le même *arc géométrique de classe C^1 par morceaux* si et seulement si il existe une bijection φ de classe C^1 par morceaux entre $[a_1, b_1]$ et $[a_2, b_2]$ telle que $\gamma_2 \circ \varphi = \gamma_1$; notons que deux chemins paramétrés continus définissant le même arc géométrique ne constituent pas forcément deux paramétrisations du même chemin continu paramétré (une bijection entre deux intervalles peut être tout aussi bien strictement croissante que strictement décroissante). Si les deux chemins de classe C^1 par morceaux $\gamma_1 : [a_1, b_1] \rightarrow \mathbb{R}^n$ et $\gamma_2 : [a_2, b_2] \rightarrow \mathbb{R}^n$ sont deux paramétrisations du même chemin paramétré de classe C^1 par morceaux, on dit que γ_1 et γ_2 définissent le même *arc géométrique C^1 par morceaux orienté*.

Le support d'un chemin paramétré continu peut présenter des pathologies par rapport à l'idée intuitive que l'on se fait de son tracé! Il existe par exemple un chemin continu $\gamma : [0, 1] \rightarrow [0, 1] \times [0, 1]$ surjectif (c'est la *courbe de Peano*), donc "noircissant" le carré $[0, 1] \times [0, 1]$. Le flocon de Von Koch est, lui, un exemple d'arc géométrique de \mathbb{R}^2 présentant une *structure fractale* : l'idée de la construction est de

partir du segment $[0, 1]$, de le subdiviser en trois segments égaux, puis de remplacer dans le trajet “direct” $[0, 1]$ le tronçon central $[1/3, 2/3]$ par le “détour” consistant à relier $(1/3, 0)$ à $(2/3, 0)$ en suivant le bord du triangle équilatéral de côté $[1/3, 2/3]$ situé dans le demi-plan $\{y \geq 0\}$, enfin en réitérant indéfiniment l’opération ci-dessus sur chacun des tronçons rectilignes au fur et à mesure qu’ils apparaissent dans le tracé (voir la figure 1.17 ci-dessous).

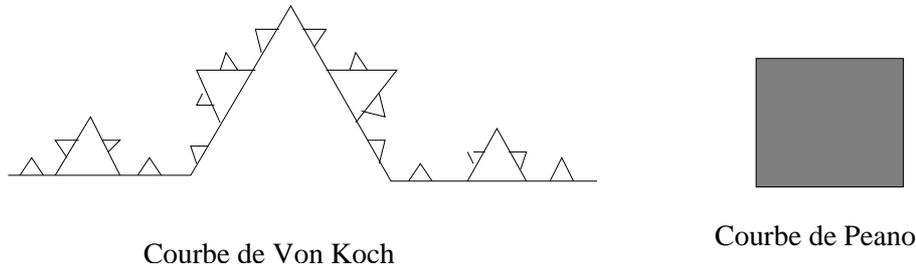


FIG. 1.17 – Deux arcs géométriques continus “pathologiques”

Les pathologies évoquées plus haut montrent qu’il y a ambiguïté à parler de “longueur” d’un arc géométrique continu ; voici à l’appui trois exemples :

- La mesure n -dimensionnelle d’un arc géométrique de classe C^1 par morceaux $\gamma : [a, b] \rightarrow \mathbb{R}^n$ est nulle si $n > 1$ (de par le calcul même du volume des ensembles quarrables dans \mathbb{R}^n) : en effet, le support de γ peut être recouvert par une union de pavés dont la somme des volumes est arbitrairement petite !
- Pour la courbe de Peano, par contre, le support a pour volume 2-dimensionnel la surface du carré $[0, 1] \times [0, 1]$, soit 1 ;
- Si l’on mesure en suivant avec notre mètre le tracé du chemin $\gamma : [a, b] \rightarrow \mathbb{R}^n$, la mesure peut être infinie, comme c’est le cas avec le flocon de Von Koch.

La dernière solution proposée ci-dessus pour mesurer un arc géométrique (en l’occurrence la courbe de Von Koch) s’avère de fait être celle que nous retiendrons.

Si $\gamma : [a, b] \rightarrow \mathbb{R}^n$ est un arc géométrique continu de \mathbb{R}^n , cet arc est dit *rectifiable* si et seulement si la borne supérieure, pour toutes les subdivisions

$$a = t_0 < t_1 < \dots < t_{N-1} < t_N = b$$

des quantités

$$\sum_{j=0}^{N-1} \|\gamma(t_{j+1}) - \gamma(t_j)\|$$

(la norme étant ici la norme euclidienne dans \mathbb{R}^n) est un nombre fini. Ce nombre est alors appelé *longueur* de l’arc géométrique continu $\gamma : [a, b] \rightarrow \mathbb{R}^n$. Cette définition a un sens puisque l’on peut s’assurer que cette expression de la longueur ne dépend pas du choix du représentant $\gamma : [a, b] \rightarrow \mathbb{R}^n$ de l’arc géométrique.

Ni la courbe de Peano, ni la courbe de Von Koch ne correspondent à des arcs géométriques continus rectifiables.

En revanche, un arc géométrique $\gamma : [a, b] \rightarrow \mathbb{R}^n$ C^1 par morceaux est rectifiable et l’on calcule sa longueur en notant que si t_0 est un point où γ est dérivable et de vecteur dérivé non nul, l’intervalle infinitésimal $[t_0, t_0 + dt]$ est transformé par γ en

l'intervalle de \mathbb{R}^n joignant les points $\gamma(t_0)$ et $\gamma(t_0) + \gamma'(t_0) \times dt$, intervalle qui est de longueur $\|\gamma'(t_0)\| dt$. En additionnant ces longueurs infinitésimales, on trouve ainsi que

$$\text{longueur de } \gamma := \int_a^b \|\gamma'(t)\| dt.$$

Exemple 1.4 Si $\gamma : t \in [0, 2\pi] \rightarrow (a \cos t, b \sin t)$, avec $a > b > 0$, la longueur de l'arc géométrique ainsi représenté (correspondant au périmètre d'une ellipse de grand axe a , de petit axe b , parcouru une fois) vaut

$$\int_0^{2\pi} \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} dt = 4\sqrt{a^2 + b^2} \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 t} dt,$$

où $k = b/\sqrt{a^2 + b^2}$, intégrale dite *intégrale elliptique* dont le calcul *via* une formule explicite s'avère difficile, voire impossible; cet exemple souligne qu'il n'est pas en général aisé (du fait de la présence quasi obligée de radicaux) de calculer explicitement (en l'exprimant grâce à une formule close et non au terme d'un calcul numérique approché) la longueur d'un arc géométrique de classe C^1 par morceaux.

1.11 Intégrale curviligne d'une 1-forme le long d'un chemin paramétré C^1 par morceaux

Si U est un sous-ensemble ouvert de \mathbb{R}^n , une 1-forme différentielle continue sur U consiste en la donnée, pour chaque point (x_1, \dots, x_n) de U , d'une application linéaire $\omega(x_1, \dots, x_n)$ de \mathbb{R}^n dans \mathbb{R} , ce de manière à ce que la forme linéaire $\omega(x_1, \dots, x_n)$ dépende continuellement des variables d'espace (x_1, \dots, x_n) dans le sous-ensemble ouvert U .

Une telle application linéaire peut se réaliser de la manière suivante :

$$\omega(x_1, \dots, x_n) : (h_1, \dots, h_n) \rightarrow \omega(x)[h] = F_1(x)h_1 + \dots + F_n(x)h_n,$$

où $\vec{F}(x) = (F_1(x), \dots, F_n(x))$ est un vecteur de \mathbb{R}^n dépendant du point courant x de manière continue; le physicien appelle en général une application du type $x \rightarrow \vec{F}(x)$ un *champ de forces* continu dans U , une 1-forme différentielle continue dans U devenant donc dans cette terminologie *l'action d'un champ de forces continu* dans l'ouvert U . Si l'application $x \rightarrow \vec{F}(x)$ est de classe C^k dans U pour $k \in \mathbb{N}^*$, on dira que la 1-forme correspondante sera une forme de classe C^k dans U .

On convient, si ω est la 1-forme différentielle correspondant à l'action du champ de forces continu \vec{F} , de noter

$$\omega(x_1, \dots, x_n) := \sum_{j=1}^n F_j(x_1, \dots, x_n) dx_j;$$

ainsi dx_j est l'application linéaire de \mathbb{R}^n dans \mathbb{R} correspondant à la prise de la j -ème coordonnée.

Le *travail du champ de forces* continu \vec{F} (défini dans un sous-ensemble ouvert U de \mathbb{R}^n) le long d'un chemin paramétré de classe C^1 par morceaux γ dont le support est inclus dans U fait apparaître un concept intégral nouveau, celui d'*intégrale curviligne*, concept dans lequel se trouve impliqué cette fois le chemin paramétré

(c'est-à-dire l'arc géométrique orienté) et non plus l'arc géométrique (comme c'était le cas dans le calcul de la longueur au paragraphe précédent).

Par définition, l'intégrale curviligne de la 1-forme continue ω le long du chemin paramétré C^1 par morceaux $\gamma : [a, b] \rightarrow U$ est l'intégrale

$$\int_{\gamma} \omega = \sum_{j=0}^{N-1} \int_{t_j}^{t_{j+1}} \omega(\gamma|_{[t_j, t_{j+1}]}(t))[\gamma'|_{[t_j, t_{j+1}]}(t)] dt,$$

si $a = t_0 < t_1 < \dots < t_{N-1} < t_N = b$ désigne la subdivision correspondant à γ . Comme

$$\omega(\gamma(t_0))[\gamma'(t_0)] = \sum_{j=1}^n F_j(\gamma(t_0))\gamma'_j(t_0)$$

au point courant t_0 de $[a, b]$, la quantité

$$\omega(\gamma(t_0))[\gamma'(t_0)] dt$$

représente le travail infinitésimal du champ de vecteurs \vec{F} le long du déplacement sur le support de l'arc paramétré correspondant à la variation du paramètre t dans $[t_0, t_0 + dt]$. Ainsi, l'intégrale curviligne

$$\int_{\gamma} \omega$$

correspond-t-elle au *travail* du champ de forces \vec{F} le long du déplacement le long du support de γ , déplacement matérialisé par le parcours balisé par un arc géométrique orienté. Cette intégrale curviligne reste d'ailleurs inchangée si l'on remplace γ par une autre paramétrisation du même chemin paramétré C^1 par morceaux (comme le montre la formule de changement de variables dans les intégrales).

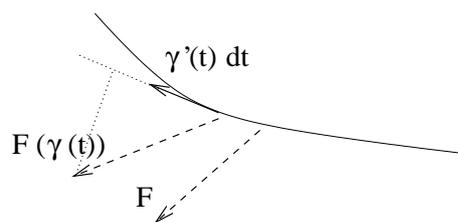


FIG. 1.18 – Intégrale curviligne et travail

1.12 La formule de Green-Riemann dans le plan

Soit A un domaine de Jordan généralisé fermé du plan \mathbb{R}^2 dont le bord est constitué d'une union finie de supports de lacets simples $\gamma_1, \dots, \gamma_N$, comme sur la figure 1.19.

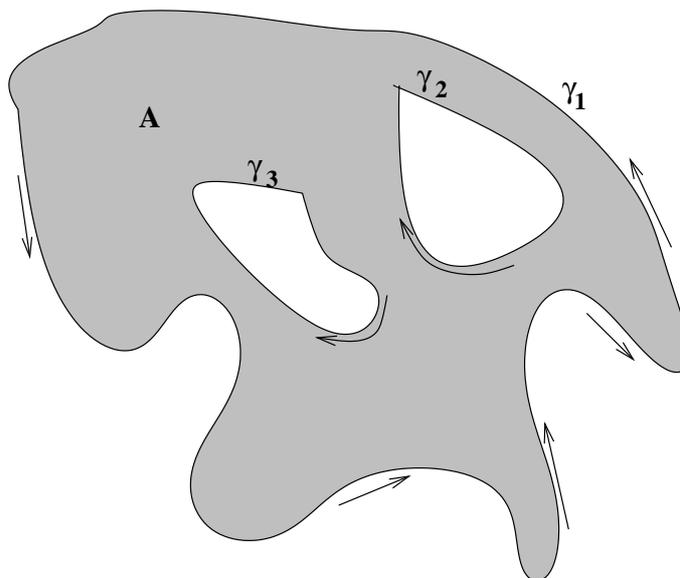


FIG. 1.19 – Un domaine de Jordan “généralisé”

Les bords $\gamma_1, \dots, \gamma_N$ de A peuvent être considérés comme les supports non seulement d’arcs géométriques de classe C^1 par morceaux, mais d’arcs géométriques orientés de classe C^1 par morceaux, en convenant d’une orientation sur chacun d’eux : on décide que chaque arc γ_j est parcouru une seule fois et que tout au long du parcours, l’on tient le domaine A à main gauche (règle du bonhomme d’Ampère). Les lacets correspondant au bord externe sont donc parcourus une fois dans le sens trigonométrique, ceux correspondant au bord interne une fois dans le sens inverse du sens trigonométrique (sens des aiguilles d’une montre). Ce choix d’orientation étant fait, on peut considérer les lacets $\gamma_1, \dots, \gamma_N$ bordant A comme les supports de chemins paramétrés de classe C^1 (que pour simplifier on notera aussi $\gamma_1, \dots, \gamma_N$).

Si $\omega = Pdx + Qdy$ est une 1-forme de classe C^1 au voisinage de A , on a l’importante *formule de Green-Riemann*, dont on verra plus loin qu’elle est l’incarnation en dimension deux du théorème fondamental de l’analyse mentionné en section 1.4. Cette formule s’énonce ainsi :

$$\sum_{j=1}^N \int_{\gamma_j} \omega = \iint_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy$$

(attention au respect des orientations pour les chemins paramétrés γ_j , c’est bien sûr essentiel!). En particulier, la surface de A s’obtient comme

$$\text{vol}_2(A) = \sum_{j=1}^N \int_{\gamma_j} x dy = - \sum_{j=1}^N \int_{\gamma_j} y dx = \frac{1}{2} \sum_{j=1}^N \int_{\gamma_j} (x dy - y dx).$$

Voici une esquisse de preuve pour la formule de Green-Riemann : on remarque tout d’abord que c’est très facile lorsque A est le triangle élémentaire de sommets $(0, 0)$, $(1, 0)$, $(0, 1)$ que l’on appelle aussi 2-simplexe élémentaire (voir la figure 1.20).

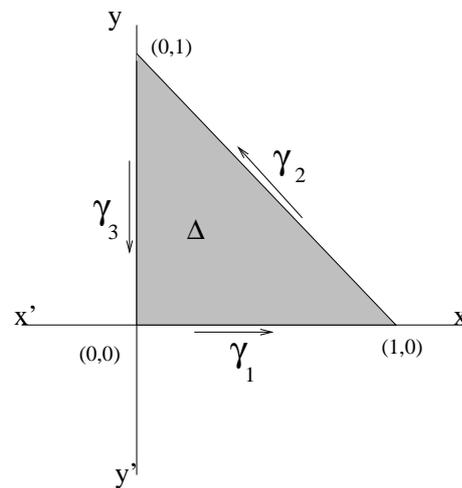


FIG. 1.20 – Le cas du triangle élémentaire

Le calcul de l'intégrale de Pdx sur les chemins paramétrés correspondant aux trois côtés du bord (orientés comme sur la figure) donne respectivement (pour $\gamma_1, \gamma_2, \gamma_3$ comme indiqué sur la figure 1.20)

$$\int_0^1 P(t, 0)dt, - \int_0^1 P(t, 1-t)dt, 0$$

et l'on voit que la somme de ces trois nombres vaut bien (si l'on utilise le théorème de Fubini)

$$- \int_0^1 (P(t, 1-t) - P(t, 0))dt = - \iint_{\Delta} \frac{\partial P}{\partial y} dx dy,$$

le calcul étant en tout point semblable lorsque l'on remplace la forme Pdx par la forme Qdy .

Ensuite, on effectue le calcul lorsque A est l'image de Δ par une application $\Phi = (\Phi_1, \Phi_2)$ de classe C^1 au voisinage de Δ , à valeurs dans \mathbb{R}^2 , et dont le jacobien

$$\text{jac}(\Phi) = \begin{vmatrix} \frac{\partial \Phi_1}{\partial x} & \frac{\partial \Phi_1}{\partial y} \\ \frac{\partial \Phi_2}{\partial x} & \frac{\partial \Phi_2}{\partial y} \end{vmatrix}$$

reste strictement positif sur Δ ; le nouveau domaine A se présente alors comme un "triangle déformé" de sommets a, b, c , comme sur la figure 1.21, la condition sur le jacobien de Φ impliquant que l'orientation dans le sens trigonométrique du bord de Δ se transporte *via* Φ en orientation dans le sens trigonométrique du bord de A (on dit que Φ *respecte l'orientation du plan*, si $\text{jac}(\Phi)$ était resté strictement négatif sur Δ , on aurait dit qu'elle *renverse l'orientation du plan*).

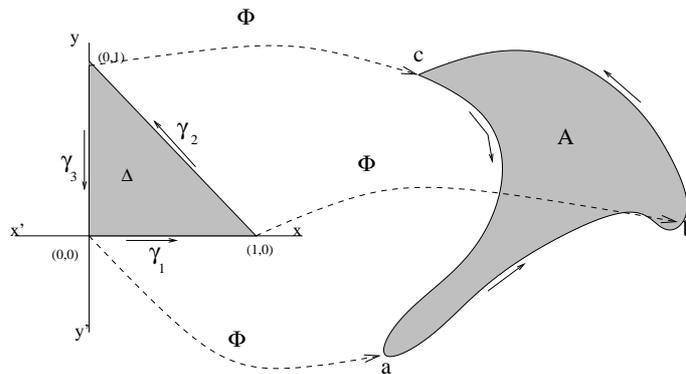


FIG. 1.21 – Le cas du triangle “déformé”

La formule de Green Riemann dans le cas d'un tel domaine A se déduit alors de la formule de Green-Riemann dans le domaine Δ en utilisant la formule de changement de variables à la fois pour les intégrales doubles et pour les intégrales simples (on transforme ainsi les deux membres de l'égalité de Green-Riemann).

Pour prouver le cas général (ou tout au moins esquisser une preuve), on remarque qu'au moins heuristiquement, il est possible de “triangler” A en une mosaïque de triangles curvilignes fermés A_j d'intérieurs disjoints, comme sur la figure 1.22 (deux triangles A_j d'intersection non vide ont soit en commun un seul point qui est sommet de chacun des triangles, soit en commun une seule arête, entière, qui est arête commune des deux triangles).

