

Analyse fonctionnelle

Partie 2

par **Gilles GODEFROY**
Directeur de recherches au Centre national de la recherche scientifique

1. Espaces fonctionnels non normables	AF 101 - 2
1.1 Lemme de Wielandt	— 2
1.2 Distributions.....	— 2
2. Transformation de Fourier.....	— 3
3. Calcul des probabilités.....	— 5
3.1 Espaces de probabilité et variables aléatoires.....	— 5
3.2 Indépendance.....	— 7
3.3 Lois limites	— 7

Les opérateurs de dérivation ne se représentent pas de façon naturelle comme opérateurs continus sur des espaces normés. Le bon cadre pour le calcul différentiel est fourni par la **théorie des distributions**, qui impose l'utilisation d'espaces non normables mais permet de donner un sens à la « dérivée » de fonctions très générales.

La **transformation de Fourier** déploie toute sa puissance dans ce cadre élargi et permet de résoudre effectivement de nombreuses équations aux dérivées partielles, en donnant l'existence et la forme générale des solutions.

C'est encore l'analyse de Fourier qui procure le bon outil pour établir les **théorèmes limites du calcul des probabilités**, et faire apparaître le rôle central des variables gaussiennes aux interfaces entre le calcul sur les sphères de grande dimension, la distribution des grandeurs physiques ou biologiques et l'incertitude des mesures.

Pour aborder sans difficultés cette deuxième partie de l'analyse fonctionnelle, le lecteur consultera, dans ce traité :

- [AF 99] - *Topologie et mesure* ;
- [AF 100] - *Analyse fonctionnelle. Partie 1.*

1. Espaces fonctionnels non normables

1.1 Lemme de Wielandt

Les espaces de Banach constituent un cadre plus général que les espaces de Hilbert, cadre qui contient une grande partie de l'analyse fonctionnelle, puisque comme nous l'avons vu, de nombreux espaces fonctionnels sont munis d'une norme complète naturelle (cf. article [AF 100]).

Pourtant, les espaces de fonctions \mathcal{C}^∞ ne peuvent pas être munis d'une structure naturelle d'espace de Banach. Cela provient d'une équation très simple : si f est une fonction dérivable, on a :

$$(xf)' - xf' = f$$

Donc si $E = \mathcal{C}^\infty(\mathbb{R})$ est l'espace des fonctions indéfiniment dérivables de \mathbb{R} dans \mathbb{R} , si $M_x : E \rightarrow E$ est la multiplication par x et $D : E \rightarrow E$ est la dérivation, on a :

$$D M_x - M_x D = I_E \quad (1)$$

Énonçons maintenant le **lemme de Wielandt** :

Lemme 1

Soit X un espace de Banach non réduit à $\{0\}$, et soit $L(X)$ l'espace des opérateurs linéaires continus sur X .

L'équation $AB - BA = I_X$ n'a pas de solution dans $L(X)$.

Preuve ♦ Supposons que $A \in L(X)$ et $B \in L(X)$ satisfont $AB - BA = I_X$. On définit $D_A : L(X) \rightarrow L(X)$ par $D_A(T) = AT - TA$.

L'application linéaire D_A est une « dérivation », au sens suivant :

$$\begin{aligned} D_A(UV) &= AUV - UVA = AUV - UAV + UAV - UVA \\ &= (AU - UA)V + U(AV - VA) = D_A(U)V + UD_A(V) \end{aligned}$$

Puisque B satisfait $D_A(B) = I_X$, on a :

$$D_A(B^n) = D_A(B^{n-1}B) + B^{n-1}D_A(B) = D_A(B^{n-1})B + B^{n-1}$$

d'où on déduit, par récurrence, que :

$$D_A(B^n) = AB^n - B^nA = nB^{n-1} \quad (2)$$

En prenant les normes dans (2), on voit que, pour tout n :

$$2\|A\| \cdot \|B\| \cdot \|B^{n-1}\| \geq n\|B^{n-1}\|$$

ce qui implique que $B^n = 0$ pour n assez grand ; mais (2) montre $B = 0$, d'où $X = \{0\}$. ♦

D'après (1), on déduit immédiatement du lemme 1 :

Proposition 1

Il n'existe pas de norme sur l'espace $\mathcal{C}^\infty(\mathbb{R})$ telle que les applications M_x et D soient toutes deux continues.

Toutes les applications linéaires « raisonnables » sur un espace de Banach sont continues. Il s'ensuit donc de la proposition 1 qu'aucune norme utilisable sur $\mathcal{C}^\infty(\mathbb{R})$ n'est complète.

Pour manipuler l'opérateur de dérivation, nous pouvons le considérer : soit comme un opérateur non continu sur un espace normé (non complet), soit comme un opérateur continu sur un espace métrisable complet non normable. La première option conduit à la théorie des « **générateurs des semi-groupes** », mais nous allons nous concentrer sur la deuxième option, qui va nous conduire à la **théorie des distributions**.

1.2 Distributions

La théorie des distributions a été créée par L. Schwartz en 1948, pour justifier certains procédés de calcul utilisés par les physiciens, et permettre en particulier de donner un sens précis à la « dérivée » d'une fonction non dérivable au sens usuel. L'une des idées de base consiste à voir une fonction intégrable $f \in L^1(\mathbb{R})$ comme une application linéaire Λ définie sur l'espace $\mathcal{D}(\mathbb{R})$ des fonctions \mathcal{C}^∞ à support compact par :

$$\Lambda(\phi) = \int_{-\infty}^{+\infty} \phi(t) f(t) dt \quad (3)$$

Si la fonction f est dérivable, une intégration par parties montre que :

$$\int_{-\infty}^{+\infty} f'(t) \phi(t) dt = - \int_{-\infty}^{+\infty} f(t) \phi'(t) dt \quad (4)$$

Nous remarquons maintenant que le membre de droite de (4) a un sens pour **toute** fonction $f \in L^1(\mathbb{R})$. Il est donc naturel de poser :

$$\Lambda'(\phi) = - \int_{-\infty}^{+\infty} f(t) \phi'(t) dt \quad (5)$$

et de considérer Λ' comme la **dérivée** de l'application linéaire Λ . Plus généralement encore, si F est une forme linéaire sur $\mathcal{D}(\mathbb{R})$, on définira sa dérivée F' par $F'(\phi) = -F(\phi')$.

Cette idée permet d'étendre le calcul différentiel classique à des fonctions très générales, en acceptant toutefois que la dérivée d'une fonction (non dérivable, ni même continue !) soit non pas une fonction mais une **distribution**. Remarquons que (3) constitue un changement de point de vue sur la fonction f : on s'intéresse moins à ses valeurs en tout point qu'à ses moyennes pondérées par des fonctions lisses ϕ . Ce nouveau point de vue est mieux adapté à la physique, où les données sont le plus souvent mesurées par des moyennes.

Nous avons maintenant besoin d'un cadre topologique précis. Dans un but de simplicité, nous nous limiterons à présenter les distributions tempérées sur l'espace \mathbb{R}^n .

Soit $f \in \mathcal{C}^\infty(\mathbb{R}^n)$. Si $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, on pose $|\alpha| = |\alpha_1| + |\alpha_2| + \dots + |\alpha_n|$ et :

$$D^\alpha f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \dots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} (f) \quad (6)$$

Définition 1

L'espace de Schwartz \mathcal{S}_n est l'espace des fonctions $f \in \mathcal{C}^\infty(\mathbb{R}^n)$ telles que pour tout $N \in \mathbb{N}$:

$$\|f\|_N = \sup_{|\alpha| \leq N} \sup_{x \in \mathbb{R}^n} (1 + |x|^2)^N |D^\alpha f(x)| < \infty$$

Il est clair que toute fonction $f \in \mathcal{C}^\infty(\mathbb{R}^n)$ à support compact appartient à \mathcal{S}_n . La famille des normes $\|\cdot\|_N$ ($N \in \mathbb{N}$) définit une distance, par exemple par la formule :

$$d(f, g) = \sum_{N=0}^{\infty} 2^{-N} \inf(1, \|f - g\|_N)$$

Si τ est la topologie correspondante, on a :

$$\tau\text{-}\lim_{k \rightarrow +\infty} f_k = f \Leftrightarrow \|f_k - f\|_N \rightarrow 0 \text{ pour tout } N$$

L'espace \mathcal{S}_n est métrique complet pour cette structure ; il s'ensuit de (1) et du lemme 1 que \mathcal{S}_n n'est **pas** un espace de Banach.

Nous pouvons maintenant définir certaines distributions.

Définition 2

Une distribution tempérée sur \mathbb{R}^n est une forme linéaire continue sur l'espace \mathcal{S}_n

Si g est une fonction mesurable telle qu'il existe un entier $k \geq 1$ tel que :

$$\int_{\mathbb{R}^n} (1 + |x|^2)^{-k} |g(x)| dm_n(x) < \infty \quad (7)$$

alors g s'identifie à une distribution tempérée, par :

$$f \in \mathcal{S}_n \mapsto \int_{\mathbb{R}^n} g(x) f(x) dm_n(x) \quad (8)$$

En particulier, toute fonction intégrable sur \mathbb{R}^n est une distribution tempérée.