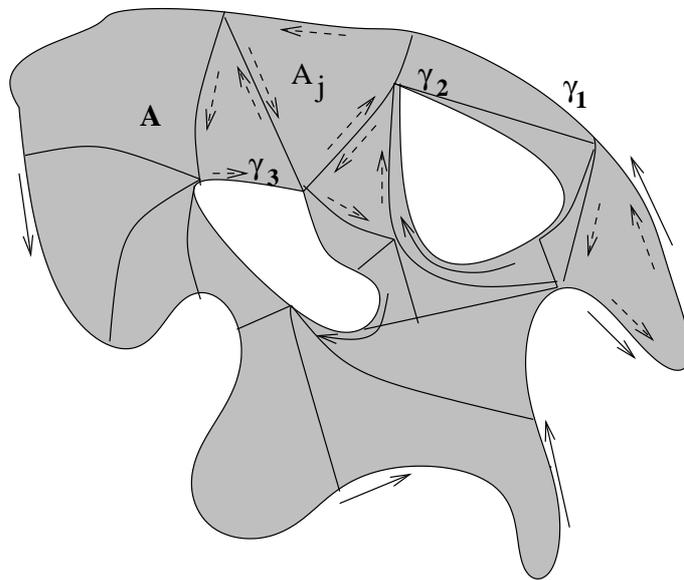


FIG. 1.22 – Un domaine de Jordan “généralisé” triangulé

L'orientation du bord de A induit en cascade des orientations sur les bords des triangles curvilignes A_j , orientations qui sont telles que les orientations sur des arêtes communes à deux triangles A_j se détruisent (voir la figure). La formule de Green-Riemann pour A s'obtient alors en ajoutant les formules de Green-Riemann pour les A_j .

Remarque 1.1 (une première interprétation physique préparant le terrain ultérieurement)

Si (P, Q) est considéré comme un champ de forces dans le plan (rapporté au repère orthonormé direct $(0, \vec{i} = (1, 0), \vec{j} = (0, 1))$) au voisinage du domaine plan A (que l'on peut voir comme une plaque métallique infiniment mince placée dans le plan $\{z = 0\}$ de l'espace \mathbb{R}^3), le vecteur

$$\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \vec{i} \wedge \vec{j}$$

représente le *rotationnel* du champ de vecteurs $P\vec{i} + Q\vec{j} + 0\vec{i} \wedge \vec{j}$ (considéré comme champ de vecteurs dans \mathbb{R}^3 , où $\vec{i} \wedge \vec{j} = (0, 0, 1)$). L'intégrale double

$$\iint_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy$$

peut alors s'interpréter comme le *flux* du rotationnel du champ $P\vec{i} + Q\vec{j} + 0\vec{i} \wedge \vec{j}$ à travers la plaque métallique infiniment fine A , considérée cette fois comme une surface dans \mathbb{R}^3 , le flux étant calculé avec la convention que la normale à la plaque est donnée par le vecteur $\vec{k} = \vec{i} \wedge \vec{j}$. On verra que ceci s'avère être un cas particulier de la formule de Stokes, où la plaque "plane" A sera remplacée par le support de ce que l'on appellera une nappe paramétrée (c'est-à-dire, intuitivement, une plaque originellement plane mais que l'on a ensuite "tordu" dans l'espace à trois dimensions).

La circulation du champ de forces $P\vec{i} + Q\vec{j} + 0\vec{i} \wedge \vec{j}$ le long du bord (convenablement orienté) de la plaque plane A est égale au flux du rotationnel de ce champ à travers cette même plaque A , vue cette fois dans \mathbb{R}^3 , le flux étant calculé avec la convention que la normale à la surface est dirigée vers le haut (c'est-à-dire par le vecteur $(0, 0, 1) = \vec{i} \wedge \vec{j}$).

Remarque 1.2 (une seconde interprétation physique préparant le terrain ultérieurement)

On peut aussi considérer A comme un domaine plan fermé dont le bord (cette fois de dimension 1, c'est une union de supports de lacets C^1 par morceaux) peut être orienté comme on oriente une surface dans \mathbb{R}^3 en convenant d'une orientation pour la normale au bord (mais en transposant ceci dans \mathbb{R}^2 cette fois). La normale que l'on décide de choisir est la normale pointant vers l'extérieur du domaine (figure 1.23).

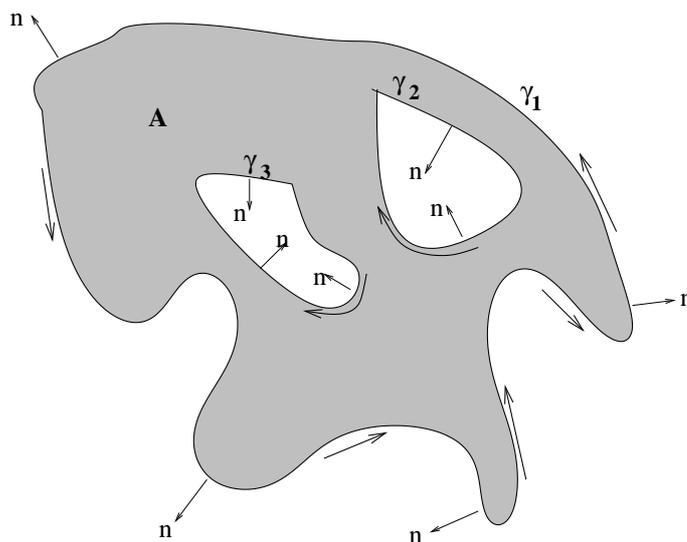


FIG. 1.23 – Le calcul du flux sortant d'un champ de vecteurs au bord d'un domaine fermé en dimension deux

Si $\gamma_j : [a_j, b_j] \rightarrow \mathbb{R}^2$ correspond à l'un des chemins paramétrés en jeu dans la formule de Green-Riemann pour A et si $\gamma'_j(t) = \|\gamma'_j(t)\|(\cos(\theta_j(t)), \sin(\theta_j(t))) \neq 0$ avec pour $t \in [a_j, b_j]$, alors la

normale extérieure à A au point $\gamma_j(t)$ est (si ce n'est pas un point anguleux du bord) dirigée par le vecteur $(\sin(\theta_j(t)), -\cos(\theta_j(t)))$. On voit alors que, si \vec{F} est un champ de forces dans le plan de classe C^1 au voisinage de A ,

$$\sum_{j=1}^N \int_{a_j}^{b_j} \|\gamma'_j(t)\| (P(\gamma_j(t)) \sin(\theta_j(t)) - Q(\gamma_j(t)) \cos(\theta_j(t))) dt = \sum_{j=1}^N \int_{\gamma_j} (P dy - Q dx)$$

Mais, comme $\|\gamma'_j(t)\| dt$ représente la mesure de l'élément infinitésimal de longueur sur la portion du bord de A correspondant au lacet géométrique γ_j et que $P(\gamma_j(t)) \sin(\theta_j(t)) - Q(\gamma_j(t)) \cos(\theta_j(t))$ représente le produit scalaire $\langle \vec{F}, n_{\text{ext}} \rangle$ du champ de vecteurs \vec{F} avec la normale extérieure de norme 1 au point courant du bord, la quantité

$$\sum_{j=1}^N \int_{a_j}^{b_j} \|\gamma'_j(t)\| (P(\gamma_j(t)) \sin(\theta_j(t)) - Q(\gamma_j(t)) \cos(\theta_j(t))) dt$$

représente physiquement le flux sortant du champ de vecteurs \vec{F} à travers le bord de A (A étant cette fois pensé dans \mathbb{R}^2). On note ce flux

$$\int_{\partial A} \langle \vec{F}, n_{\text{ext}} \rangle d\sigma$$

et la formule de Green-Riemann nous donne en dimension deux ce qui deviendra pour nous plus loin la *formule de Green-Ostrogradski* en dimension trois, à savoir

$$\int_{\partial A} \langle \vec{F}, n_{\text{ext}} \rangle d\sigma = \iint_A \operatorname{div}(\vec{F}) dx dy,$$

la *divergence* du champ $P\vec{i} + Q\vec{j}$ étant par définition

$$\operatorname{div}(\vec{F}) = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}.$$

1.13 Formes exactes et fermées

Si $\omega = Pdx + Qdy$ est une 1-forme de classe C^1 dans un ouvert U de \mathbb{R}^2 , on dit que ω est *fermée* dans U si l'on a identiquement dans U l'identité

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \equiv 0$$

(ou, ce qui revient au même, si le rotationnel du champ de forces correspondant est identiquement nul dans U).

La même définition vaut pour les 1-formes de classe C^1 dans un ouvert U de \mathbb{R}^3 ; une forme

$$P_1 dx_1 + \cdots + P_n dx_n$$

de classe C^1 dans un ouvert U de \mathbb{R}^n est dite *fermée* si et seulement si

$$\frac{\partial P_i}{\partial x_j} = \frac{\partial P_j}{\partial x_i}$$

pour tout couple (i, j) de $\{1, \dots, n\} \times \{1, \dots, n\}$. Dans le cas $n = 3$, on retrouve le fait que la forme $Pdx + Qdy + Rdz$ induite par l'action du champ de forces de classe C^1 $\vec{F} = (P, Q, R)$ est fermée dans U si et seulement si le rotationnel de ce champ est identiquement nul dans U .

Parmi les 1-formes de classe C^1 fermées dans un ouvert U de \mathbb{R}^n , il y a celles induites par un champ de forces qui se présente comme le gradient d'une fonction de classe C^2 de U dans \mathbb{R} (on dit alors que le champ *dérive d'un potentiel réel*). Ceci résulte du lemme de Schwarz qui assure que si V est une fonction de classe C^2 dans un ouvert U de \mathbb{R}^n ,

$$\frac{\partial^2 V}{\partial x_j \partial x_i} \equiv \frac{\partial^2 V}{\partial x_i \partial x_j}$$

pour tout couple (i, j) de $\{1, \dots, n\} \times \{1, \dots, n\}$. De telles 1-formes sont dites *exactes*; si la 1-forme dérive sur U du potentiel V , on convient de la noter dV .

L'intégrale curviligne de la 1-forme exacte dV sur un chemin paramétré de classe C^1 par morceaux $\gamma : [a, b] \rightarrow U$ se calcule par la formule

$$\int_{\gamma} dV = V(\gamma(b)) - V(\gamma(a))$$

et est donc nulle si γ est un lacet; plus généralement, elle ne dépend que de la différence de potentiel aux extrémités $\gamma(b)$ et $\gamma(a)$ du support du chemin et non du chemin lui-même.

On vient de voir que toute 1-forme exacte dans un ouvert U de \mathbb{R}^n est automatiquement fermée. La réciproque est fautive; dans $\mathbb{R}^2 \setminus \{(0, 0)\}$, si la forme $\omega := \frac{xdy - ydx}{x^2 + y^2}$ (qui est fermée) était exacte (et s'écrivait dV), on aurait, si $\gamma : t \in [0, 2\pi] \rightarrow (\cos t, \sin t)$,

$$\int_{\gamma} \omega = \int_0^{2\pi} V'(\gamma(t))\gamma'(t) dt = V(e^{2i\pi}) - V(1) = 0$$

(car γ est un lacet) tandis que, d'après Green-Riemann,

$$\int_{\gamma} \omega = \int_{\gamma} (xdy - ydx) = 2 \iint_{D(0,1)} dx dy = 2\pi \neq 0 !$$

Par contre, la réciproque est vraie (c'est un cas particulier du *lemme de Poincaré*) dans un ouvert U tel que tout lacet fermé de U puisse de manière continue (et en restant dans U !) se déformer en un point. Un tel ouvert est dit *simplement connexe*; c'est le cas par exemple des ouverts *étoilés* par rapport à un de leurs points, des ouverts "sans trous" de \mathbb{R}^2 , des ouverts de \mathbb{R}^3 intérieurs de domaines quarrables tels deux de la section 1.7.3. Le potentiel V dont dérive la forme fermée ω s'obtient en posant

$$V(x) = \int_{\gamma_{x_0, x}} \omega,$$

où x_0 est un point de U et $\gamma_{x_0, x}$ un chemin paramétré C^1 par morceaux de support dans U ayant pour origine x_0 et pour extrémité x (la clause de simple connexité assure que cette intégrale curviligne ne dépend pas du chemin $\gamma_{x_0, x}$ choisi pourvu que la forme ω soit fermée).

Conclusion : dans un ouvert simplement connexe de \mathbb{R}^n , toute 1-forme de classe C^1 fermée dérive automatiquement d'un potentiel. Attention cependant à la clause de *simple connexité* qui par exemple dans le cas du plan interdit des trous dans l'ouvert !

1.14 Nappes paramétrées

On appelle *nappe paramétrée de classe C^1 régulière* toute application Φ classe C^1 d'un sous-ensemble élémentaire fermé A de \mathbb{R}^2 , à valeurs dans \mathbb{R}^3 , telle que la fonction vectorielle

$$(u, v) \rightarrow \frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v}$$

ne s'annule pas sur A . Le *support* de la nappe paramétrée $\Phi : A \rightarrow \mathbb{R}^3$ est par définition le sous-ensemble $\Phi(A)$. Le support d'une nappe paramétrée de classe C^1 régulière est aussi parfois appelé plus concrètement *portion de surface dans l'espace \mathbb{R}^3* . Dans la suite, on dira pour simplifier *nappe paramétrée* pour "nappe paramétrée de classe C^1 régulière".

Une *paramétrisation* de la nappe paramétrée $\Phi : A \rightarrow \mathbb{R}^3$ est par définition une application de classe C^1 $\tilde{\Phi}$ d'un sous-ensemble élémentaire $\tilde{A} \subset \mathbb{R}^2$, à valeurs dans \mathbb{R}^3 , telle qu'il existe une application bijective Ψ de classe C^1 entre $A \subset \mathbb{R}^2$ et $\tilde{A} \subset \mathbb{R}^2$ dont le jacobien reste strictement positif sur A , avec $\tilde{\Phi} \circ \Psi = \Phi$.

Deux nappes paramétrées $\Phi_1 : A_1 \rightarrow \mathbb{R}^3$ et $\Phi_2 : A_2 \rightarrow \mathbb{R}^3$ définissent la même *nappe géométrique* s'il existe une bijection Ψ de classe C^1 entre A_1 et A_2 telle que $\Phi_2 \circ \Psi = \Phi_1$. Si $\Phi_1 : A_1 \rightarrow \mathbb{R}^3$ et $\Phi_2 : A_2 \rightarrow \mathbb{R}^3$ sont deux paramétrisations de la même nappe paramétrée, on dit que ces deux nappes paramétrées définissent la même *nappe géométrique orientée*. La différence entre nappe paramétrée et nappe géométrique est que dans le cas de la nappe paramétrée, on a imposé une *orientation de la nappe*.

On soulignera l'analogie entre les notions de nappes paramétrées et de nappes géométriques (ce classe C^1 régulières) avec les notions de chemin paramétrée et d'arc géométrique (de classe C^1) vues dans la section 1.10.

Un exemple important de nappe paramétrée est celui où l'application $\Phi : A \rightarrow \mathbb{R}^3$ se présente comme un *graphe*, c'est-à-dire est de la forme

$$\Phi : (x, y) \rightarrow (x, y, \varphi(x, y)),$$

exemple que nous avons visualisé sur la figure 1.24. Si (x_0, y_0) est un point de A , la surface $\Phi(A)$ colle de manière infinitésimale au voisinage de $M_0 = (x_0, y_0, \varphi(x_0, y_0))$ au plan d'équation

$$-\frac{\partial \varphi}{\partial x}(x_0, y_0)(x - x_0) - \frac{\partial \varphi}{\partial y}(x_0, y_0)(y - y_0) + (z - z_0) = 0,$$

dit *plan tangent* à la portion de surface $\Phi(A)$ au point M_0 ; le vecteur non nul

$$\frac{\partial \Phi}{\partial x}(x_0, y_0) \wedge \frac{\partial \Phi}{\partial y}(x_0, y_0) = \begin{pmatrix} 1 \\ 0 \\ -\frac{\partial \varphi}{\partial x}(x_0, y_0) \end{pmatrix} \wedge \begin{pmatrix} 0 \\ 1 \\ -\frac{\partial \varphi}{\partial y}(x_0, y_0) \end{pmatrix} = \begin{pmatrix} -\frac{\partial \varphi}{\partial x}(x_0, y_0) \\ -\frac{\partial \varphi}{\partial y}(x_0, y_0) \\ 1 \end{pmatrix}$$

représente un *vecteur normal* à la surface au point M_0 , vecteur qui pointe ici vers les $z > 0$. Si \tilde{A} est l'image de A par l'application $(x, y) \rightarrow (y, x)$, l'application

$$\tilde{\Phi} : (x, y) \in \tilde{A} \rightarrow (y, x, \varphi(y, x)) = (y, x, \tilde{\varphi}(x, y))$$

définit la même nappe géométrique de $\Phi : A \rightarrow \mathbb{R}^3$, mais, du point de vue paramétrisation, la nappe géométrique orientée ayant une orientation opposée à celle induite par $\Phi : A \rightarrow \mathbb{R}^3$; le vecteur normal

$$\frac{\partial \tilde{\Phi}}{\partial x} \wedge \frac{\partial \tilde{\Phi}}{\partial y} = \begin{pmatrix} -\frac{\partial \tilde{\varphi}}{\partial x} \\ \frac{\partial \tilde{\varphi}}{\partial x} \\ -\frac{\partial \tilde{\varphi}}{\partial y} \\ -1 \end{pmatrix}$$

pointe cette fois dans la direction des $z < 0$!

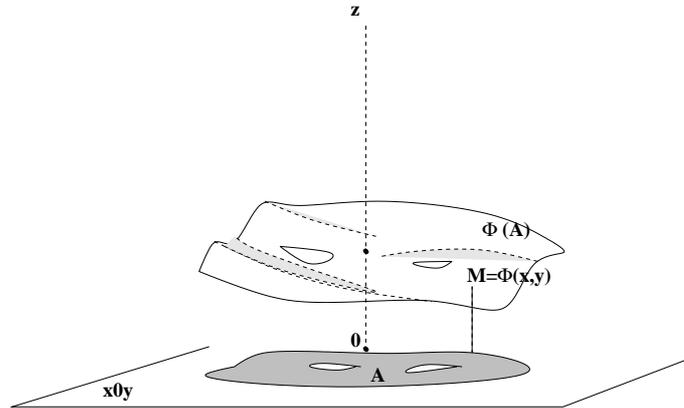


FIG. 1.24 – Un exemple de nappe paramétrée (cas des graphes)

Plus généralement, si $\Phi : (u, v) \in A \rightarrow \Phi(u, v) = (\varphi_1(u, v), \varphi_2(u, v), \varphi_3(u, v))$ est une nappe paramétrée, le vecteur

$$\frac{\partial \Phi}{\partial u}(u_0, v_0) \wedge \frac{\partial \Phi}{\partial v}(u_0, v_0) = \begin{pmatrix} \frac{\partial \varphi_2}{\partial u} \frac{\partial \varphi_3}{\partial v} - \frac{\partial \varphi_3}{\partial u} \frac{\partial \varphi_2}{\partial v} \\ \frac{\partial \varphi_3}{\partial u} \frac{\partial \varphi_1}{\partial v} - \frac{\partial \varphi_1}{\partial u} \frac{\partial \varphi_3}{\partial v} \\ \frac{\partial \varphi_1}{\partial u} \frac{\partial \varphi_2}{\partial v} - \frac{\partial \varphi_2}{\partial u} \frac{\partial \varphi_1}{\partial v} \end{pmatrix} (u_0, v_0)$$

définit un vecteur normal à la portion de surface $\Phi(A)$ au point $M_0 = \Phi(u_0, v_0)$ (c'est-à-dire un vecteur normal au plan tangent en M_0 à la portion de surface, plan représentant le plan "collant" à la portion de surface au niveau infinitésimal au voisinage de M_0).