Si $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{N}^n$ est un multi-indice et que $u \in \mathcal{S}_n^*$ est une distribution tempérée, la formule :

$$(D^\alpha u)(f) = (-1)^{|\alpha|} u(D^\alpha f) \quad (9)$$

où $f \in \mathcal{S}_n$ et $D^\alpha f$ est donnée par (6), définit une distribution tempérée $(D^\alpha u)$.

L'opérateur $D^\alpha : \mathcal{S}_n^* \rightarrow \mathcal{S}_n^*$ vérifie les règles usuelles (telles que la formule de Leibniz) des opérateurs de dérivation. De plus, il prolonge en effet le calcul différentiel usuel : si une fonction f est partout dérivable et que sa dérivée f' est intégrable (et définit donc une distribution), cette dérivée au sens classique coïncide avec la dérivée de f au sens des distributions.

Exemple 1 : soit $h : \mathbb{R} \rightarrow \mathbb{R}$ la « fonction de Heaviside », définie par $h(x) = 1$ si $x \geq 0$ et $h(x) = 0$ si $x < 0$. C'est une distribution tempérée sur \mathbb{R} . Pour toute $f \in \mathcal{S}_1$, on a :

$$(Dh)(f) = - \int_0^{+\infty} f'(t) dt = f(0)$$

et, par conséquent, $Dh = \delta_{\{0\}}$, la mesure de Dirac en 0.

Dans le langage des physiciens, nous avons montré que la dérivée de la « fonction de Heaviside » h est l'impulsion unité en 0.

D'après (9), on a pour tout $k \geq 1$:

$$(D^k \delta_{\{0\}})(f) = f^{(k)}(0)$$

Nous voyons, sur ce simple exemple, que les distributions permettent de représenter les opérateurs différentiels, ainsi que les opérateurs intégraux [cf. relation (8)].

Une classe importante d'opérateurs est constituée par les opérateurs qui commutent avec les translations : si f est une fonction de $L^1(\mathbb{R}^n)$ et $x \in \mathbb{R}^n$, on définit :

$$(\tau_x f)(y) = f(y - x) \quad \text{et} \quad \check{f}(y) = f(-y)$$

Si $g \in L^1(\mathbb{R}^n)$ et que :

$$(g * f)(x) = C_g(f)(x) = \int_{\mathbb{R}^n} g(y) f(x - y) dm_n(y) \quad (10)$$

On a, par un changement de variable :

$$C_g(\tau_x f) = \tau_x[C_g(f)] \quad (11)$$

Pour étendre (10) et définir la convolution d'une distribution Λ et d'une fonction f , il est naturel de poser :

$$(\Lambda * f)(x) = \Lambda[\tau_x \check{f}] \quad (12)$$

L'équation (12) définit un opérateur $C_\Lambda(f) = \Lambda * f$ de \mathcal{S}_n dans $\mathcal{C}^\infty(\mathbb{R}^n)$ qui vérifie (11). Inversement, les opérateurs qui vérifient (11) sont de la forme C_Λ . Par exemple, si δ_0 est la mesure de Dirac en 0 $\in \mathbb{R}^n$ et $\alpha \in \mathbb{N}^n$ est un multi-indice, on a pour toute fonction $f \in \mathcal{S}_n$:

$$D^\alpha(f) = (D^\alpha \delta_0) * f$$

Les opérateurs de convolution, en synergie avec la transformation de Fourier, jouent un rôle fondamental dans la résolution des équations aux dérivées partielles, comme nous allons le voir au paragraphe 2.

2. Transformation de Fourier

Nous avons déjà rencontré (cf. article [AF 100, § 2]) certains aspects de l'analyse de Fourier. Nous allons aborder dans ce paragraphe quelques applications des méthodes de Fourier à la résolution d'équations aux dérivées partielles. Rappelons d'ailleurs que J.B. Fourier a été amené à considérer les séries qui portent son nom en cherchant (dans son mémoire de 1807) à résoudre l'« équation de la chaleur ».

Nous aurons besoin des notations suivantes.

Notations

■ La mesure de Lebesgue normalisée m_n^* sur \mathbb{R}^n est définie par :

$$dm_n^*(x) = (2\pi)^{-n/2} dm_n(x)$$

■ Pour tout $t = (t_1, t_2, \dots, t_n) \in \mathbb{R}^n$, la fonction $e_t : \mathbb{R}^n \rightarrow \mathbb{C}$ est définie par :

$$e_t(x) = e^{it \cdot x} = \exp\left(i \cdot \sum_{k=1}^n t_k x_k\right)$$

et pour tout $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{N}^n$, on pose :

$$D_\alpha = (i)^{-|\alpha|} D^\alpha = \left(\frac{1}{i} \frac{\partial}{\partial x_1}\right)^{\alpha_1} \left(\frac{1}{i} \frac{\partial}{\partial x_2}\right)^{\alpha_2} \dots \left(\frac{1}{i} \frac{\partial}{\partial x_n}\right)^{\alpha_n}$$

■ Si $P(\xi) = \sum c_\alpha \xi^\alpha = \sum c_\alpha \xi_1^{\alpha_1} \xi_2^{\alpha_2} \dots \xi_n^{\alpha_n}$ est un polynôme à n variables à coefficients complexes, on pose :

$$P(D) = \sum c_\alpha D_\alpha$$

et donc :

$$P(-D) = \sum (-1)^{|\alpha|} c_\alpha D_\alpha$$

■ Les opérateurs de translation sont toujours définis par $(\tau_x f)(y) = f(y - x)$.

Pour des raisons de normalisation, la définition ci-dessous diffère légèrement des notations précédentes (cf. article [AF 100, relation (47)]).

Définition 3

Soit f une fonction intégrable sur \mathbb{R}^n . La transformée de Fourier \hat{f} de f est définie pour tout $t \in \mathbb{R}^n$ par :

$$\hat{f}(t) = \int_{\mathbb{R}^n} f e_{-t} dm_n^* = \int_{\mathbb{R}^n} f(x) e^{-it \cdot x} dm_n^*(x)$$

Le résultat suivant illustre l'utilité de cette notion dans la résolution des équations aux dérivées partielles.

Théorème 1

Pour toute fonction $f \in \mathcal{S}_n$ et tout polynôme P à n variables, on a les équations :

$$(P(D) f)^\wedge = P \cdot \hat{f} \quad (13)$$

$$(P \cdot f)^\wedge = P(-D) \hat{f} \quad (14)$$

Preuve ♦ En dérivant l'exponentielle, on vérifie immédiatement que :

$$P(D) e_t = P(t) e_t$$

Il s'ensuit de (11) que les opérateurs de convolution commutent avec les dérivations. Pour toute fonction $f \in \mathcal{S}_n$, on a $P(D) f \in \mathcal{S}_n$ et de plus :

$$[P(D)f] * e_t = f * (P(D) e_t) = P(t) [f * e_t] \quad (15)$$

et comme $\hat{g}(t) = (g * e_t)(0)$ pour toute fonction $g \in \mathcal{S}_n$, la relation (15) en $x = 0$ montre que :

$$[P(D)f]^\wedge(t) = P(t) \hat{f}(t)$$

pour tout $t \in \mathbb{R}^n$, ce qui est identique à (13).

Si $t = (t_1, t_2, \dots, t_n)$ et $t' = (t_1 + \varepsilon, t_2, \dots, t_n)$, on a :

$$\frac{\hat{f}(t') - \hat{f}(t)}{i\varepsilon} = \int_{\mathbb{R}^n} x_1 f(x) \frac{e^{-ix_1\varepsilon} - 1}{ix_1\varepsilon} e^{-ix \cdot t} dm_n^*(x)$$

Le théorème de convergence dominée (cf. article [AF 99, § 4]) permet de passer à la limite quand ε tend vers 0, pour obtenir :

$$-\frac{1}{i} \frac{\partial \hat{f}}{\partial t_1}(t) = \int_{\mathbb{R}^n} x_1 f(x) e^{-ix \cdot t} dm_n^*(x)$$

ce qui est le cas $P(x) = x_1$ de (14). Le cas général se montre de façon identique. ♦

En quelques mots, le théorème 1 affirme que la **transformation de Fourier échange dérivation et multiplication**. L'analogie entre (13) et (14) suggère aussi une symétrie entre la transformée de Fourier et son inverse. On a en effet :

Lemme 2

Soit $\phi_n : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par $\phi_n(x) = \exp\left(-\frac{1}{2}\|x\|^2\right)$

On a $\phi_n \in \mathcal{S}_n$ et $\hat{\phi}_n(t) = \phi_n(t)$ pour tout $t \in \mathbb{R}^n$.