Comme la quantité

$$\left\| \frac{\partial \Phi}{\partial u}(u, v) \wedge \frac{\partial \Phi}{\partial v}(u, v) \right\| dudv$$

représente la surface de l'image du pavé $[u, u + du] \times [v, v + dv]$ par l'application linéaire

$$\begin{pmatrix} h \\ k \end{pmatrix} \rightarrow d\Phi(u, v) \cdot \begin{pmatrix} h \\ k \end{pmatrix} = \begin{pmatrix} \frac{\partial \varphi_1}{\partial u} & \frac{\partial \varphi_1}{\partial v} \\ \frac{\partial \varphi_2}{\partial u} & \frac{\partial \varphi_2}{\partial v} \\ \frac{\partial \varphi_3}{\partial u} & \frac{\partial \varphi_3}{\partial v} \end{pmatrix} \cdot \begin{pmatrix} h \\ k \end{pmatrix}$$

(voir la section 1.6.3), il est naturel de définir l'aire de la portion de surface $\Phi(A)$ par

$$\text{aire}(\Phi(A)) = \iint_A \left\| \frac{\partial \Phi}{\partial u}(u, v) \wedge \frac{\partial \Phi}{\partial v}(u, v) \right\| dudv.$$

La formule de changement de variable dans les intégrales doubles montre que ce nombre ne dépend que de la nappe géométrique correspondant à la nappe paramétrée

Φ (en particulier, ce nombre ne dépend pas de l'orientation de Φ). Du point de vue pratique, le fait que l'on doive extraire une racine carrée rend en général difficile, voire impossible, le calcul de l'aire d'un élément de surface par une formule close (c'est la même difficulté que celle rencontrée pour calculer la longueur d'un arc géométrique).

Exemple 1.5 La demi-sphère de rayon 1 paramétrée par $x = \cos \theta \sin \varphi$, $y = r \sin \theta \sin \varphi$, $z = \cos \varphi$, $(\theta, \varphi) \in [0, 2\pi] \times [0, \pi/2]$ (θ pour la longitude, φ pour la co-latitude) a pour aire

$$\iint_{[0, 2\pi] \times [0, \pi/2]} \sqrt{\sin^4 \varphi \cos^2 \theta + \sin^4 \varphi \sin^2 \theta + \sin^2 \varphi \cos^2 \varphi} d\theta d\varphi = 2\pi,$$

et l'on retrouve l'aire de la sphère égale à 4π .

1.15 La formule de Stokes pour les nappes paramétrées

Soit $\Phi : A \rightarrow \mathbb{R}^3$ une nappe paramétrée et $\vec{F} = (P, Q, R)$ un champ de vecteurs dans l'espace défini et de classe C^2 au voisinage de la portion de surface correspondant au support de Φ . Comme le bord (orienté) de A correspond à une collection de chemins paramétrés $\gamma_1, \dots, \gamma_N$ (l'orientation étant toujours conforme à la règle du bonhomme d'Ampère qui exige que le domaine A soit à notre gauche lorsque l'on suit l'un des supports des chemins qui marquent son bord), les chemins $\Phi \circ \gamma_j$, $j = 1, \dots, N$, qui sont cette fois des chemins de \mathbb{R}^3 , sont tels que leurs supports constituent le bord de la portion de surface $\Phi(A)$. Le fait que les γ_j soient des chemins paramétrés, donc orientés, induit que le bord de $\Phi(A)$ hérite d'une orientation.

La *formule de Stokes* relie le calcul de l'intégrale curviligne

$$\sum_{j=1}^N \int_{\Phi \circ \gamma_j} (Pdx + Qdy + Rdz)$$

à une certaine intégrale de surface sur la nappe géométrique correspondant à Φ . Cette formule s'énonce sous la forme

$$\begin{aligned} \sum_{j=1}^N \int_{\Phi \circ \gamma_j} (Pdx + Qdy + Rdz) &= \iint_A \left(\text{rot}(\vec{F}) \bullet \left(\frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v} \right) \right) dudv \\ &= \iint_A \left(\text{rot}(\vec{F}) \bullet \left(\frac{\frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v}}{\left\| \frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v} \right\|} \right) \right) \left\| \frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v} \right\| dudv \quad (\dagger) \end{aligned}$$

le rotationnel $\text{rot}(\vec{F})$ du champ \vec{F} étant par définition le champ de vecteurs défini par

$$\text{rot}(\vec{F}) := \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$$

et \bullet désignant le produit scalaire dans \mathbb{R}^3 .

On a noté

$$n(u, v) = \frac{\frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v}}{\left\| \frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v} \right\|}$$

le vecteur normal unitaire à la portion de surface $\Phi(A)$, orienté suivant le choix de la nappe paramétrée Φ (ceci est important!) et on a représenté sur la figure 1.25 le vecteur normal $n(u, v)$ et le rotationnel de \vec{F} au point courant $\Phi(u, v)$ de la portion de surface.

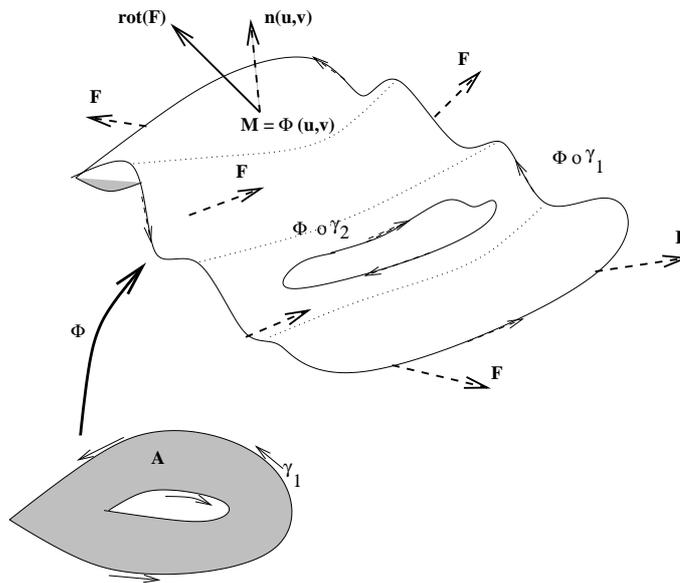


FIG. 1.25 – La formule de Stokes pour une nappe paramétrée

Si l'on choisit une nappe paramétrée $\tilde{\Phi} : \tilde{A} \rightarrow \mathbb{R}^3$ correspondant à la même nappe géométrique, mais avec des orientations opposées, le membre de droite de la formule est changé en son opposé.

La formule (†) se déduit de la formule de Green-Riemann de la section 1.12; on remarque simplement que

$$\sum_{j=1}^N \int_{\Phi \circ \gamma_j} (Pdx + Qdy + Rdz) = \sum_{j=1}^N \int_{\gamma_j} \omega,$$

avec

$$\omega = \left(P \frac{\partial \Phi_1}{\partial u} + Q \frac{\partial \Phi_2}{\partial u} + R \frac{\partial \Phi_3}{\partial u} \right) du + \left(P \frac{\partial \Phi_1}{\partial v} + Q \frac{\partial \Phi_2}{\partial v} + R \frac{\partial \Phi_3}{\partial v} \right) dv,$$

puis on applique la formule de Green-Riemann dans le domaine élémentaire plan A .

Voici l'interprétation "physique" de cette importante formule de Stokes : le membre de gauche

$$\sum_{j=1}^N \int_{\Phi \circ \gamma_j} (Pdx + Qdy + Rdz)$$

représente la *circulation* (ou le *travail*) du champ de forces \vec{F} le long du bord (orienté) de la portion de surface $\Phi(A)$ (le bord étant parcouru de manière à avoir constamment la portion de surface $\Phi(A)$ à gauche). Le membre de droite

$$\iint_A \left(\text{rot}(\vec{F}) \cdot \left(\frac{\frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v}}{\left\| \frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v} \right\|} \right) \right) \left\| \frac{\partial \Phi}{\partial u} \wedge \frac{\partial \Phi}{\partial v} \right\| dudv$$

représente le *flux* du champ $\text{rot}(\vec{F})$ à travers la nappe géométrique orientée définie par la nappe paramétrée Φ . Ce flux ne dépend pas de la paramétrisation de la nappe ; par contre, si l'on prend une paramétrisation correspondant à la même nappe géométrique mais avec l'orientation opposée, le flux est changé en son opposé.

On remarque que la formule de Green-Riemann est une conséquence de cette formule lorsque $\Phi : (x, y) \rightarrow (x, y, 0)$ (la portion de surface A est une portion de surface plane dans ce cas).

En fait, de même que l'on peut définir la différentielle d'un potentiel (dV est par définition la 1 forme correspondant à l'action du champ de forces $\text{grad } V$), on peut définir la différentielle d'une 1-forme $\omega = Pdx + Qdy + Rdz$; nous nous contenterons de faire ceci formellement en posant

$$d\omega = dP \wedge dx + dQ \wedge dy + dR \wedge dz$$

et en tirant parti des relations sous-jacentes à l'opération de "produit extérieur" (notée \wedge), à savoir $dx \wedge dx = dy \wedge dy = dz \wedge dz = 0$, $dx \wedge dy = -dy \wedge dx$, $dy \wedge dz = -dz \wedge dy$, $dz \wedge dx = -dx \wedge dz$. On obtient ainsi

$$d\omega = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy \wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz \wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy ;$$

on remarque que

$$d\omega = \tilde{P}dy \wedge dz + \tilde{Q}dz \wedge dx + \tilde{R}dx \wedge dy ,$$

où $(\tilde{P}, \tilde{Q}, \tilde{R})$ est le rotationnel du champ de vecteurs (P, Q, R) dont l'action correspond à la 1-forme ω .

Si $\Phi : A \rightarrow \mathbb{R}^3$ est une nappe paramétrée et ω une 1-forme au voisinage de la portion de surface $\Phi(A)$ correspondant à l'action du champ de forces (P, Q, R) (donc $\omega = Pdx + Qdy + Rdz$), le calcul du flux du rotationnel de (P, Q, R) à travers la nappe géométrique orientée correspondant à Φ s'effectue formellement comme suit : on exprime la forme $d\omega$ en remplaçant dx par $d\Phi_1 = (\partial\Phi_1/\partial u) du + (\partial\Phi_1/\partial v) dv$, dy par $d\Phi_2$, dz par $d\Phi_3$, ce qui nous donne comme nouvelle expression une expression du type

$$\Psi(u, v) du \wedge dv$$

(en utilisant les règles de calcul impliquées par le symbole \wedge), puis l'on constate que le flux du rotationnel de (P, Q, R) à travers la nappe géométrique orientée définie par Φ vaut

$$\iint_A \Psi(u, v) du dv .$$

Si l'on convient de l'identification formelle

$$\iint_A \Psi(u, v) du \wedge dv := \iint_A \Psi(u, v) du dv ,$$

alors on constate que la formule de Stokes se lit

$$\sum_{j=1}^N \int_{\Phi \circ \gamma_j} \omega = \iint_{\Phi(A)} d\omega$$

avec la convention

$$\iint_{\Phi(A)} d\omega = \iint_A \Psi(u, v) du \wedge dv = \iint_A \Psi(u, v) du dv .$$

C'est donc bien le pendant en dimension deux du théorème fondamental de l'analyse de la section 1.4.

1.16 La formule de Green-Ostrogradski

Soit Δ le simplexe de \mathbb{R}^3 défini comme le tétraèdre

$$\Delta := \{(x, y, z) \in \mathbb{R}^3 ; x \geq 0, y \geq 0, z \geq 0, x + y + z \leq 1\},$$

avec ses quatre faces $\Delta \cap \{x = 0\}$ (le vecteur normal unitaire dirigé vers l'extérieur de Δ à cette face étant $(-1, 0, 0)$), $\Delta \cap \{y = 0\}$ (le vecteur normal unitaire dirigé vers l'extérieur de Δ à cette face étant $(0, -1, 0)$), $\Delta \cap \{z = 0\}$ (le vecteur normal unitaire dirigé vers l'extérieur de Δ à cette face étant $(0, 0, -1)$), $\Delta \cap \{x + y + z = 1\}$ (le vecteur normal unitaire dirigé vers l'extérieur de Δ à cette face étant $(1, 1, 1)/\sqrt{3}$) (voir la figure 1.26 (a)).

Si $\vec{F} = P\vec{i} + Q\vec{j} + R\vec{k}$ est un champ de vecteurs de classe C^1 au voisinage de Δ , on peut calculer la somme des flux de ce champ à travers les quatre faces mentionnées, l'orientation choisie pour la normale étant celle qui pointe vers l'extérieur de Δ . On appelle cette somme le *flux sortant* du champ \vec{F} au travers du simplexe Δ et on la note $\Psi_{\Delta}(\vec{F})$.

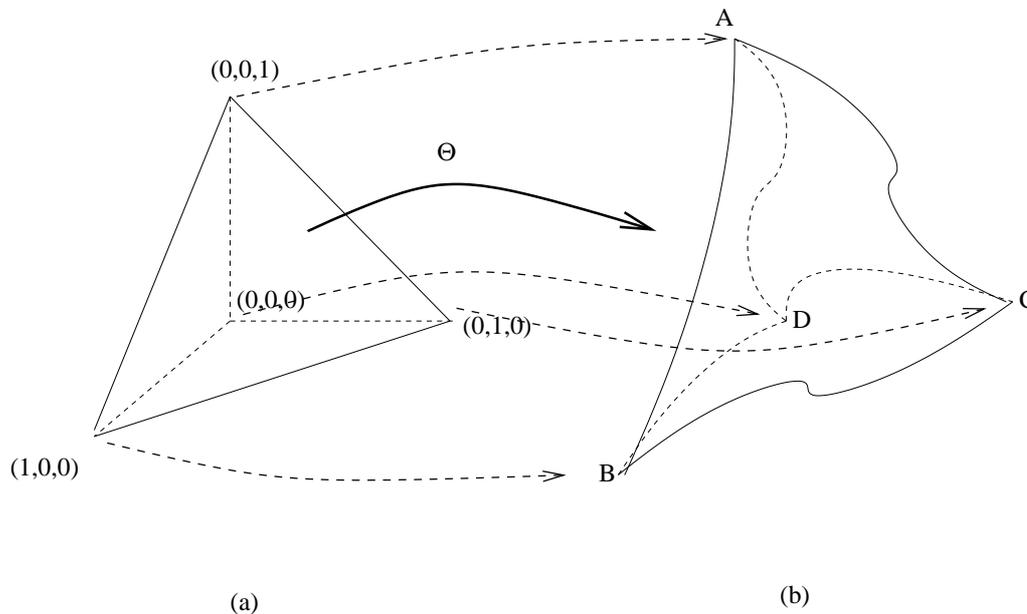


FIG. 1.26 – Simplexe et simplexe déformé dans \mathbb{R}^3

On remarque que l'on a la formule :

$$\Psi_{\Delta}(\vec{F}) = \iiint_{\Delta} \operatorname{div} \vec{F}(x, y, z) \, dx dy dz,$$

où

$$\operatorname{div} \vec{F} := \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$$

dénote la *divergence* du champ \vec{F} .

Ceci se voit par un calcul simple, que l'on peut se contenter de faire si $\vec{F} = R\vec{k}$ (par raison de symétrie), auquel cas pour les calculs de flux, il n'y a que deux faces

(correspondant à des nappes paramétrées convenablement orientées) à considérer, celles qui correspondent aux nappes paramétrées :

$$\begin{aligned}(x, y) \in \Delta_2 &\rightarrow (y, x, 0) \\ (x, y) \in \Delta_2 &\rightarrow (x, y, 1 - x - y),\end{aligned}$$

où $\Delta_2 := \{(x, y) \in \mathbb{R}^2; x \geq 0, y \geq 0, x + y \leq 1\}$. Le flux sortant se réduit (par Fubini) à l'expression

$$\int_0^1 \left(\int_0^{1-x} (R(x, y, 1 - x - y) - R(x, y, 0)) dy \right) dx,$$

expression que l'on peut réécrire en utilisant le théorème fondamental de l'analyse sous la forme

$$\int_0^1 \left(\int_0^{1-x} \left(\int_0^{1-x-y} \frac{\partial R}{\partial z}(x, y, z) dz \right) dy \right) dx;$$

par le théorème de Fubini encore, cette intégrale vaut bien

$$\iiint_{\Delta} \frac{\partial R}{\partial z}(x, y, z) dx dy dz,$$

ce qui correspond bien à l'intégrale sur Δ de la divergence de $\vec{F} = R\vec{k}$.

Si maintenant Θ est une application de classe C^1 de Δ dans \mathbb{R}^3 dont le jacobien reste strictement positif et si $\vec{F} = \tilde{P}\vec{i} + \tilde{Q}\vec{j} + \tilde{R}\vec{k}$ est un champ au voisinage du tétraèdre déformé $\tilde{\Delta} = \Theta(\Delta)$ (voir la figure 1.26 (b)), la formule de changement de variable dans les intégrales triples de la section 1.9 montre que

$$\iiint_{\tilde{\Delta}} \operatorname{div} \vec{F}(u, v, w) dudvdw = \iiint_{\Theta(\Delta)} \operatorname{div} \vec{F}(x, y, z) dx dy dz$$

où les composantes du champ \vec{F} sont définies par :

$$\begin{aligned}P &= (\tilde{P} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_2}{\partial v} & \frac{\partial \theta_2}{\partial w} \\ \frac{\partial \theta_3}{\partial v} & \frac{\partial \theta_3}{\partial w} \end{vmatrix} + (\tilde{Q} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_3}{\partial v} & \frac{\partial \theta_3}{\partial w} \\ \frac{\partial \theta_1}{\partial v} & \frac{\partial \theta_1}{\partial w} \end{vmatrix} + (\tilde{R} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_1}{\partial v} & \frac{\partial \theta_1}{\partial w} \\ \frac{\partial \theta_2}{\partial v} & \frac{\partial \theta_2}{\partial w} \end{vmatrix} \\ Q &= (\tilde{P} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_2}{\partial w} & \frac{\partial \theta_2}{\partial u} \\ \frac{\partial \theta_3}{\partial w} & \frac{\partial \theta_3}{\partial u} \end{vmatrix} + (\tilde{Q} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_3}{\partial w} & \frac{\partial \theta_3}{\partial u} \\ \frac{\partial \theta_1}{\partial w} & \frac{\partial \theta_1}{\partial u} \end{vmatrix} + (\tilde{R} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_1}{\partial w} & \frac{\partial \theta_1}{\partial u} \\ \frac{\partial \theta_2}{\partial w} & \frac{\partial \theta_2}{\partial u} \end{vmatrix} \\ R &= (\tilde{P} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_2}{\partial u} & \frac{\partial \theta_2}{\partial v} \\ \frac{\partial \theta_3}{\partial u} & \frac{\partial \theta_3}{\partial v} \end{vmatrix} + (\tilde{Q} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_3}{\partial u} & \frac{\partial \theta_3}{\partial v} \\ \frac{\partial \theta_1}{\partial u} & \frac{\partial \theta_1}{\partial v} \end{vmatrix} + (\tilde{R} \circ \Theta) \begin{vmatrix} \frac{\partial \theta_1}{\partial u} & \frac{\partial \theta_1}{\partial v} \\ \frac{\partial \theta_2}{\partial u} & \frac{\partial \theta_2}{\partial v} \end{vmatrix};\end{aligned}$$

le calcul est un peu pénible, mais on y arrive !

On peut aussi calculer la somme des flux de \vec{F} à travers les quatre nappes paramétrées correspondant à l'image par Θ des quatre faces du simplexe Δ , soit les quatre nappes :

$$\begin{aligned}(s, t) \in \Delta_2 &\rightarrow \Theta(0, t, s) \\ (s, t) \in \Delta_2 &\rightarrow \Theta(t, 0, s) \\ (s, t) \in \Delta_2 &\rightarrow \Theta(t, s, 0) \\ (s, t) \in \Delta_2 &\rightarrow \Theta(s, t, 1 - t - s).\end{aligned}$$

Un calcul tout aussi fastidieux que le précédent, mais basé cette fois sur la formule de changement de variable dans les intégrales doubles, montre que la somme de ces quatre flux, que l'on note aussi $\Psi_{\tilde{\Delta}}(\vec{F})$ vaut

$$\Psi_{\tilde{\Delta}}(\vec{F}) = \Psi_{\Delta}(\vec{F}).$$

On a en combinant les deux formules :

$$\Psi_{\tilde{\Delta}}(\vec{F}) = \iiint_{\Theta(\tilde{\Delta})} \operatorname{div} \vec{F}(x, y, z) \, dx dy dz . \quad (\dagger)$$

Si maintenant A est un domaine comme dans la section 1.7.3, c'est-à-dire un ouvert borné avec éventuellement des cavités dont la frontière se présente comme une union de supports de nappes paramétrées de classe C^1 , on peut, exactement comme on a dans la section 1.11 triangulé un domaine de Jordan généralisé dans \mathbb{R}^2 pour prouver la formule de Green-Riemann (voir la figure 1.22), décomposer le fermé \bar{A} comme une union de tétraèdres déformés $\tilde{\Delta}_j$, d'intérieurs disjoints, avec la règle suivante : si $\tilde{\Delta}_i$ et $\tilde{\Delta}_j$ se coupent, alors l'intersection de ces deux polytopes est exactement soit un sommet commun (et un seul), soit une arête commune (et une seule), soit une face commune (et une seule).

Si \vec{F} est un champ de vecteurs de classe C^1 au voisinage du domaine fermé \bar{A} , on voit, en additionnant les formules (\dagger) pour les divers polytopes de la triangulation que le flux sortant du champ \vec{F} à travers le domaine A (ce flux est donc calculé en prenant la normale pointant vers l'extérieur du domaine), flux que l'on note $\Psi_A(\vec{F})$, se calcule par la formule dite *de Green-Ostrogradski* :

$$\Psi_A(\vec{F}) = \iiint_A \operatorname{div} \vec{F}(x, y, z) \, dx dy dz .$$

En termes de formes différentielles, on voit que, si ω est la 2-forme

$$\omega = P dy \wedge dz + Q dz \wedge dx + R dx \wedge dy ,$$

alors

$$d\omega = dP \wedge dx \wedge dy + dQ \wedge dy \wedge dz + dR \wedge dx \wedge dy = \operatorname{div} \vec{F} \, dx \wedge dy \wedge dz$$

(en respectant les règles formelles du calcul extérieur établies dans la section 1.15). Notre formule de Green-Ostrogradski se lit encore, si l'on convient que le bord ∂A de A est orienté de manière à ce que la normale pointe vers l'extérieur,

$$\iint_{\partial A} \omega = \iiint_A d\omega ,$$

si l'on convient que

$$\iiint_A f(x, y, z) dx \wedge dy \wedge dz = \iiint_A f(x, y, z) dx dy dz ;$$

sous cette forme, la formule de Green-Ostrogradski apparaît comme la transcription dans \mathbb{R}^3 du théorème fondamental de l'analyse de la section 1.4.

Chapitre 2

Initiation à la théorie des probabilités

2.1 Epreuve et ensemble d'évènements

Le lancer d'un dé à six faces constitue un exemple de ce que l'on appelle une *épreuve*; si le dé n'est pas pipé, tout non-mathématicien prévoira qu'il a une chance sur six de réaliser *six* lors d'une telle épreuve. Il étayera son résultat en vous faisant constater que, s'il itère N fois cette épreuve (avec N très grand), le quotient $N(6)/N$ (où $N(6)$ désigne le nombre de six obtenus lors de la série de N coups) tend asymptotiquement vers $1/6$ lorsque N tend vers l'infini. Son argumentation relève, on le verra, du point de vue qualifié de *statistique*.

Formulons mathématiquement le problème en considérant l'ensemble Ω de tous les résultats possibles de l'épreuve, ici

$$\Omega = \{1, \dots, 6\}.$$

Le résultat de l'épreuve est *aléatoire*; l'épreuve consiste à choisir un élément dans l'ensemble Ω , espace que l'on qualifie ici d'*espace d'évènements* (on devrait en fait plutôt dire "d'évènements élémentaires"). Se donner la *loi* probabiliste à laquelle obéit cette épreuve aléatoire, à savoir ici le lancer du dé, consiste ici à se donner une collection de 6 nombres réels positifs $p_j, j = 1, \dots, 6$ de somme 1. La quantité p_j représente ce que notre non-mathématicien interprétait comme *le nombre de chances de réaliser le résultat j* en lançant le dé.

Dans le cas de dés non pipés, il est naturel de travailler avec le modèle mathématique :

$$p_j = \frac{1}{6}, j = 1, \dots, 6.$$

Ce modèle est un cas particulier du cas où Ω est un ensemble fini et où chaque singleton (ou encore évènement élémentaire) $\{a\}$ de Ω est "chargé" avec la masse

$$P(\{a\}) = \frac{1}{\text{card } \Omega},$$

auquel cas, une partie A de Ω (que l'on appelle un évènement) est "chargée" avec la masse

$$P(A) = \frac{\text{card } A}{\text{card } \Omega};$$

cette distribution de probabilité est dite *distribution uniforme* sur Ω .

Dans le cas de dés pipés, on doit envisager d'autres modèles, comme par exemple

$$p_1 = p_2 = p_3 = p_4 = \frac{1}{12}, \quad p_5 = p_6 = \frac{1}{3}.$$

L'ensemble des évènements n'est pas nécessairement fini, comme le montre l'exemple *géométrique* suivant : l'épreuve consiste à prendre au hasard deux points distincts sur un cercle Γ donné de rayon 1 et à tracer la corde qui les joint (comme sur la figure 2.1).

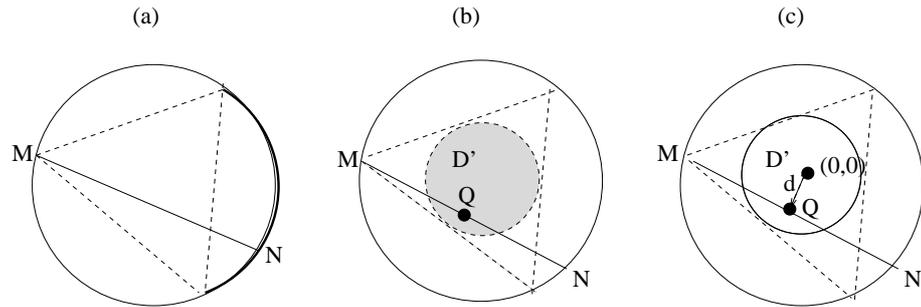


FIG. 2.1 – Le paradoxe de Bertrand

On peut poser la question naïve suivante : les choix de points étant équiprobables sur $\Gamma \times \Gamma$ (on va préciser ceci ultérieurement), quel "nombre de chance" ou encore "probabilité" a-t-on que la longueur de la corde joignant deux points M et N de Γ soit supérieure à celle du côté d'un triangle équilatéral inscrit dans Γ ? (voir la figure 2.1 (a)). Notre non-spécialiste pourra répondre en disant qu'il y a une chance sur trois que ce qu'il souhaite se réalise lorsqu'il choisit ses deux points au hasard (et indépendamment l'un de l'autre) : son raisonnement s'appuie sur le fait que, dès que l'un des points est fixé, le second doit être dans un arc de cercle de longueur $\pi/3$. Ce calcul intuitif de probabilité est fondé sur le fait que l'ensemble d'évènements Ω choisi soit ici le produit $\Omega = \Gamma \times \Gamma$ du cercle par lui-même, la probabilité d'un produit d'intervalles cuvilignes $I \times J$ étant

$$\frac{\text{longueur}(I)}{\pi} \times \frac{\text{longueur}(J)}{\pi}.$$

Ceci est cependant ambigu, comme le faisait remarquer le mathématicien Joseph Bertrand en 1899. On peut en effet aussi repérer la corde par son milieu Q et considérer comme ensemble d'évènements Ω l'ensemble des positions possibles de ce milieu après tracé de la corde, soit le disque D dont Γ est le bord ; cette fois, l'ensemble des évènements n'est plus l'ensemble des paires de points de Γ , mais l'ensemble des points du disque fermé $D(0, 1)$ (ces points repèrent le milieu de la corde, voir la figure 2.1 (b)) ; avec ce choix, le fait que la corde ait une longueur supérieure à celle du triangle équilatéral inscrit se rephrase en *le milieu de la corde est dans le disque D' de rayon $1/2$ et de centre l'origine* ; notre non-spécialiste pourrait tout autant affirmer qu'il y a une chance sur 4 (puisque $1/4$ représente le quotient de la surface de D' par la surface de $D(0, 1)$) que son souhait se réalise. On pourrait aussi en repérant la corde par sa distance $d \in [0, 1]$ à l'origine (voir la figure 2.1 (c))

et (en prenant comme ensemble des évènements $[0, 1]$) dire que l'évènement que l'on cherche est celui qui correspond au fait que cette distance soit inférieure ou égale à $1/2$, soit l'intervalle $[0, 1/2]$, qui est de longueur $1/2$; comme la longueur de $[0, 1]$ est normalisée égale à 1, la probabilité de notre évènement suivant cette troisième interprétation vaudrait $1/2$!

Cet exemple montre bien que la mise sous forme mathématique d'un problème de nature probabiliste exige au préalable la définition de trois choses :

- un ensemble Ω d'évènements élémentaires (les points de Ω peuvent par exemple permettre de repérer les résultats d'une épreuve); c'est la non-définition claire de cet ensemble d'évènements qui explique le paradoxe de Bertrand;
- une famille de sous-ensembles de Ω candidats à pouvoir être *mesurés* (la contrainte étant que le mathématicien n'a que l'accès au dénombrable pour faire ses calculs de mesure);
- une règle enfin pour mesurer, avec la convention que la mesure de Ω est 1; si Ω représente l'ensemble des résultats d'une certaine épreuve, alors la "mesure" de A sera interprétée par notre non-spécialiste comme le pourcentage de chance qu'a le résultat de l'épreuve de tomber dans A .

Nous allons dans la section suivante définir proprement les notions mathématiques consolidant ces idées intuitives.

2.2 Tribus et probabilités

Une *tribu* \mathcal{T} sur un ensemble abstrait Ω (appelé à être un ensemble d'évènements, par exemple les résultats d'une épreuve) est une sous-famille \mathcal{T} de l'ensemble des parties de Ω telle que :

- $\Omega \in \mathcal{T}$;
- si $A \in \mathcal{T}$, $\Omega \setminus A$ aussi (donc l'ensemble vide est en particulier dans \mathcal{T});
- si $A_1, A_2 \in \mathcal{T}$, alors $A_1 \cap A_2 \in \mathcal{T}$;
- si $(A_n)_{n \in \mathbb{N}}$ est une famille dénombrable d'éléments de \mathcal{T} , l'union de tous les A_n est encore dans \mathcal{T} .

Si Ω est un ensemble équipé d'une tribu, on dit que le couple (Ω, \mathcal{T}) est un *espace probabilisable*.

Exemple 2.1

- un exemple basique est celui où Ω est un ensemble fini ou dénombrable et \mathcal{T} la famille de toutes des parties de Ω (ensemble vide inclus);
- si $\Omega = \mathbb{R}^n$, la collection de tous les pavés du type $]a_1, b_1] \times \cdots \times]a_n, b_n]$ n'est pas une tribu, mais on appelle *tribu borélienne* (du nom du mathématicien français Emile Borel, 1871-1956) la plus petite tribu contenant tous ces pavés (c'est aussi la plus petite tribu contenant tous les ouverts de \mathbb{R}^n) et *tribu de Lebesgue* (du nom cette fois d'Henri Lebesgue, 1875-1941) la plus petite tribu contenant la tribu borélienne et tous les sous-ensembles de \mathbb{R}^n de mesure extérieure nulle; la tribu de Lebesgue est en fait constituée de tous les sous-ensembles quarrables de \mathbb{R}^n . (on l'admettra); l'axiome du choix en logique permet la construction de sous-ensembles de \mathbb{R}^n non quarrables (mais ce n'est pas évident!), ce qui fait que ni la tribu borélienne, ni même la tribu de Lebesgue, ne forment la tribu de toutes les parties.

Signalons que la nécessité de travailler avec des ensembles d'évènements non dénombrables est très souvent chose courante en probabilité, même si le problème

semble de nature discrète : par exemple, l'ensemble d'évènements attaché à une partie de pile ou face est $\{0, 1\}^{\mathbb{N}}$, ensemble des suites infinies de d'éléments de $\{0, 1\}$, ensemble dont le cardinal a la puissance du continu (il y a correspondance entre les points de $[0, 1]$ et leur écriture dans le système en base deux).

Une *mesure de probabilité* sur un espace probabilisable (Ω, \mathcal{T}) est une application P de \mathcal{T} dans $[0, \infty]$ telle que l'on ait $P(\Omega) = 1$ et que, si les ensembles A_n , $n \in \mathbb{N}$, sont des éléments de \mathcal{T} disjoints, on ait

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n).$$

Ainsi, si A et B sont deux éléments de \mathcal{T} , on a la formule

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

formule que l'on peut généraliser au cas plus général du calcul de la probabilité d'une des union d'éléments de \mathcal{T} non nécessairement disjoints.

Les calculs de probabilités sur un ensemble fini (la distribution de probabilité étant la distribution uniforme) font intervenir des calculs de dénombrement. Si k et n sont deux entiers strictement positifs tels que $k \leq n$, le nombre d'applications injectives d'un ensemble à k éléments dans un ensemble à n éléments est aussi égal au nombre de k -uplets (n_1, \dots, n_k) , avec $1 \leq n_j \leq n$ et les n_j distincts (attention ! l'ordre des n_j est ici important) ; ce nombre, qui vaut

$$A_n^k = \frac{n!}{(n-k)!}$$

est aussi appelé *nombre d'arrangements* de l'ensemble à k éléments dans l'ensemble à n éléments. Le nombre de manières de choisir k éléments parmi N se déduit du précédent par la formule

$$C_n^k = \binom{n}{k} = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}$$

puisque le nombre de permutations d'un ensemble à k éléments est égal au cardinal du groupe symétrique \mathcal{S}_k , soit à k . Le nombre C_n^k (noté aussi $\binom{n}{k}$) est dit *nombre de combinaisons* de k éléments parmi n ; si $k = 0$, on convient que $A_n^0 = C_n^0 = 1$. La suite des nombres C_n^k est régie par les identités du *triangle de Pascal* :

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \quad k, n \in \mathbb{N},$$

avec les conventions $\binom{m}{p} = 0$ si $p > m$ ou si $p < 0$. On a aussi la classique *formule du binôme* :

$$2^n = \sum_{k=0}^n \binom{n}{k}, \quad n \in \mathbb{N}.$$

Exemple 2.2

- Sur $\Omega = 1, \dots, n$ (ou plus généralement si Ω est un ensemble à n éléments), la probabilité affectant à chaque singleton le nombre $1/n$ est appelée *distribution de probabilité uniforme sur Ω* ; la probabilité de A vaut dans ce cas $\text{card } A / \text{card } \Omega$.
- Autre exemple plus sophistiqué : sur $\{0, \dots, n\}$, on définit une probabilité P_τ associée à un paramètre $\tau \in [0, 1]$ en posant

$$P_\tau(k) := \binom{n}{k} \tau^k (1 - \tau)^{n-k}.$$

c'est la *distribution de probabilité binomiale de paramètres (n, τ)* , celle que l'on retrouve dans les problèmes de tirage avec remise : considérons une urne avec N boules, dont R rouges et B blanches ; on décide de tirer k boules de manière indépendante (ceci doit être précisé par la suite) et avec remise. La probabilité que l'on tire k boules rouges (si on prend comme ensemble d'évènements $\Omega = \{1, \dots, N\}^k$ et comme distribution de probabilité la loi uniforme, ce qui, on le verra ultérieurement rend compte de l'indépendance) est alors

$$P(R_k) = \frac{\binom{n}{k} R^k (N - R)^{N-k}}{N^n}$$

ce qui correspond à une loi de probabilité binomiale avec $n = N$ et $\tau = R/N$. Notons que c'est aussi le résultat que l'on trouve (avec $\tau = 1/2$) lorsque, dans une suite de N épreuves indépendantes de pile ou face avec une pièce honnête, on évalue la probabilité d'obtenir k fois pile lors de ces n épreuves. Ici encore, pour corroborer les hypothèses de l'expérience (l'*indépendance* des jets successifs), on travaille avec $\Omega = \{pile, face\}^N$ et la distribution uniforme.