Preuve ♦ Il est facile de vérifier que $\phi_n \in \mathcal{S}_n$. La fonction ϕ_1 est solution de l'équation différentielle :

$$y' + xy = 0 \quad (16)$$

et d'après le théorème 1, on a :

$$(x\phi_1)^\wedge = i(\hat{\phi}_1)'; \quad (\phi_1')^\wedge = ix\hat{\phi}_1$$

Par linéarité de l'application $\phi \rightarrow \hat{\phi}$, il s'ensuit de $\hat{\phi}_1$ est également solution de (16), donc ϕ_1 et $\hat{\phi}_1$ sont proportionnelles. Mais $\phi_1(0) = 1$ et :

$$\hat{\phi}_1(0) = \int_{-\infty}^{+\infty} \phi_1(t) dm_1^*(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1$$

d'où $\hat{\phi}_1 = \phi_1$. On remarque ensuite que :

$$\phi_n(x) = \phi_1(x_1) \phi_1(x_2) \dots \phi_1(x_n)$$

Donc, en intégrant :

$$\hat{\phi}_n(t) = \hat{\phi}_1(t_1) \hat{\phi}_1(t_2) \dots \hat{\phi}_1(t_n)$$

d'où $\hat{\phi}_n = \phi_n$. ♦

Ce lemme permet de montrer le théorème fondamental d'**inversion de Fourier** :

Théorème 2

Pour toute fonction $g \in \mathcal{S}_n$ et tout $x \in \mathbb{R}^n$, on a :

$$g(x) = \int_{\mathbb{R}^n} \hat{g}(t) e^{ix \cdot t} dm_n^*(t) = \hat{\hat{g}}(-x) \quad (17)$$

Preuve ♦ Soit f et g dans \mathcal{S}_n . En échangeant l'ordre des intégrations dans :

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x) g(y) e^{-ix \cdot y} dm_n^*(x) dm_n^*(y)$$

on constate que :

$$\int_{\mathbb{R}^n} \hat{f}(y) g(y) dm_n^*(y) = \int_{\mathbb{R}^n} f(x) \hat{g}(x) dm_n^*(x) \quad (18)$$

Si $\phi \in \mathcal{S}_n$, $\lambda > 0$ et $f(x) = \phi(x/\lambda)$, on voit par un changement de variable que :

$$\hat{f}(y) = \lambda^n \hat{\phi}(\lambda y) \quad (19)$$

D'après (18) et (19) et le changement de variable $t = \lambda y$, on a donc :

$$\int_{\mathbb{R}^n} g\left(\frac{t}{\lambda}\right) \hat{\phi}(t) dm_n^*(t) = \int_{\mathbb{R}^n} \phi\left(\frac{x}{\lambda}\right) \hat{g}(x) dm_n^*(x)$$

et, en faisant tendre λ vers $+\infty$, on en déduit par convergence dominée que :

$$g(0) \int_{\mathbb{R}^n} \hat{\phi}(t) dm_n^*(t) = \phi(0) \int_{\mathbb{R}^n} \hat{g}(x) dm_n^*(x) \quad (20)$$

Si on applique (20) à la fonction ϕ_n du lemme 2, on en déduit que pour toute fonction $g \in \mathcal{S}_n$:

$$g(0) = \int_{\mathbb{R}^n} \hat{g}(x) dm_n^*(x) \quad (21)$$

Pour terminer la démonstration, on remarque que :

$$\begin{aligned} (\tau_{-y} g)^\wedge(t) &= \int_{\mathbb{R}^n} g(x+y) e^{-it \cdot x} dm_n^*(x) \\ &= \int_{\mathbb{R}^n} g(u) e^{-itu} e^{ity} dm_n^*(u) = e_{iy}(t) \hat{g}(t) \end{aligned}$$

et par conséquent, on a en appliquant (21) à $(\tau_{-y} g) \in \mathcal{S}_n$:

$$g(y) = (\tau_{-y} g)(0) = \int_{\mathbb{R}^n} (\tau_{-y} g)^\wedge dm_n^* = \int_{\mathbb{R}^n} \hat{g} e_{iy} dm_n^*$$

et donc

$$g(y) = \hat{\hat{g}}(-y) \quad \blacklozenge$$

Nous avons déjà observé (cf. article [AF 100, relation (24)]) que $(f * g)^\wedge = \hat{f} \hat{g}$. Le théorème 2 permet d'en déduire que $(fg)^\wedge = \hat{f} * \hat{g}$. Donc, la **transformation de Fourier échange la convolution et le produit ponctuel**.