- Le dernier exemple que nous proposons sur un ensemble fini correspond, lui, aux résultats que l'on obtient lorsque l'on envisage des tirages hors d'une urne, mais cette fois sans remise. Considérons 3 nombres entiers R, N, n avec $R < N$ et $n \leq N$; on définit une probabilité sur $\{0, \dots, n\}$ en posant, pour tout k dans $\{0, \dots, n\}$ tel que

$$\max(0, n - (N - R)) \leq k \leq \min(n, R),$$

$$p_k = \frac{\binom{R}{k} \binom{N - R}{n - k}}{\binom{N}{n}}$$

et 0 sinon. Cette loi de probabilité est la loi *hypergéométrique* de paramètres $n, \tau_1 = R/N, \tau_2 = (N - R)/N$; cette définition peut être élargie au cas où le nombre des paramètres τ_j (avec $\sum \tau_j = 1$) dépasse 2 ; on a alors la loi *polyhypergéométrique*. On retrouve (on le verra en exercice dans la section 2.3) les p_k lorsque l'on effectue n tirages successifs (et indépendants) hors d'une urne contenant N boules, dont R rouges et $N - R$ blanches ; en considérant comme espace probabilisé l'espace de toutes les suites à n termes de 0, 1 possibles (1 si la boule tirée est rouge, 0 si elle est blanche), équipé de la distribution de probabilité uniforme, le calcul de la probabilité de l'évènement *il y a exactement k boules rouges tirées* donne le résultat p_k .

- On peut également donner des exemples dans le cas où $\Omega = \mathbf{N}$; se donner une distribution de probabilité (la tribu étant celle de toutes les parties) équivaut à se donner une suite (p_n) de réels positifs telle que la série $\sum_n p_n$ converge et ait pour somme 1 : par exemple la suite

$$p_n = (1 - p)p^n, \quad 0 \leq p < 1$$

(loi *géométrique* ou de *Pascal* de paramètre p) ou encore

$$p_n = \frac{\lambda^n}{n!} \exp(-\lambda), \quad \lambda \in \mathbf{R}^+$$

(loi de *Poisson* de paramètre λ).

Pour construire des probabilités sur \mathbf{R} ou \mathbf{R}^n équipés de la tribu borélienne, on utilise un résultat d'extension permettant de prolonger à la tribu borélienne une

”mesure” sur la collection des pavés du type $]a_1, b_1] \times \cdots \times]a_n, b_n]$. Pourvu que la mesure obtenue soit telle que $P(\mathbb{R}^n) = 1$, on construit bien ainsi des distributions de probabilité. Le ”prototype” de telles distributions de probabilité sur \mathbb{R}^n repose sur la donnée d’une fonction intégrable $f : \mathbb{R}^n \rightarrow [0, +\infty]$ telle que

$$\int \cdots \int_{\mathbb{R}^n} f(x_1, \dots, x_n) dx_1 dx_2 \cdots dx_n = 1.$$

Une telle fonction est dite *densité de probabilité* sur \mathbb{R}^n . Si f est une telle fonction densité sur \mathbb{R}^n , la probabilité induite par f sur l’espace probabilisable \mathbb{R}^n équipé de la tribu borélienne telle que, pour tout borélien de \mathbb{R}^n , on ait

$$P(A) = \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n;$$

on dit que cette distribution de probabilité sur \mathbb{R}^n est une *distribution de densité de probabilité* f .

Un exemple majeur de telle distribution sur \mathbb{R} est la *distribution de Gauss*, correspondant à la fonction

$$t \rightarrow f(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$$

(dite *gaussienne réduite centrée*). Cette distribution joue un rôle crucial en physique car la gaussienne est la fonction réalisant le meilleur compromis dans l’incoutournable principe d’incertitude (impossible de localiser simultanément une particule et son spectre!); on la retrouvera aussi dans les théorèmes limite de la théorie des probabilités : la célèbre *expérience de Galton* consistant à laisser tomber un stock de billes au dessus d’une grille avec des plots (la bille pouvant tomber indifféremment à gauche ou à droite lorsqu’elle rencontre un plot), que l’on expliquera plus loin dans ce cours, montre que le tas de billes se répartit dans le réceptacle suivant une distribution de Gauss (voir la figure 2.2).

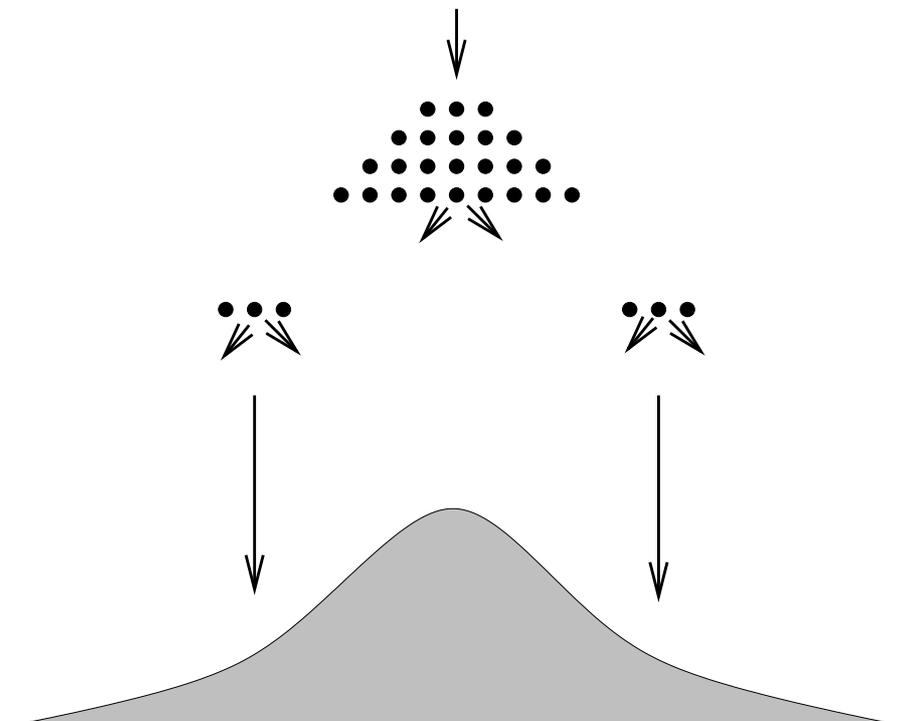


FIG. 2.2 – Le triangle de Galton

Si P est une probabilité sur \mathbf{R} (équipé de la tribu borélienne), la fonction

$$x \in \mathbf{R} \rightarrow P(] - \infty, x])$$

est dite *fonction de répartition* de la distribution de probabilité f sur \mathbf{R} ; cette fonction de répartition est une fonction croissante sur \mathbf{R} , continue à gauche, tendant vers 0 en $-\infty$ et vers 1 en $+\infty$. On peut d'ailleurs montrer que toute fonction F sur \mathbf{R} ayant ces trois propriétés est la fonction de répartition d'une distribution de probabilité sur \mathbf{R} ; s'il existe de plus une fonction positive intégrable f (d'intégrale 1) telle que

$$F(x) = \int_{]-\infty, x[} f(t) dt,$$

(par exemple, si F , en plus des trois propriétés mentionnées, est dérivable sur \mathbf{R} et de dérivée continue), la distribution de probabilité correspondant à F est une distribution à densité, de densité $f = F'$.

Tout triplet (Ω, \mathcal{T}, P) , où \mathcal{T} est une tribu sur l'ensemble d'évènements Ω et P une probabilité sur \mathcal{T} , est dit *espace probabilisé*.

2.3 Notions de probabilité induite et conditionnelle ; indépendance

Soit (Ω, \mathcal{T}, P) un espace probabilisé et A un élément de \mathcal{T} tel que $P(A) \neq 0$. La famille

$$\mathcal{T}_A = \{E \cap A, E \in \mathcal{T}\}$$

est encore une tribu, cette fois considérée comme famille de parties de A ; on définit une probabilité P_A sur (A, \mathcal{T}_A) en posant

$$P_A(B) := \frac{P(B)}{P(A)}, \quad B \in \mathcal{T}_A.$$

On dit que (A, \mathcal{T}_A) est ainsi équipé de la *probabilité induite par P sur A* .

On peut aussi définir une autre probabilité que P sur (Ω, \mathcal{T}) , liée cette fois à l'évènement A ; on la notera $P(\cdot | A)$ et on la définit par

$$P(B | A) := \frac{P(A \cap B)}{P(A)};$$

c'est la *probabilité conditionnelle attachée à P sous le conditionnement A* .

Si A_1, \dots, A_n sont n évènements d'un espace probabilisé (Ω, \mathcal{T}, P) tels que $P(A_1 \cap \dots \cap A_n) > 0$, on prouve par récurrence la formule suivante :

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1) \cdots P(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Exemple 2.3 On dispose d'un jeu de 32 cartes; on en tire 8 (et ce de manière indépendante (disons que ceci signifie, on le verra un peu plus loin, que l'espace des évènements est l'espace de

tous les choix de 8 cartes possibles, la distribution de probabilité étant la distribution uniforme). L'évènement $A := \text{on tire 2 valets}$ a pour probabilité

$$P(A) = \frac{\binom{4}{2} \binom{28}{6}}{\binom{32}{8}};$$

(on calcule son cardinal et l'on divise par le cardinal de l'ensemble des évènements, à savoir le nombre de combinaisons de 8 cartes parmi 32). L'évènement $B := \text{une carte tirée au moins est un valet}$ a pour probabilité

$$P(B) = 1 - \frac{\binom{28}{8}}{\binom{32}{8}}$$

et l'on a donc

$$P(A | B) = \frac{P(A)}{P(B)} = \frac{\binom{4}{2} \binom{28}{6}}{\binom{32}{8} - \binom{28}{8}}.$$

Une formule importante impliquant le conditionnement est la *formule de Bayes* ou encore *formule des probabilités totales*. Elle repose sur le fait très simple suivant ; si (Ω, \mathcal{T}, P) est un espace probabilisé et $(A_k)_{1 \leq k \leq m}$ une collection de m éléments de \mathcal{T} deux à deux disjoints et tels que

$$P(\Omega \setminus \bigcup_{k=1}^m A_k) = 0,$$

alors, pour tout évènement B de \mathcal{T} , on a

$$P(B) = \sum_{\{j, P(A_j) \neq 0\}} P(B | A_j) P(A_j).$$

Couplée avec la formule de Bayes, on a aussi la *formule des preuves* que l'on peut énoncer ainsi :

Soit (Ω, \mathcal{T}) un espace probabilisé et B_1, \dots, B_m, \dots une suite d'éléments deux à deux disjoints de \mathcal{T} dont l'union est de probabilité 1 (on dit encore un système complet d'évènements); on suppose tous les B_j de probabilité strictement positive; alors, pour tout élément A de \mathcal{T} de probabilité strictement positive, pour tout entier positif k , on a

$$P(B_k | A) = \frac{P(A | B_k) P(B_k)}{\sum_j P(A | B_j) P(B_j)}.$$

Voici l'interprétation de la formule des preuves (que l'on prouve en utilisant la formule de Bayes ainsi que le fait que $P(B_k | A) = P(A \cap B_k) / P(A) = P(A | B_k) P(B_k) / P(A)$) et à quoi elle sert concrètement : si A est un évènement de probabilité strictement positive et si les $B_k, k = 1, \dots, m$, constituent un jeu d'hypothèses, le problème que l'on se pose est celui de savoir si la réalisation de A confirme ou infirme la liste d'hypothèses $B_k, k = 1, \dots, m$. On connaît les probabilités *a priori* des $B_k, k = 1, \dots, m$, ainsi que les probabilités conditionnelles $P(A | B_k), k = 1, \dots, m$, et l'on souhaite calculer les probabilités *a posteriori* $P(B_k | A), k = 1, \dots, m$; le fait

que $P(B_k | A)$ soit voisin de 1 est la preuve que la réalisation de A confirme l'hypothèse B_k ; le fait que $P(B_k | A)$ soit voisin de 0 est, en revanche, une indication que la réalisation de A infirme l'hypothèse B_k .

Exemple 2.4 (le processus de désintégration atomique) Supposons que les atomes d'un corps radioactif se désintègrent de manière aléatoire. On considère un espace probabilisé (Ω, \mathcal{T}, P) adapté aux règles imposées par le physicien, à savoir la règle suivante : la probabilité que, si un atome n'est pas désintégré à l'instant t_0 , il se désintègre dans l'intervalle de temps $[t_0, t_0 + t]$, est une fonction *qui ne dépend que de t* (fonction que l'on appellera $F(t)$) ; regardons l'évolution d'un atome de la population, intact à l'instant $t = 0$; appelons A_s l'évènement *l'atome est encore intact à l'instant $t \geq 0$* . On a $P(A_s) = 1 - F(s)$ et

$$P(A_{s+t}) = P(A_{s+t} | A_s)P(A_s), \quad s, t \geq 0$$

soit, en tenant compte de ce qui régit le processus,

$$F(s+t) = F(s)F(t) \quad s, t \geq 0.$$

En résolvant cette équation fonctionnelle classique, on voit qu'il existe un réel (forcément positif) λ tel que $F(t) = e^{-\lambda t}$, d'où

$$P(A_t) = \exp(-\lambda t).$$

On a $F(t) = 1 - e^{-\lambda t}$; ce type de processus est dit *processus exponentiel*.

Si A et B sont deux évènements d'un espace probabilité (Ω, \mathcal{T}, P) avec $P(A) > 0$, alors $P(A | B) = P(A)$ signifie intuitivement que la réalisation ou la non réalisation de B n'influe pas sur celle de A ; les évènements A et B sont alors dits indépendants. Plus généralement, on a la définition suivante :

Étant donné un espace probabilisé (Ω, \mathcal{T}, P) , deux évènements A et B , éléments de \mathcal{T} sont dits indépendants si et seulement si $P(A \cap B) = P(A)P(B)$.

Attention ! Il ne faut pas confondre cette notion d'indépendance (qui est une notion probabiliste faisant intervenir la probabilité P) avec la notion d'*incompatibilité*, qui elle est une notion purement ensembliste : deux évènements A et B sont dits *incompatibles* si et seulement si $A \cap B$ est vide.

Cette notion s'étend au cadre des familles d'évènements :

*Étant donné un espace probabilisé (Ω, \mathcal{T}, P) et une collection finie ou dénombrable d'évènements $(A_j)_{j \in J}$, les A_j sont dits *mutuellement indépendants* si et seulement si, pour tout sous ensemble fini K dans J , on a*

$$P\left(\bigcap_{j \in K} A_j\right) = \prod_{j \in K} P(A_j).$$

*Étant donnée une collection finie ou dénombrable $(\mathcal{T}_j)_{j \in J}$ de sous-tribus de \mathcal{T} , les \mathcal{T}_j sont dites *mutuellement indépendantes* si et seulement si toute collection (A_j) avec $A_j \in \mathcal{T}_j$ est une collection d'évènements mutuellement indépendants.*

Si (Ω, \mathcal{T}, P) est un espace probabilisé, deux évènements A et B de \mathcal{T} sont indépendants si et seulement si leurs complémentaires A^c et B^c dans Ω sont indépendants : on écrit en effet

$$P(A^c)P(B^c) = (1 - P(A))(1 - P(B)) = 1 - P(A) - P(B) + P(A)P(B) ;$$

cette expression vaut $P((A \cup B)^c) = P(A^c \cap B^c)$ si et seulement si $P(A)P(B) = P(A \cap B)$, ce qui prouve l'assertion.

Etant donné un espace probabilisé (Ω, \mathcal{T}, P) , deux évènements A et B tels que $P(A)P(A^c) > 0$ sont indépendants si et seulement si

$$P(B) = P(B | A) = P(B | A^c).$$

Exemple 2.5 Considérons l'épreuve consistant à tirer un nombre au hasard entre 1 et n , tous les choix étant équiprobables. L'ensemble des évènements élémentaires est $\Omega = \{1, \dots, n\}$, la probabilité est la loi uniforme. Si p est un diviseur premier de n , notons E_p l'évènement *le nombre tiré est divisible par p* . Il est facile de constater, en utilisant le lemme de Gauss (si un nombre premier divise un produit de facteurs, il divise automatiquement l'un au moins des facteurs) que la collection d'évènements (E_{p_j}) , où $(p_j)_j$ est la famille des diviseurs premiers de n , est une collection d'évènements mutuellement indépendants. Alors il en est de même pour la collection des $(E_{p_j}^c)$; ceci implique que la probabilité de l'évènement $A :=$ *le nombre tiré est premier avec n* vaut

$$P(A) = \prod_j \left(1 - \frac{1}{p_j}\right);$$

en évaluant $P(A)$ comme le quotient du nombre de cas favorables sur le nombre de cas possibles, on retrouve ainsi la formule d'Euler : si $\varphi(n)$ désigne le nombre d'entiers entre 1 et n premiers avec n , alors

$$\varphi(n) = n \prod_{p|n, p \text{ premier}} \left(1 - \frac{1}{p_j}\right).$$

2.4 Variables aléatoires.

2.4.1 Variables aléatoires discrètes

Soit un espace probabilisé (Ω, \mathcal{T}, P) ; on appelle *variable aléatoire discrète* sur (Ω, \mathcal{T}, P) toute application de Ω dans un ensemble E fini ou dénombrable telle que l'image réciproque de tout sous-ensemble de E soit un élément de \mathcal{T} .

On pourrait penser, à lire cette définition, que la probabilité P ne joue aucun rôle; de fait, ceci n'est pas le cas, car la donnée d'une variable aléatoire présuppose toujours la donnée d'une distribution de probabilité P sur l'ensemble Ω initial; en effet, la connaissance de cette distribution de probabilité P sur l'espace des évènements, distribution de probabilité conditionnant la règle du jeu, joue un rôle essentiel dans la définition de ce que l'on appelle la *loi de la variable aléatoire X* (définition que l'on donnera un peu plus loin). Or ce qui est important lorsque l'on considère une variable aléatoire n'est pas tant la fonction elle-même que la loi de probabilité à laquelle cette fonction obéit (loi qui dépend d'une part de la fonction, d'autre part de la probabilité P sur l'espace Ω des évènements).

L'ensemble des évènements élémentaires est un ensemble correspondant à tous les résultats possibles de l'épreuve. Donnons deux exemples de variables aléatoires, attachés à deux épreuves souvent rencontrées : les tirages dans une urne avec remise et le jeu de pile ou face.