L'équation (17) est vérifiée pour presque tout x lorsque g et \hat{g} sont intégrables. Elle permet aussi de déduire de (18) que pour toutes f et $g \in \mathcal{S}_n$, on a la **formule de Parseval** :

$$\int_{\mathbb{R}^n} f(y) \overline{g(y)} dm_n^*(y) = \int_{\mathbb{R}^n} \hat{f}(x) \overline{\hat{g}(x)} dm_n^*(x) \quad (22)$$

qui montre en particulier que $\|f\|_2 = \|\hat{f}\|_2$ pour toute $f \in \mathcal{S}_n$. Mais on peut alors prolonger par continuité la transformation de Fourier et montrer :

Corollaire 1

Il existe une unique isométrie inversible \mathcal{F} de $L^2(\mathbb{R}^n)$ sur $L^2(\mathbb{R}^n)$ telle que $\mathcal{F}(f) = \hat{f}$ pour toute $f \in \mathcal{S}_n$.

Nous pouvons donc voir la transformation de Fourier comme un **opérateur unitaire** sur l'espace de Hilbert $L^2(\mathbb{R}^n)$. Notons que (22) reste vraie si f et g sont dans L^2 , et que l'on peut également donner un sens à la définition 3 et à (17) dans L^2 , bien qu'une fonction $f \in L^2$ ne soit pas en général intégrable sur \mathbb{R}^n .

La transformation de Fourier s'étend encore, en respectant les équations (13) et (14), aux distributions tempérées $u \in \mathcal{S}'_n$, si l'on définit $\hat{u} \in \mathcal{S}'_n$ par :

$$\hat{u}(\phi) = u(\hat{\phi})$$

pour toute $\phi \in \mathcal{S}_n$. Elle échange toujours convolution et produit ponctuel lorsque ceux-ci ont un sens.

Énonçons à présent un théorème montré indépendamment par Malgrange et Ehrenpreis en 1954.

Théorème 3

Soit P un polynôme à n variables à coefficients complexes. Il existe une distribution u_0 sur \mathbb{R}^n telle que $P(D)(u_0) = \delta_{\{0\}}$.

Remarquons que si v est \mathcal{C}^∞ à support compact, on peut écrire :

$$P(D)(u_0 * v) = (P(D) u_0) * v = \delta_{\{0\}} * v = v$$

et donc l'équation $P(D)(y) = v$ a une solution $y = u_0 * v$, qui est une fonction \mathcal{C}^∞ . Le théorème de Malgrange-Ehrenpreis est donc un résultat fondamental d'existence de solutions pour les équations aux dérivées partielles à coefficients constants.

Esquissons une preuve de ce théorème dans le cas où la fonction $(1/P)$ est une distribution tempérée et où $v \in \mathcal{S}_n$. On cherche $u \in \mathcal{C}^\infty(\mathbb{R}^n)$ telle que :

$$P(D) u = v \quad (23)$$

D'après (13), u doit vérifier $(P(D) u)^\wedge = P \cdot \hat{u} = \hat{v}$, donc $\hat{u} = \frac{1}{P} \cdot \hat{v}$, mais comme \mathcal{F} échange produit ponctuel et convolution, il s'ensuit que :

$$\hat{u} = \left(\frac{1}{P} \cdot \hat{v} \right)^\wedge = \left(\frac{1}{P} \right) * \hat{v}$$

et, d'après le théorème 2 que $u = \hat{Q} * v$ avec $Q(x) = \frac{1}{P(-x)}$.

La distribution tempérée $u_0 = \hat{Q}$ satisfait donc la conclusion du théorème 3. Remarquons qu'une solution de l'équation $P(D) u = v$ s'obtient en convolant une solution fondamentale u_0 avec le second membre v de l'équation. Il faut donc s'attendre, lorsque l'on résout une équation différentielle ou, plus généralement, une équation aux dérivées partielles, à trouver une solution obtenue par convolution avec le second membre.

L'argument ci-dessus suggère que l'équation $P(D)(u) = v$ se comporte mieux lorsque l'on a des estimations de croissance sur $\left(\frac{1}{P}\right)$, donc lorsque P s'annule peu. Il en est bien ainsi : les solutions des équations elliptiques, que nous définissons maintenant, ont une propriété de régularité automatique exprimée par le théorème 4.