Exemple 2.6

- **1.** On considère une urne contenant B boules blanches et R boules rouges, hors de laquelle on puise n fois de suite une boule, avec remise à chaque fois, les divers tirages successifs étant indépendants (ce qui ne serait bien sûr pas le cas si les tirages s'effectuaient sans remise!) Le modèle d'espace d'évènements Ω associé à cette épreuve est l'espace $\{1, \dots, B + R\}^n$, de cardinal $(B + R)^n$, la tribu étant la famille de toutes les parties de $\{1, B + R\}^n$ (les boules

sont numérotées de 1 à $B + R$; la probabilité que l'on met sur cet espace Ω est la probabilité uniforme [on pourrait d'ailleurs aussi prendre comme modèle d'espace d'événements $\{0, 1\}^N$, avec cette fois $P(E_1 \times E_2 \cdots \times E_n) = P(E_1) \cdots P(E_n)$ si $E_k \subset \{0, 1\}$, avec $P(\{0\}) = B/(B + R)$ et $P(\{1\}) = R/(B + R)$]. Un exemple de variable aléatoire discrète sur $(\Omega, \mathcal{P}(\Omega), P)$ est par exemple le nombre de boules rouges tirées au terme des n tirages successifs ; c'est une variable aléatoire à valeurs dans $\{0, \dots, n\}$, ensemble fini à $n + 1$ éléments.

- 2. Soit $\Omega = \{0, 1\}^{\mathbf{N}^*}$ l'ensemble de toutes les suites indexées par \mathbf{N}^* constituées de 0 ou de 1, équipé de la plus petite tribu contenant les ensembles du type $E_1 \times E_2 \times E_3 \times \cdots$, où $E_k \subset \{0, 1\}$ et $E_k = \{0, 1\}$ sauf pour un nombre fini d'indices k . Cet ensemble Ω modélise par exemple l'ensemble de tous les résultats possibles d'un jeu sans fin à pile ou face. La probabilité que l'on met est celle qui consiste à poser

$$P(E_1 \times E_2 \times \cdots) = P(E_1) \times P(E_2) \times \cdots ;$$

la règle du jeu est que les résultats des lancers correspondant à des jets différents sont des événements indépendants et l'on se reporte à la définition de la mutuelle indépendance entre événements introduite dans la section précédente pour justifier la définition que l'on vient de prendre pour la probabilité sur Ω ; on convient ensuite de ce que, pour un jet donné

$$\begin{aligned} P(\text{on obtient pile}) &= p \\ P(\text{on obtient face}) &= 1 - p, \end{aligned}$$

où $p \in]0, 1[$ (la pièce peut être truquée si $p \neq 1/2$, mais le jeu de pile ou face n'est pas, lui, biaisé, au vu de la probabilité que l'on a mis sur $\{0, 1\}^{\mathbf{N}}$). Un exemple de variable aléatoire discrète sur (Ω, \mathcal{T}, P) (à valeurs dans $\mathbf{N} \cup \{+\infty\}$) est la variable aléatoire discrète correspondant au numéro du jet où l'on obtient *pile* pour la première fois.

Par définition, la loi d'une variable aléatoire $X : (\Omega, \mathcal{P}, P) \rightarrow E$ (où E est un ensemble fini ou dénombrable) est la probabilité sur E (équipé de la tribu de toutes ses parties) que l'on définit en transportant *via* X la probabilité P (il était donc bien indispensable de disposer d'une probabilité, c'est-à-dire en quelque sorte d'une règle du jeu) sur l'espace probabilisable (Ω, \mathcal{T}) . La loi de X est donc la distribution de probabilité P_X sur $(E, \mathcal{P}(E))$ définie par

$$P_X(A) := P(\{\omega \in \Omega ; X(\omega) \in A\}) = P(X^{-1}(A)).$$

Reprenons les deux exemples ci-dessus :

Exemple 2.7

- 1. Dans le premier exemple, on a

$$P(X = k) = \binom{n}{k} \left(\frac{R}{B + R}\right)^k \times \left(\frac{B}{B + R}\right)^{n-k}$$

car le dénombrement des cas favorables donne

$$\binom{n}{k} \times R^k \times B^{n-k} ;$$

si l'on pose $\tau = R/(B + R)$, la loi de X est donnée par

$$P_X(\{k\}) = \binom{n}{k} \tau^k (1 - \tau)^{n-k}, \quad k = 0, \dots, n.$$

(on vérifie que la somme de ces nombres vaut bien 1 à cause de la formule du binôme). Cette loi de probabilité très importante (régissant ici la variable aléatoire X du premier exemple) est dite *loi binomiale* de paramètre $\tau \in [0, 1]$.

– 2. Dans le second exemple, on a

$$P(X = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots;$$

on note que

$$\sum_{k=1}^{\infty} (1 - p)^{k-1}p = p \times \sum_{k=0}^{\infty} (1 - p)^k = \frac{p}{1 - (1 - p)} = 1,$$

ce qui montre que $P(X = \infty) = 0$ (ce d'ailleurs quelque soit la valeur de p , c'est-à-dire que la pièce soit truquée ou non, ce qui intuitivement n'est pas évident!). La loi de probabilité P_X sur \mathbb{N}^* définie par

$$P_X(\{k\}) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots$$

est dite *loi de Pascal* ou encore *loi géométrique* de raison $1 - p$; c'est la version discrète de la *loi exponentielle* que l'on retrouvera plus loin.

Une loi de probabilité sur \mathbb{N} joue un rôle très important en physique; cette loi est dite *loi de Poisson de paramètre $\lambda > 0$* ; elle correspond à la distribution de probabilité sur \mathbb{N} définie par

$$P(\{k\}) = \exp(-\lambda) \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

Voici la raison pour laquelle cette loi est si importante, en même temps que sa modélisation comme loi d'une certaine variable aléatoire sur un espace probabilisé *ad hoc*. Nous supposons que nous disposons d'une liste infinie (dénombrable) d'instantanés $\tau_i, i \in \mathcal{I}$. Ces instantanés seront des instantanés dits "marqués" sur un intervalle de l'axe réel de longueur $T > 0$, par exemple $[0, T]$, ce marquage s'effectuant de manière stochastique. La règle initiale de marquage est simple : on décide que la probabilité, lorsque l'on marque N points de cet intervalle, pour que k d'entre eux soient dans l'intervalle (t_1, t_2) , est

$$p_{t_1, t_2, N}(k) = \binom{N}{k} \left(\frac{t_2 - t_1}{T}\right)^k \left(1 - \left(\frac{t_2 - t_1}{T}\right)\right)^{N-k},$$

ce qui correspond à l'idée que le marquage des instantanés s'effectue de manière uniforme. Si l'on suppose N très grand et $t_2 - t_1$ petit devant T , on a l'approximation classique de la loi binomiale

$$p_{t_1, t_2, N}(k) \simeq \exp\left(-\frac{N(t_2 - t_1)}{T}\right) \times \frac{1}{k!} \left(\frac{N(t_2 - t_1)}{T}\right)^k,$$

approximation qui résulte de la formule de Stirling

$$n! \simeq \sqrt{2\pi} n^{n+1/2} e^{-n}.$$

Si l'on suppose maintenant que N et T tendent vers l'infini, mais que le nombre N/T reste constant et égal à un paramètre $\lambda > 0$, l'approximation devient

$$p_{t_1, t_2}(k) \simeq e^{-\lambda(t_2 - t_1)} \frac{1}{k!} (\lambda(t_2 - t_1))^k.$$

Le paramètre λ peut être interprété comme la densité de l'ensemble des instantanés marqués. Ce nombre $p_{t_1, t_2}(k)$ s'interprète comme la probabilité que l'on marque k points dans l'intervalle (t_1, t_2) , λ représentant la densité des points (ou particules) marquées.

On retrouve la loi de Poisson dans les processus d'émission de particules. Les signaux enregistrés dans les techniques de CAT-Scanner ou d'échographie par exemple sont des bruits poissonniens, c'est-à-dire des signaux correspondant à des variables aléatoires suivant une loi de Poisson (en déterminant le paramètre, on retrouve la densité de rayonnement de l'organe).

2.4.2 Variables aléatoires réelles ou vecteurs de variables aléatoires réelles

Soit (Ω, \mathcal{T}, P) un espace probabilisé; une variable aléatoire sur (Ω, \mathcal{T}, P) et à valeurs réelles (resp. à valeurs dans \mathbb{R}^n) est une application X de l'espace des événements Ω dans \mathbb{R} (resp. dans \mathbb{R}^n) telle que, pour tout sous-ensemble quarrable A de \mathbb{R} (resp. \mathbb{R}^n), l'ensemble

$$\{\omega \in \Omega; X(\omega) \in A\}$$

appartienne à la tribu \mathcal{T} .

Une telle variable aléatoire sur (Ω, \mathcal{T}, P) induit par transport une probabilité sur \mathbb{R} (resp. \mathbb{R}^n) équipé de la tribu des ensembles quarrables (tribu de Lebesgue). Cette distribution de probabilité P_X sur \mathbb{R} (resp. \mathbb{R}^n) est dite *loi de la variable aléatoire réelle X* (resp. *loi du vecteur de variables aléatoires X*).

Une loi de probabilité sur \mathbb{R} (modélisable comme la loi d'une certaine variable aléatoire) joue un rôle important en physique, la *loi de Gauss normale*, définie par

$$P(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-t^2/2} dt$$

si A est un sous-ensemble quarrable de \mathbb{R} ; on a aussi une loi de probabilité sur \mathbb{R}^n en posant

$$P(A) = (2\pi)^{-n/2} \int \dots \int_A \exp(-(x_1^2 + \dots + x_n^2)/2) dx_1 \dots dx_n$$

pour tout sous-ensemble quarrable A de \mathbb{R}^n . Il s'agit d'une distribution de probabilité car

$$\left(\int_{\mathbb{R}} \exp(-t^2/2) dt \right)^2 = \iint_{\mathbb{R}^2} \exp(-(t^2 + u^2)/2) dt du = \int_0^{+\infty} \int_0^{2\pi} e^{-r^2/2} r dr d\theta = 2\pi$$

grâce au théorème de Fubini et à la formule de changement de variables (passage aux coordonnées polaires).

L'expérience du triangle de Galton (figure 2.2) modélise la distribution normale de Gauss comme la loi d'une variable aléatoire; cette expérience illustre d'ailleurs un théorème limite de la théorie des probabilités, le *théorème limite centrale* que nous verrons plus loin.

Une raison importante qui explique le rôle majeur de la distribution de Gauss en physique est que la fonction

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

se transforme par la prise de spectre essentiellement en elle-même, plus précisément en la fonction

$$\omega \rightarrow e^{-\omega^2/2}.$$

Cette fonction réalise le meilleur compromis dans le *principe d'incertitude* d'Heisenberg (impossibilité de localiser simultanément une particule et son spectre) et fournit donc un modèle raisonnable pour les particules en mécanique quantique (que ce soit sous l'angle déterministe ou sous l'angle probabiliste).

2.5 Variables aléatoires réelles à densité

Soit (Ω, \mathcal{T}, P) un espace probabilisé et X une variable aléatoire réelle (resp. un vecteur de variables aléatoires réelles, c'est-à-dire une variable aléatoire à valeurs dans \mathbb{R}^n).

On dit que X est une variable aléatoire à densité (resp. un vecteur variables aléatoires à densité) s'il existe une fonction $f : t \rightarrow f(t)$, positive, intégrable sur \mathbb{R} et d'intégrale 1 sur \mathbb{R} (resp. une fonction $f : (x_1, \dots, x_n) \rightarrow f(x_1, \dots, x_n)$) telle que la loi de X soit donnée par

$$P_X(A) = \int_A f(t) dt$$

pour tout sous-ensemble quarrable A de \mathbb{R} (resp.

$$P_X(A) = \int \dots \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n$$

pour tout sous-ensemble quarrable A de \mathbb{R}^n). La fonction f est alors appelée *densité* de la variable aléatoire X (resp. *densité* du vecteur de variables aléatoires X). On dit que la variable aléatoire X (resp. le vecteur de variables aléatoires X) est une *variable aléatoire réelle à densité* (resp. un *vecteur de variables aléatoires réelles à densité*).

Bien sûr, toutes les variables aléatoires réelles ne sont pas à densité ; il est d'ailleurs très important de remarquer que si X est une variable aléatoire réelle à densité (resp. un vecteur de variables aléatoires réelles à densité) sur un espace probabilisé (Ω, \mathcal{T}, P) , alors, pour tout $t \in \mathbb{R}$, on a $P(\{\omega ; X(\omega) = t\}) = 0$ (resp. pour tout $x \in \mathbb{R}^n$, $P(\{\omega ; X(\omega) = x\}) = 0$). Une variable aléatoire (resp. un vecteur de variables aléatoires) prenant ses valeurs dans un sous-ensemble dénombrable de \mathbb{R} (resp. de \mathbb{R}^n) ne saurait être à densité ; si par exemple X suit une loi de Poisson de paramètre $\lambda > 0$ sur (Ω, \mathcal{T}, P) , on a, pour tout $n \in \mathbb{N}$, $P(X = n) = e^{-\lambda} \lambda^n / n! > 0$.

Exemple 2.8

- une variable aléatoire suivant une loi normale ($P_X(A) = (2\pi)^{-1/2} \int_A e^{-t^2/2} dt$) est une variable aléatoire à densité ; plus généralement, si m_1, \dots, m_n sont des nombres réels, $\sigma_1, \dots, \sigma_n$ des nombres réels strictement positifs

$$f(x_1, \dots, x_n) = (2\pi)^{-n/2} (\sigma_1 \dots \sigma_n)^{-1/2} \exp \left(- \sum_{j=1}^n \frac{(x_j - m_j)^2}{2\sigma_j} \right),$$

- un vecteur de variables aléatoires ayant pour densité f est dit *vecteur gaussien* ; de même que les variables normales jouent un rôle clef en physique (particulièrement, on l'a vu, en physique quantique), les vecteurs gaussiens (qui en sont la généralisation naturelle après translation, dilatation, et passage d'une variable à plusieurs) ont aussi un rôle essentiel ;
- on retrouve la variable aléatoire réelle de densité

$$f(t) := \begin{cases} \lambda e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$$

dans le processus de désintégration atomique (voir l'exemple 2.4).

2.6 Fonction de répartition d'une variable aléatoire réelle

Si X est une variable aléatoire réelle sur un espace probabilisé (Ω, \mathcal{T}, P) , on préfère souvent retenir sa loi en retenant sa *fonction de répartition*, c'est-à-dire la fonction

$$t \rightarrow P_X(]-\infty, t]) = P(X \in]-\infty, t]).$$

Cette fonction est croissante, de limite 0 en $-\infty$, de limite 1 en $+\infty$, et est continue à gauche; d'ailleurs une fonction de \mathbf{R} dans $[0, 1]$ ayant ces quatre propriétés est la fonction de répartition d'une variable aléatoire.

Si la fonction de répartition est de classe C^1 et de dérivée f , la variable aléatoire réelle correspondante est une variable à densité, de densité précisément f .

Exemple 2.9

- La fonction de répartition d'une variable aléatoire gaussienne suivant une loi normale est

$$t \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du ;$$

seule une table fournit les valeurs de cette fonction (inexprimable en termes de fonctions simples).

- La fonction de répartition d'une variable aléatoire à densité $t \rightarrow \lambda e^{-\lambda t}$ pour $t \geq 0$ est

$$t \rightarrow \lambda \int_0^t e^{-\lambda u} du = [-e^{-\lambda u}]_0^t = 1 - e^{-\lambda t}$$

si $t \geq 0$ et vaut 0 si $t < 0$.

2.7 Espérance et variance d'une variable aléatoire

2.7.1 Le cas des variables aléatoires à valeurs dans un sous-ensemble fini ou dénombrable de \mathbf{R}^n

Si X est une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{T}, P) et à valeurs dans un sous-ensemble fini de \mathbf{R}^n , on appelle *espérance* de X la "valeur moyenne" que prend X ; comme X prend une valeur donnée x_k avec la probabilité $p_k = P(X = x_k) = P_X(\{x_k\})$, cette valeur moyenne vaut naturellement

$$E[X] := \sum_k x_k P(X = x_k)$$

(la somme des valeurs prises, ces valeurs étant "pondérées" de coefficients correspondant aux probabilités respectives avec lesquelles elles sont prises). On retrouve bien ici le procédé classique pour calculer une moyenne en tenant compte de coefficients.

Si maintenant X prend ses valeurs dans un sous-ensemble dénombrable de \mathbf{R}^n , $X(\Omega) = \{x_k; k \in \mathbf{N}\}$, on dit que X a une *espérance* si et seulement si la quantité

$$\sum_k \|x_k\| P(X = x_k)$$

est finie ; si c'est le cas, on peut poser sans ambiguïté

$$E(X) = \lim_{N \rightarrow \infty} \sum_{k=0}^N x_k P(X = x_k)$$

et l'espérance (ou plutôt le *vecteur des espérances*) si X est à valeurs vectorielles est bien définie sur le modèle d'une valeur moyenne comme dans le cas où X prend ses valeurs dans un ensemble fini.

Dans le cas particulier où X est une variable aléatoire réelle positive sur un espace probabilisé (Ω, \mathcal{T}, P) et telle que

$$\sum_k |x_k| P(X = x_k) = \sum_k x_k P(X = x_k) = +\infty,$$

on peut encore définir l'espérance de X et l'on dit que X est d'espérance infinie. Si X est à valeurs vectorielles ou à valeurs réelles non positives, la clause de sécurité

$$\sum_k \|x_k\| P(X = x_k) < \infty$$

est essentielle pour pouvoir parler de l'espérance de X .

Exemple 2.10

- Si X suit une loi de Poisson de paramètre $\lambda > 0$, on a

$$E[X] := \sum_{k=0}^{\infty} k \times e^{-\lambda} (\lambda^k / k!) = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \lambda^{k-1} / (k-1)! = \lambda e^{-\lambda} \times e^{\lambda} = \lambda ;$$

l'espérance est donc égale au paramètre pour une telle loi.

- Supposons que X suive une loi binomiale de paramètres (n, τ) ; pour calculer l'espérance, on utilise une méthode classique empruntée à Polya, dite *méthode des séries génératrices* ; soit $\tau' = 1 - \tau$; on calcule

$$(\tau x + \tau')^n = \sum_{k=0}^n \binom{n}{k} \tau^k (\tau')^{n-k} x^k$$

puis on dérive, ce qui donne

$$n\tau(\tau x + \tau')^{n-1} = \sum_{k=1}^n k \binom{n}{k} \tau^k (\tau')^{n-k} x^{k-1} ;$$

on évalue enfin en 1, ce qui donne $E[X] = \sum_{k=0}^n k P(\{X = k\}) = np$.

- Supposons que X suive une loi géométrique de paramètre p ; on écrit, pour $x \in]-1, 1[$,

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k,$$

puis on dérive, ce qui donne

$$\left(\frac{1}{1-x} \right)^2 = \sum_{k=1}^{\infty} k x^{k-1}$$

d'où

$$E[X] = (1-p) \sum_{k=1}^{\infty} k p^k = \frac{p}{1-p}.$$

- Le mécanisme est un peu plus complexe pour la loi hypergéométrique de paramètres rationnels ($n, \tau_1 = R/N, \tau_2 = (N - R)/N$) ; on part de

$$(1 + xy)^R(1 + y)^{N-R} = \sum_{n=0}^{\infty} \left(\sum_{k_1+k_2=n} \binom{R}{k_1} \binom{N-R}{k_2} x^{k_1} \right) y^n$$

avec les conventions

$$\binom{R}{k_1} = 0, \quad k_1 > R$$

et

$$\binom{N-R}{k_2} = 0, \quad k_2 > N - R;$$

en dérivant par rapport à x , on obtient

$$yR(1 + xy)^{R-1}(1 + y)^{N-R} = \sum_{n=0}^{\infty} \left(\sum_{\substack{k_1+k_2=n \\ k_1 > 0}} k_1 \binom{R}{k_1} \binom{N-R}{k_2} x^{k_1-1} \right) y^n;$$

faisant $x = 1$ et identifiant les coefficients de y^n , on trouve

$$E[X] = nR/N;$$

la loi hypergéométrique de paramètres $(n, \tau, 1 - \tau)$ a donc, comme la loi binomiale de paramètres n, τ , pour espérance $n\tau$ (que τ soit rationnel ou non).

Connaître l'espérance d'une variable aléatoire réelle (ou d'un vecteur de variables aléatoires réelles) prenant au plus une infinité dénombrable de valeurs ne suffit pas à l'analyse fine du résultat de l'épreuve dont cette variable aléatoire affiche le résultat. Il est naturel de concevoir, lorsque vous disposez par exemple de la totalité des notes d'une épreuve d'examen, que la moyenne de ces notes n'est pas une information suffisante sur le résultat : comment par exemple discerner, si la moyenne est 10/20 le cas où toutes les notes sont agglutinées à la moyenne et le cas où les copies se rangent en deux paquets, la moitié des notes se situant entre 0 et 5, l'autre moitié entre 15 et 20 ? Pour compléter l'information relative à une variable aléatoire réelle donnée (ou à un vecteur de variables aléatoires réelles, les variables prenant toujours ici au plus une infinité dénombrable de valeurs), il faut définir, lorsque cela est possible, une autre notion, celle de *variance*.

Soit X une variable aléatoire réelle comme ci-dessus ou un vecteur de variables aléatoires réelles défini sur un espace probabilisé (Ω, \mathcal{T}, P) . On dit que X *admet une variance* (ou encore est à *variance finie*) si et seulement si la variable positive $\|X\|^2$ admet une espérance ; ceci implique que le vecteur X a une espérance et l'on définit la variance de X par

$$V(X) := E[\|X - E[X]\|^2].$$

L'*écart type* est alors la racine carrée de la variance.

$$\sigma(X) := \sqrt{E[\|X - E[X]\|^2]}.$$

L'écart type mesure donc ce en quoi diffère la variable X d'une variable prenant la valeur $E[X]$, soit la valeur moyenne de X , avec la probabilité 1.

Exemple 2.11

- Un calcul inspiré des techniques utilisant les fonctions génératrices (comme pour le calcul de l'espérance) donne la variance d'une loi binomiale de paramètres (n, τ) ; on a alors pour une telle loi $V(X) = n\tau(1 - \tau)$.
- Une variable suivant une loi de Poisson de paramètre λ a pour variance $E[X] = \lambda$.

2.7.2 Le cas général des variables aléatoires réelles ou des vecteurs de variables aléatoires à densité

Supposons maintenant que X soit une variable aléatoire réelle positive à densité sur un espace probabilisé (Ω, \mathcal{T}, P) ; pour chaque $t \in [0, \infty[$, on sait que $P(X = t) = 0$; par contre, la quantité infinitésimale $P(t \leq X \leq t + dt)$ vaut, elle, $f(t) dt$ (c'est précisément la définition de la densité). L'analogie continue de l'expression qui nous a permis la définition de l'espérance est donc

$$E[X] = \sum_{t \geq 0} tP(X \in [t, t + dt]) \simeq \int_0^{\infty} tf(t)dt \in [0, \infty].$$

Si maintenant X est une variable aléatoire réelle à densité, que l'on peut écrire $X = X^+ - X^-$, où $X^+ := \sup(X, 0)$ est à densité $f \times \chi_{[0, \infty[}$ et $X^- := \sup(-X, 0)$ est à densité $f(-t) \times \chi_{]0, \infty[}$, on dit que X admet une espérance si et seulement si les quantités $E[X^+]$ et $E[X^-]$ sont finies, ce qui équivaut à dire que

$$\int_{\mathbb{R}} |t| f(t) dt < \infty;$$

l'espérance de X est alors

$$E[X] = E[X^+] - E[X^-] = \int_{\mathbb{R}} t f(t) dt.$$

Si enfin $X = (X_1, \dots, X_n)$ est un vecteur de variables aléatoires à densité sur (Ω, \mathcal{T}, P) , de densité la fonction intégrable $f(x_1, \dots, x_n)$, on dit que X admet une espérance si

$$\int \dots \int_{\mathbb{R}^n} \|x\| f(x_1, \dots, x_n) dx_1 \dots dx_n < \infty;$$

on appelle alors espérance de (X_1, \dots, X_n) (ou plutôt *vecteur des espérances de X*) le vecteur de \mathbb{R}^n suivant :

$$\left(\int \dots \int_{\mathbb{R}^n} x_1 f(x_1, \dots, x_n) dx_1 \dots dx_n, \dots, \int \dots \int_{\mathbb{R}^n} x_n f(x_1, \dots, x_n) dx_1 \dots dx_n \right).$$

Exemple 2.12

- Le vecteur des espérances d'un vecteur gaussien (X_1, \dots, X_n) de densité

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\sigma_1 \dots \sigma_n}} \exp\left(-\sum_{j=1}^n \frac{(x_j - m_j)^2}{2\sigma_j}\right)$$

vaut (m_1, \dots, m_n) ; un tel vecteur modélise une particule vivant statistiquement près du point (m_1, \dots, m_n) .

- Une variable aléatoire réelle de densité

$$f(t) = \frac{1}{\pi(1+t^2)}$$

(loi de Poisson continue) n'a pas d'espérance car

$$\int_{\mathbb{R}} \frac{|t|}{1+t^2} dt = +\infty.$$

- Une variable aléatoire réelle à densité $\lambda e^{-\lambda t}$ pour $t > 0$ (comme dans le processus de désintégration atomique) a pour espérance

$$E[X] = \lambda \int_0^{\infty} t e^{-\lambda t} dt = \left[-t e^{-\lambda t} \right]_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt = 1/\lambda.$$

On se contentera dans ce cours de définir l'espérance des variables aléatoires réelles à densité, mais on notera néanmoins qu'en couplant ce sous-paragraphe avec le précédent, on peut définir l'espérance d'une variable aléatoire réelle se présentant comme la somme d'une variable aléatoire X_{cont} réelle à densité et d'une variable aléatoire réelle X_{disc} prenant au plus une infinité dénombrable de valeurs. L'espérance de X existe si et seulement si les quantités $E[|X_{\text{cont}}|]$ et $E[|X_{\text{disc}}|]$ sont finies et vaut alors

$$E[X] := E[X_{\text{cont}}] + E[X_{\text{disc}}],$$

ces deux quantités ayant été définies respectivement dans ce qui précède et dans la sous-section précédente.

Ceci se trouve justifié par une propriété importante de l'espérance : le fait que la prise d'espérance soit une opération linéaire, c'est-à-dire que, si X et Y sont deux vecteurs de variables aléatoires réelles (de (Ω, \mathcal{T}, P) dans \mathbb{R}^n) et si $E[||X||]$ et $E[||Y||]$ sont des quantités finies, alors

$$E[\lambda X + \mu Y] = \lambda E[X] + \mu E[Y], \quad \forall \lambda, \mu \in \mathbb{R}.$$

Prouvons juste cela au moins heuristiquement si $n = 1$. On a

$$\begin{aligned} E[\lambda X + \mu Y] &\simeq \sum_{t,s} (\lambda t + \mu s) P(X \in [t, t + dt], Y \in [s, s + ds]) \\ &\simeq \lambda \sum_t t \left(\sum_s P(X \in [t, t + dt], Y \in [s, s + ds]) \right) \\ &\quad + \mu \sum_s s \left(\sum_t P(X \in [t, t + dt], Y \in [s, s + ds]) \right) \\ &\simeq \lambda \sum_t t P(X \in (t, t + dt]) + \mu \sum_s s P(X \in [s, s + ds]) \\ &\simeq \lambda E[X] + \mu E[Y]. \end{aligned}$$

La linéarité de l'espérance est un fait très important. Si l'on y réfléchit bien, on voit que cette linéarité s'explique exactement comme la linéarité de l'intégrale, mais cette fois l'ensemble de départ des fonctions n'est plus \mathbb{R}^n comme au chapitre 1, mais un ensemble abstrait Ω sur lequel, comme pour \mathbb{R}^n , on a le moyen de "mesurer" certains ensembles dits quarrables (ici en l'occurrence les ensembles éléments de la tribu \mathcal{T}).

Si X est un vecteur de variables aléatoires réelles à densité (ou plus généralement la somme d'un tel vecteur $X_{\text{cont}} : (\Omega, \mathcal{T}, P) \rightarrow \mathbb{R}^n$ et d'un vecteur X_{disc} de variables aléatoires de (Ω, \mathcal{T}, P) dans un sous-ensemble au plus dénombrable de \mathbb{R}^n), on dit que X a une variance si et seulement si $E[||X_{\text{cont}}||^2]$ et $E[||X_{\text{disc}}||^2]$ sont finies. La variance de X est alors

$$V(X) := E[||X - E[X]||^2].$$

On remarque que si X est une variable aléatoire réelle ayant une variance, alors la linéarité de l'espérance implique :

$$\begin{aligned} V(X) = E[(X - E[X])^2] &= E[X^2 - 2E[X]X + (E[X])^2] \\ &= E[X^2] - 2(E[X])^2 + (E[X])^2 = E[X^2] - (E[X])^2 \end{aligned}$$

puisque l'espérance d'une variable aléatoire constante $X = C$ vaut C (on applique ceci avec $C = E[X]$).

Contrairement à la prise d'espérance, la prise de variance n'est pas une opération linéaire ; comme la prise d'énergie en physique, c'est une opération quadratique ; le calcul de la variance d'une somme de deux variables aléatoires réelles ayant toutes les deux une variance fait apparaître des termes croisés (ou encore d'interférence) que dans le vocabulaire des probabilités, on appelle *termes de corrélation* ; plus précisément, on a, si X et Y sont deux variables aléatoires réelles ayant toutes les deux une variance et définies sur le même espace probabilisé (Ω, \mathcal{T}, P) :

$$\begin{aligned} V(X + Y) &= E[(X + Y)^2] - (E[X] + E[Y])^2 \\ &= V(X) + V(Y) + 2E[(X - E[X])(Y - E[Y])]; \end{aligned}$$

la quantité "mixte"

$$E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

(qui existe dès que X et Y sont deux variables aléatoires réelles telles que X^2 et Y^2 soient d'espérance finie), est dite *covariance* des variables X et Y .

La notion de covariance se généralise au cas d'un vecteur $X = (X_1, \dots, X_n)$ de variables aléatoires réelles tels que $E[\|X\|^2]$ soit fini : la *matrice de covariance* de X symétrique

$$\text{covar}(X_1, \dots, X_n) := \left(E[(X_i - E[X_i])(X_j - E[X_j])] \right)_{1 \leq i, j \leq n}.$$

La variance de $\lambda_1 X_1 + \dots + \lambda_n X_n$ vaut

$$\text{Var}(\lambda_1 X_1 + \dots + \lambda_n X_n) = (\lambda_1, \dots, \lambda_n) \bullet \text{covar}(X_1, \dots, X_n) \bullet \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

Si X_1, \dots, X_n sont à valeurs complexes, la matrice de covariance est

$$\text{covar}(X_1, \dots, X_n) := \left(E[(X_i - E[X_i])(\overline{X_j - E[X_j]})] \right)_{1 \leq i, j \leq n}$$

et l'on a dans ce cas

$$\text{Var}(\lambda_1 X_1 + \dots + \lambda_n X_n) = (\overline{\lambda_1}, \dots, \overline{\lambda_n}) \bullet \text{covar}(X_1, \dots, X_n) \bullet \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}, \quad \forall \lambda \in \mathbb{C}^n.$$

2.7.3 Les inégalités de Markov et de Bienaymé-Tchebychev

Soit (Ω, \mathcal{T}, P) un espace probabilisé et X une variable aléatoire réelle positive sur cet espace ; alors, si λ est un nombre réel strictement positif donné, on a, formellement :

$$\lambda \sum_{t > \lambda} P(t \leq X \leq t + dt) \leq \sum_{t > \lambda} t P(t \leq X \leq t + dt) \leq \sum_t t P(t \leq X \leq t + dt) \simeq E[X];$$

mais le membre de gauche de cette inégalité est une manière approchée d'écrire $\lambda P(X > \lambda)$, d'où il résulte l'importante *inégalité de Markov* :

$$P(X > \lambda) \leq \frac{E[X]}{\lambda}, \quad \forall \lambda > 0,$$

valable pour toute variable aléatoire réelle positive sur un espace probabilisé.

Si X est une variable aléatoire réelle admettant une variance, on peut utiliser cette inégalité (en l'appliquant à la variable aléatoire réelle positive $|X - E[X]|^2$) pour contrôler quantitativement comment X diffère statistiquement de sa moyenne ; on a en effet

$$P(|X - E[X]|^2 > \lambda^2) = P(|X - E[X]| > \lambda) \leq \frac{E[|X - E[X]|^2]}{\lambda^2} = \frac{V(X)}{\lambda^2}.$$

C'est la célèbre *inégalité de Bienaymé-Tchebychev*, ingrédient essentiel de l'argumentation probabiliste. Cette inégalité est intéressante pour les valeurs de λ inférieures à l'écart type de X .

2.8 Indépendance de variables aléatoires réelles

2.8.1 Indépendance de deux variables aléatoires réelles

Deux variables aléatoires réelles X et Y sur un même espace probabilisé (Ω, \mathcal{T}, P) sont dites *indépendantes* si et seulement si, pour toute paire (A, B) de sous-ensembles quarrables de \mathbb{R} , on a

$$P(\{\omega \in \Omega; x \in A, y \in B\}) = P_X(A) \times P_Y(B).$$

Si X et Y sont indépendantes et ont toutes deux une variance, on a donc formellement

$$\begin{aligned} E[XY] &= \sum_t \sum_s ts P(X \in [t, t + dt], Y \in [s, s + ds]) \\ &= \sum_t \sum_s ts P(X \in [t, t + dt]) P(Y \in [s, s + ds]) \\ &= \left(\sum_t t P(X \in [t, t + dt]) \right) \times \left(\sum_s s P(Y \in [s, s + ds]) \right) \\ &= E[X] \times E[Y]. \end{aligned}$$

Si deux variables aléatoires réelles ayant toutes les deux une variance sont indépendantes, on a

$$\text{covar}(X, Y) = E[XY] - E[X]E[Y] = 0.$$

La réciproque de cette proposition est fautive : deux variables aléatoires réelles peuvent avoir une covariance nulle sans être indépendantes.

Si (X_1, \dots, X_n) sont n variables aléatoires réelles deux à deux indépendantes, la matrice de covariance $\text{covar}(X_1, \dots, X_n)$ est une matrice diagonale ; comme dans le cas de deux variables, la réciproque est fautive.

Exemple 2.13 Si X et Y sont deux variables indépendantes, à valeurs dans \mathbb{N} et suivant toutes deux une loi de Poisson (de paramètres respectifs λ_X, λ_Y), on a, pour tout $n \in \mathbb{N}$,

$$P(\{X + Y = n\}) = \sum_{k=0}^n P(\{X = k\})P(\{Y = n - k\}) = e^{-\lambda_X - \lambda_Y} \frac{(\lambda_X + \lambda_Y)^n}{n!},$$

ce qui montre que $X + Y$ suit aussi une loi de Poisson de paramètre $\lambda_X + \lambda_Y$. A propos de cet exemple, examinons un phénomène physique où intervient la loi de Poisson : il s'agit de l'émission de particules. Etant donnés deux instants $t < s$, notons $A_k(t, s)$ l'évènement "*k* particules exactement sont émises pendant le laps de temps $]t, s[$ ". L'expérience est censée obéir aux règles suivantes :

- (a) si les intervalles $]t_j, s_j]$ sont disjoints, les évènements correspondants $A_{k_j}(t_j, s_j)$ sont indépendants ;
- (b) pour tout k , il existe une fonction ϕ_k telle que $P(A_k(t, s)) = \phi_k(s - t)$ (c'est ce qu'on appelle une condition de stationnarité sur le processus) ;
- (c) si l'on note $B(t) := \bigcup_{k=2}^{\infty} P(A_k(0, t))$, on a

$$\lim_{\substack{t \rightarrow 0 \\ t > 0}} \frac{P(B(t))}{t} = 0.$$

Si l'on suppose $\alpha = \phi_0(1) > 0$ (ce que nous ferons), il est facile de voir, compte tenu des hypothèses (a) et (b), que $\phi_0(n) = \alpha^n$ pour tout entier positif n ; pour tout rationnel positif ξ , on en déduit $\phi_0(\xi) = \alpha^\xi$. La continuité à gauche de ϕ_0 (propriété des mesures de probabilité) permet d'étendre ceci à tout ξ réel positif. Mais

$$\phi_1(t) = 1 - \alpha^t + o(t) = -\log(\alpha)t + o(t)$$

du fait de l'hypothèse (c). Mais, grâce à l'indépendance, on a, pour tout $k \in \mathbf{N}$, pour tout $t, s > 0$,

$$\phi_k(t + s) = \sum_{j=0}^k \phi_j(t)\phi_{k-j}(s),$$

ce que l'on peut aussi écrire, toujours grâce à (c),

$$\phi_k(t + s) = \phi_k(t)\alpha^s + \phi_{k-1}(t)((-\log(\alpha)s + o(s)) + o_1(s)) ;$$

on voit ainsi que ϕ_k est dérivable à droite sur $[0, \infty[$; la fonction étant croissante, elle est dérivable sur $]0, \infty[$ et solution de l'équation différentielle

$$\phi_k'(t) = -\log(\alpha)(\phi_{k-1}(t) - \phi_k(t)).$$

Par une induction immédiate (sur k), on voit que, si $\lambda = -\log(\alpha) > 0$, on a

$$\phi_k(t) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t)$$

et l'on voit apparaître la loi de Poisson de paramètre λ .