Définition 4

Un opérateur différentiel sur \mathbb{R}^n de la forme $L = \sum_{|\alpha| \leq N} f_\alpha D_\alpha$

où $f_\alpha \in \mathcal{C}^\infty(\mathbb{R}^n)$ pour tout α est dit elliptique si la fonction :

$$P(x, y) = \sum_{|\alpha| = N} f_\alpha(x) y^\alpha$$

satisfait $P(x, y) \neq 0$ pour tout $x \in \mathbb{R}^n$ et tout $y \in \mathbb{R}^n \setminus \{0\}$.

La démonstration du théorème ci-dessous utilise le lemme (et les espaces) de Sobolev, qui dépassent le cadre de cet exposé.

Théorème 4

Soit L un opérateur elliptique. Si u est une distribution telle que $L(u) = v \in \mathcal{C}^\infty(\mathbb{R}^n)$, alors $u \in \mathcal{C}^\infty(\mathbb{R}^n)$.

Un cas particulier important du théorème 4 est : si $L(u) = 0$, alors $u \in \mathcal{C}^\infty(\mathbb{R}^n)$.

Exemples 2 :

a) Si $n = 2$ et $L(u) = \frac{\partial^2 u}{\partial x_1 \partial x_2}$ (24)

alors $P(x, y) = -y_1 y_2$ et L n'est pas elliptique. Nous remarquons que toute fonction de la forme :

$$u(x_1, x_2) = u_1(x_1) + u_2(x_2)$$

est solution de (24) et que les fonctions u_1 et u_2 peuvent être, par exemple, dérivables mais pas \mathcal{C}^1 .

b) Si $n = 2$ et $L(u) = \frac{\partial^2 u}{\partial x_1^2} - \frac{\partial^2 u}{\partial x_2^2}$ (25)

alors $P(x, y) = -y_1^2 + y_2^2$ et L n'est pas elliptique. Les solutions de $L(u) = 0$ sont les fonctions :

$$u(x_1, x_2) = u_1(x_1 - x_2) + u_2(x_1 + x_2)$$

et la remarque ci-dessus s'applique. (25) est l'« équation des cordes vibrantes ».

c) L'opérateur de Laplace $\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \dots + \frac{\partial^2}{\partial x_n^2}$

est elliptique puisque son polynôme caractéristique est $P(x, y) = -(y_1^2 + y_2^2 + \dots + y_n^2)$.

Les solutions de $\Delta u = 0$ sont les **fonctions harmoniques**, qui d'après le théorème 4 sont nécessairement \mathcal{C}^∞ .

d) L'équation de la chaleur dans \mathbb{R}^3 (sans conditions au bord) s'écrit :

$$\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0 \quad (26)$$

où t représente le temps et (x, y, z) un point de \mathbb{R}^3 ; $u(t, x, y, z)$ est la température en (x, y, z) au temps $t > 0$, sachant que :

$$u(0, x, y, z) = f(x, y, z) \quad (27)$$

est la température initiale. La fonction u qui satisfait (26) et (27) pour $t > 0$ est donnée par :

$$u(t, x, y, z) = h_t * f \quad (28)$$

où $h_t(v) = (4\pi t)^{-3/2} e^{-\|v\|^2/4t}$

et, pour tout $w \in \mathbb{R}^3$, $h_t * f(w) = \int_{\mathbb{R}^3} h_t(w-v) f(v) dv$

Le caractère \mathcal{C}^∞ des fonctions de Gauss (h_t) implique que la fonction u est \mathcal{C}^∞ pour tout $t > 0$. L'équation (28) modélise la diffusion de la chaleur par le « **semi-groupe de la chaleur** » h_t ; l'expression « semi-groupe » signifie que $h_{t+s} = h_t * h_s$. Cette équation représente le premier pas vers la modélisation mathématique du mouvement brownien, qui nécessite d'importants outils probabilistes.

3. Calcul des probabilités

3.1 Espaces de probabilité et variables aléatoires

Il est le plus souvent impossible d'effectuer une mesure, une vérification ou une expérience exhaustive sur la totalité d'une collection d'objets (ou de personnes), et cela conduit à sélectionner des échantillons significatifs sur lesquels on travaille. Cette démarche naturelle pose immédiatement deux problèmes :

comment obtenir l'échantillon ? De quelle taille doit-il être pour être en effet significatif ?

Pour éliminer si possible un choix arbitraire de l'expérimentateur, il convient de choisir les éléments de l'échantillon « **au hasard** ». Quant à la taille, elle doit permettre d'obtenir un résultat qui est « **probablement** » juste, à une « **petite erreur** » près.