Si X et Y sont deux variables aléatoires réelles indépendantes à densités (de densités respectives f_X et f_Y , le vecteur (X, Y) est un vecteur de variables aléatoires réelle à densité, de densité

$$f_{(X,Y)}(x, y) = f_X(x) \times f_Y(y).$$

Ceci résulte immédiatement de la définition de l'indépendance.

Si X_1, \dots, X_n sont n variables indépendantes deux à deux, on a, du fait des formules intervenant dans le calcul de variance,

$$V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n).$$

2.8.2 Régression linéaire

Etant données deux variables aléatoires réelles X et Y , toutes deux définies sur un même espace probabilisé (Ω, \mathcal{T}, P) , ayant toutes les deux une variance, on appelle *droite de régression de Y relativement à X* la droite d'équation $y = \alpha x + \beta$, où la variable $Z = \alpha X + \beta$ réalise la meilleure approximation de Y au sens des moindres carrés pour la fonctionnelle correspondant à la corrélation, c'est-à-dire est la combinaison linéaire des variables 1 et X telle que

$$E[|Y - \alpha X - \beta|^2]$$

soit minimale. Chercher la droite de régression de Y relativement à X , c'est tenter de déterminer les meilleurs paramètres α, β qui donnent la fonction affine en X dont la distribution statistique des valeurs approche *au mieux, et en tout cas en un sens stochastique*, la distribution statistique des valeurs de Y . Les conditions vérifiées par α et β doivent être :

$$E[Y - \alpha X - \beta] = E[(Y - \alpha X - \beta)X] = 0$$

Si $V(X) V(Y) \neq 0$ et si $\sigma(X), \sigma(Y)$ désignent les écarts-types de X et Y , l'équation de la droite de régression s'écrit

$$\frac{y - E[Y]}{\sigma(Y)} = \left(\frac{\text{covar}(X, Y)}{\sigma(X)\sigma(Y)} \right) \frac{x - E[X]}{\sigma(X)}.$$

Le coefficient

$$\rho(x, y) := \frac{\text{covar}(X, Y)}{\sigma(X)\sigma(Y)}$$

est dit *coefficient de corrélation de X et Y* ; ce coefficient peut être nul quand bien même X et Y sont indépendantes !

2.8.3 Indépendance mutuelle d'une famille de variables indépendantes

Si X est une variable aléatoire réelle sur un espace probabilisé (Ω, \mathcal{T}, P) , la tribu $\mathcal{T}(X)$ est par définition la plus petite tribu contenant tous les ensembles $X^{-1}(A)$, où A est un sous-ensemble quarrable quelconque de \mathbf{R} .

Une collection finie ou dénombrable $(X_i)_{i \in I}$ de variables aléatoires réelles toutes définies sur le même espace probabilisé (Ω, \mathcal{T}, P) est dite *collection de variables mutuellement indépendantes* si les tribus $\mathcal{T}(X_i)$, $i \in I$, forment une collection de tribus mutuellement indépendantes. La même définition vaut si les X_i sont des vecteurs de variables aléatoires réelles ; on dit que la collection des vecteurs de variables aléatoires réelles $(X_i)_i$ est une collection de vecteurs de variables aléatoires réelles mutuellement indépendants si les tribus $(\mathcal{T}(X_i))_i$ forment une collection de sous-tribus de \mathcal{T} mutuellement indépendantes.

2.9 Les théorèmes limite de la théorie des probabilités

2.9.1 La notion de convergence en probabilité et la loi faible des grands nombres

Soient X_1, \dots, X_n, \dots une suite de variables indépendantes deux à deux, toutes définies sur un même espace probabilisé (Ω, \mathcal{T}, P) et ayant même loi, de moyenne m et de variance finie. Alors la suite de variables

$$Z_n := \frac{\sum_{k=1}^n X_k}{n} - m$$

satisfait la propriété suivante :

$$\forall \lambda > 0, \quad \lim_{n \rightarrow \infty} P(|Z_n| \geq \lambda) = 0.$$

On dit que la suite $(Z_n)_n$ converge en probabilité vers 0. Plus généralement, une suite de variables aléatoires réelles $(Z_n)_n$, toutes définies sur le même espace probabilisé (Ω, \mathcal{T}, P) converge en probabilité (ou encore converge stochastiquement) vers une variable aléatoire réelle Z si et seulement si

$$\forall \lambda > 0, \quad \lim_{n \rightarrow \infty} P(|Z_n - Z| \geq \lambda) = 0.$$

Ce résultat, annonçant la *loi forte des grands nombres* que nous énoncerons ultérieurement, est connu comme la *loi faible des grands nombres*. Cette loi faible résulte de l'inégalité de Bienaymé-Tchebychev. En effet, si

$$Y_n = \frac{\sum_{j=1}^n X_j}{n}$$

on a $V(Y_n) = (1/n^2)(V(X_1) + \dots + V(X_n)) = V(X_1)/n$ (puisque les X_j sont indépendantes deux à deux et que la prise de variance est une opération quadratique) et $E[Y_n] = m$ (d'après la linéarité de la prise d'espérance) ; on a donc

$$P(|Y_n - m| > \lambda) = P(|Z_n| > \lambda) \leq \frac{V(Y_n)}{\lambda^2} = \frac{V(X_1)}{n\lambda^2},$$

ce qui montre que cette quantité tend bien vers 0 lorsque n tend vers l'infini.

2.9.2 La notion de convergence en loi et le théorème de la limite centrale

Une suite $(X_n)_n$ de variables aléatoires réelles est dite converger en loi vers une variable aléatoire X si et seulement si

$$F_{X_n}(t) \rightarrow F_X(t),$$

si F_{X_n} (resp. F_X) désigne la fonction de répartition de X_n (resp. de X), ce en tout point t de continuité de F_X .

Si $(X_n)_n$ est une suite de variables aléatoires réelles convergeant en probabilité vers une variable X (i.e. $\lim_{n \rightarrow \infty} P(|X_n - X| > \lambda) = 0$ pour tout $\lambda > 0$), alors la suite $(X_n)_n$ converge en loi vers X . On admettra ce résultat.

Une condition suffisante pour qu'une suite $(X_n)_n$ de variables aléatoires réelles converge en loi est qu'il existe une fonction Φ continue en $\omega = 0$, telle que, pour tout nombre réel ω ,

$$\lim_{n \rightarrow \infty} E[e^{i\omega X_n}] = \Phi(\omega);$$

la fonction Φ s'écrit alors

$$\Phi(\omega) = E[e^{i\omega X}],$$

où X est une variable aléatoire réelle X ; cette variable X est alors la limite en loi de la suite $(X_n)_n$. Ce résultat important est le théorème de Paul Lévy (Paul Lévy

fut l'un des initiateurs de la théorie des probabilités dans la première moitié du XX^e siècle). La notion sous-jacente de *fonction caractéristique* (on appelle *fonction caractéristique* d'une variable aléatoire réelle X la fonction $\omega \in \mathbb{R} \rightarrow E[e^{i\omega X}]$) fait entrer la transformation de Fourier (ou, pour les physiciens, la *prise de spectre*) au service de la théorie des probabilités. On notera aussi, ce qui est commode, que la fonction caractéristique d'une somme de variables aléatoires indépendantes est le produit des fonctions caractéristiques; ceci est un avatar de l'intérêt calculatoire de la transformation de Fourier qui transforme l'opération de convolution (passage à travers un filtre linéaire dont les paramètres restent immuables dans le temps, par exemple une cellule mécanique ou une cellule électrique) en l'opération autrement plus maniable de multiplication.

Voici une illustration de la convergence en loi, avec le *théorème de la limite centrale* qui précise la convergence des lois binomiales vers la loi de Gauss et éclaire l'expérience du triangle de Galton évoquée plus haut dans ce cours. Le résultat s'énonce ainsi : soit $(X_n)_n$ une suite de variables aléatoires réelles, que l'on suppose mutuellement indépendantes, toutes définies sur un même espace probabilisé (Ω, \mathcal{T}, P) ; on suppose que ces variables ont toutes même loi, avec pour espérance m et pour variance σ^2 ; Il existe alors une variable aléatoire X , elle aussi définie sur (Ω, \mathcal{T}, P) , suivant une loi de Gauss de paramètres $(0,1)$, telle que la suite de variables

$$Y_n := \frac{\sum_{k=1}^n X_k - nm}{\sqrt{n}\sigma}$$

converge en loi vers X . C'est ce théorème qui nous permet, quand $n\tau(1-\tau) > 10$, d'approcher raisonnablement bien la loi binomiale de paramètres (n, τ) par une loi gaussienne de paramètres $\mu = n\tau$, $\sigma = \sqrt{n\tau(1-\tau)}$, ce qui est une approximation très couramment utilisée.

Voici (proposée ici à titre d'exercice) une preuve rapide du théorème limite centrale : on commence par remarquer que si Y est une variable aléatoire centrée et de variance 1, on a, pour t fixé, en utilisant par exemple la formule de Taylor avec reste intégral à l'ordre 2 pour développer l'exponentielle,

$$E[\exp(itY/\sqrt{n})] = 1 - \frac{t^2}{2n} + o_t(1/n);$$

en utilisant l'indépendance mutuelle des variables X_j , on a, pour tout n , pour tout t réel,

$$E\left[\exp\left(it\frac{\sum_{k=1}^n X_k - nm}{\sqrt{n}\sigma}\right)\right] = \left(E\left[\exp\left(it\frac{X_1 - m}{\sqrt{n}\sigma}\right)\right]\right)^n;$$

on a donc, pour tout t réel,

$$E[\exp(itY_n)] = \left(1 - \frac{t^2}{2n} + o_t(1/n)\right)^n \rightarrow \exp(-t^2/2), \quad n \rightarrow \infty;$$

comme la fonction $t \mapsto \exp(-t^2/2)$ est la fonction caractéristique d'une variable aléatoire suivant une loi gaussienne réduite centrée, la conclusion du théorème limite centrale résulte du théorème de Paul Lévy.

Dans le cadre des variables aléatoires discrètes, le concept de *fonction génératrice* remplace souvent avantageusement celui de fonction caractéristique.

Soit Y une variable aléatoire à valeurs dans \mathbb{N} définie sur un espace probabilisé (Ω, \mathcal{T}, P) ; la fonction

$$z \in \bar{U} = \{z \in \mathbb{C}; |z| \leq 1\} \mapsto G_Y(z) = E[z^Y] = \sum_{n=0}^{\infty} P(\{Y = n\})z^n$$

est une fonction continue sur \bar{U} , de classe \mathcal{C}^k sur \bar{U} dès que $E[|Y|^k] < \infty$, que l'on appelle *fonction génératrice* de Y ; de plus, si $E[|Y|^k] < \infty$, pour tout $1 \leq l \leq k$,

$$\left(\frac{d}{dz}\right)^l [G(z)] = \sum_{k=0}^{\infty} (k+1) \cdots (k+l-1) P(X=k); \quad (\dagger)$$

Pour des variables discrètes, on a la version suivante du théorème de Paul Lévy : une suite de variables aléatoires $(X_n)_n$ à valeurs dans \mathbf{N} , toutes définies sur (Ω, \mathcal{T}, P) , converge en loi vers une variable aléatoire discrète X (aussi à valeurs dans \mathbf{N}) si et seulement si il existe une fonction G , définie sur \bar{U} et continue en 1, telle que

$$\lim_{n \rightarrow \infty} G_{X_n}(z) = \Phi(z), \quad z \in \bar{U}.$$

Le concept de fonction génératrice s'avère un outil commode pour calculer espérance, variance, *etc.*, d'une variable aléatoire à valeurs dans \mathbf{N} : par exemple, en évaluant le second membre de (\dagger) en $z = 1$, on trouve espérance, variance, *etc.* On notera aussi que (tout au moins formellement), comme c'est le cas pour les fonctions caractéristiques, la fonction génératrice d'une somme de variables aléatoires à valeurs dans \mathbf{N} indépendantes est le produit des fonctions génératrices.

2.10 La convergence presque sûre et la loi forte des grands nombres

Une suite de variables aléatoires réelles $(X_n)_n$, toutes définies sur le même espace probabilisé (Ω, \mathcal{T}, P) converge P -presque sûrement vers une variable aléatoire réelle X si l'ensemble A des $\omega \in \Omega$ où $X_n(\omega)$ ne converge pas vers $X(\omega)$ est tel que $P(A) = 0$.

La convergence presque sûre d'une suite de variables aléatoires $(X_n)_n$ vers une variable aléatoire X implique la convergence en probabilité de la suite $(X_n)_n$ vers cette même variable, donc *a fortiori* la convergence en loi de $(X_n)_n$ vers X . La convergence presque sûre est donc la plus forte des trois notions de convergence que nous avons introduit dans ce cours.

Avant de donner un exemple de convergence presque sûre (la loi forte des grands nombres), nous avons l'amélioration suivante de l'inégalité de Bienaymé-Tchebychev, due à Kolmogorov, dite *inégalité de Kolmogorov* : soient X_1, \dots, X_n n variables mutuellement indépendantes, toutes définies sur le même espace probabilisé (Ω, \mathcal{T}, P) , admettant chacune une variance et toutes de moyenne nulle; alors, pour tout $\epsilon > 0$,

$$P \left[\left\{ \max_{1 \leq j \leq n} \left| \sum_{l=1}^j X_l \right| \geq \epsilon \right\} \right] \leq \frac{\sum_{k=1}^n V(X_k)}{\epsilon^2}.$$

La preuve de l'inégalité de Kolmogorov peut être considérée comme un intéressant exercice sur le conditionnement; nous la mentionnons brièvement ici. On introduit les variables

$$Y_j := \sum_{l=1}^j X_l, \quad j = 1, \dots, n;$$

on introduit aussi le système complet d'événements suivants $(A_q(\epsilon)), q = 0, \dots, n$, où

$$A_0(\epsilon) = A_0 = \{ |Y_1| < \epsilon, \dots, |Y_n| < \epsilon \}$$

et, pour q entre 1 et n ,

$$A_q(\epsilon) = A_q = \{ |Y_l| < \epsilon, l < q; |Y_q| \geq \epsilon \}.$$

Utilisant ce système complet et l'indépendance des X_j , on a

$$E[Y_n^2] = \sum_{k=1}^n V(X_k) = \sum_{q=0}^n E[Y_n^2 \chi_{A_q}] \geq \sum_{q=1}^n E[Y_n^2 \chi_{A_q}]. \quad (*)$$

Soit q entre 1 et n ; si nous écrivons Y_n sous la forme

$$Y_n = Y_q + X_{q+1} + \dots + X_n,$$

nous voyons en utilisant l'indépendance mutuelle des X_j (l'indépendance deux à deux ne suffit pas ici) que

$$E[Y_n^2 \chi_{A_q}] = E[Y_q^2 \chi_{A_q}] + \sum_{l=q+1}^n E[X_l^2 \chi_{A_q}] \quad (**)$$

où χ_{A_q} désigne la fonction indicatrice de l'événement A_q , valant 1 si A_q est réalisé, 0 sinon. Mais, comme $E[Y_q^2 \chi_{A_q}] \geq \epsilon^2 P(A_q)$ (voir la définition de $A_q(\epsilon)$), on déduit de (**) que

$$E[Y_n^2 \chi_{A_q}] \geq \epsilon^2 P(A_q);$$

en ajoutant les inégalités ainsi obtenues, on déduit de (*)

$$\sum_{q=1}^n P(A_q(\epsilon)) \leq \frac{\sum_{k=1}^n V(X_k)}{\epsilon^2}$$

ce qui est la conclusion voulue si l'on se reporte à la définition des $A_q(\epsilon)$.

La *loi forte des grands nombres* est une conséquence de l'inégalité de Kolmogorov (comme la loi faible résultait de l'inégalité de Bienaymé-Tchebychev). La loi forte des grands nombres réalise l'articulation fondamentale entre la théorie des probabilités et le raisonnement fondant les statistiques. Voici, pour conclure cette brève initiation aux probabilités, l'énoncé de cette loi capitale :

Soit $(X_n)_n$ une suite de variables aléatoires réelles, mutuellement indépendantes, toutes définies sur un même espace probabilisé (Ω, \mathcal{T}, P) (conditionnant l'épreuve). On suppose que ces variables ont toutes même loi, avec pour espérance m et pour variance σ^2 (il faut penser X_k comme le résultat d'une épreuve, épreuve que l'on répète indéfiniment de manière indépendante). Alors la suite de variables

$$Z_n := \frac{\sum_{k=1}^n X_k}{n}$$

converge P -presque sûrement vers la variable P -presque partout égale à m , qui est la valeur moyenne du résultat de l'épreuve. Ainsi, si l'on jette un dé indéfiniment et si l'on divise le nombre de points obtenus par le nombre de coups, on approche presque sûrement lorsque le nombre de coups tend vers l'infini, la constante

$$\frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} = 3.5$$

(qui correspond à la valeur moyenne de l'épreuve).

Voici, à titre d'exercice, le chemin logique menant de l'inégalité de Kolmogorov à la formulation de la loi forte des grands nombres. Il n'y a pas de restriction à supposer les variables X_j toutes

centrées (sinon, on leur retranche leur espérance). On reprend les variables Y_l introduites lors de la preuve de l'inégalité de Kolmogorov. Posons, pour tout $\epsilon > 0$, pour tout $n \in \mathbf{N}$,

$$B_n(\epsilon) = \left\{ \sup_{k \geq n} \left| \frac{Y_k}{k} \right| \geq \epsilon \right\}.$$

On a

$$\begin{aligned} P(B_{2^j}(\epsilon)) &\leq \sum_{l=j}^{\infty} P \left[\sup_{2^l \leq k < 2^{l+1}} \left(\left| \frac{Y_k}{k} \right| > \epsilon \right) \right] \\ &\leq \sum_{l=j}^{\infty} P \left[\sup_{1 \leq k < 2^{l+1}} (|Y_k|) \geq \epsilon 2^l \right] \\ &\leq \sum_{l=j}^{\infty} \left[\frac{1}{\epsilon^2 2^{2l}} (2^{l+1} \sigma^2) \right]. \end{aligned}$$

On utilise ici (pour la dernière inégalité) l'inégalité de Kolmogorov. On voit ainsi que pour tout $\epsilon > 0$, on a

$$\lim_{j \rightarrow +\infty} P(B_{2^j}(\epsilon)) = 0.$$

Ceci étant acquis pour tout $\epsilon > 0$, on en déduit que la suite Z_n converge vers 0 P -presque partout, ce qui achève la preuve de la loi forte des grands nombres.