Pour remplacer les expressions entre guillemets ci-dessus par des notions précises et des données quantitatives, nous avons besoin d'une modélisation mathématique. Celle-ci est fournie par la théorie des probabilités, développée par les écoles française (autour d'E. Borel et P. Lévy) et russe (présidée par A. Kolmogorov) au XX^e siècle, après une « préhistoire » où brillent en particulier les noms de Pascal, de Laplace et de Gauss.

Nous allons voir que cette modélisation utilise largement les outils de l'analyse fonctionnelle. Les problèmes concrets qui motivent la théorie concernent bien sûr des ensembles finis ; mais il est souvent plus facile et plus efficace de « passer à la limite » et de remplacer le fini très grand par l'infini. On est alors amené à travailler dans les espaces de Banach $L_p(\mathbb{R}^n)$, introduits dans l'article [AF 99, § 5], et en particulier dans l'espace de Hilbert $L_2(\mathbb{R}^n)$.

Nous devons d'abord préciser le sens de l'expression « au hasard ». Nous avons pour cela :

Définition 5

Un espace de probabilité est un ensemble Ω muni d'une famille Σ de parties mesurables, stable par union et intersection dénombrables et passage au complémentaire, et d'une application $\mathbb{P} : \Sigma \rightarrow [0, 1]$ telle que $\mathbb{P}(\Omega) = 1$ et :

$$\mathbb{P}\left(\bigcup_{k \geq 0} A_k\right) = \sum_{k=0}^{+\infty} \mathbb{P}(A_k)$$

pour toute suite disjointe (A_k) dans Σ .

La fonctionnelle \mathbb{P} est appelée une **mesure de probabilité**. L'ensemble Ω doit être compris comme l'ensemble des échantillons, ou des résultats possibles d'une expérience.

Si $A \in \Sigma$, le nombre $\mathbb{P}(A) \in [0, 1]$ est la probabilité pour que le résultat w soit dans A .

Exemples 3 :

a) Si l'expérience est « **le jet d'un dé équilibré** », on a :

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

et pour tout $A \subseteq \Omega$, $\mathbb{P}(A) = \frac{|A|}{6}$, où $|A|$ désigne le nombre d'éléments de A .

b) Si l'expérience est « **n lancers successifs d'une pièce équilibrée** », on a :

$$\Omega = \{0, 1\}^n$$

où on convient que « pile = 0 » et « face = 1 », et pour tout $A \subseteq \Omega$, $\mathbb{P}(A) = 2^{-n} |A|$.

c) Si la pièce de l'exemple ci-dessous est truquée et que « face » sort avec une probabilité $p \neq 1/2$, on a toujours $\Omega = \{0, 1\}^n$ mais pour tout $w \in \Omega$,

$$\mathbb{P}(\{w\}) = p^k (1-p)^{n-k}$$

où k est le nombre de « 1 » dans w .

d) Si l'expérience est « **la prise au hasard d'un nombre réel** $x \in [0, 1]$ », on a $\Omega = [0, 1]$ et la probabilité \mathbb{P} est la mesure de Lebesgue m_1 sur $[0, 1]$.

Mentionnons à cette occasion que de nombreux « paradoxes » de la théorie des probabilités reposent sur le simple fait que la « prise au hasard » n'est pas précisée, et peut donc, par exemple, changer de sens.

Lorsqu'une expérience est réalisée, on lui associe souvent un nombre réel (ou une liste de k nombres réels) qui mesure son résultat. Voici la définition correspondante.

Définition 6

Une variable aléatoire est une application $X : \Omega \rightarrow \mathbb{R}^k$ telle que $X^{-1}(U) \in \Sigma$ pour tout ouvert U de \mathbb{R}^k .

Exemple 4 : dans l'exemple **3b** précédent du jeu de pile ou face, soit $X : \Omega = \{0, 1\}^n \rightarrow \mathbb{R}$ définie par $X(w) =$ « nombre de 1 dans w ». La variable X mesure le nombre de fois où « face » sort dans une suite w de n lancers.

Les variables aléatoires permettent d'utiliser pleinement la théorie de la mesure, telle qu'elle est esquissée dans l'article [AF 100, § 3] (exemple **2**), qui étend la théorie de la mesure géométrique de Lebesgue présentée paragraphe 4 de l'article [AF 99]. Pour préciser ce lien, nous avons besoin d'un jeu de terminologie.

Soit $X : \Omega \rightarrow \mathbb{R}^k$. La **loi** \mathbb{P}_X de la variable aléatoire X est la mesure de probabilité définie sur les parties boréliennes de \mathbb{R}^k par :

$$\mathbb{P}_X(A) = \mathbb{P}[X^{-1}(A)] \quad (29)$$

La variable X transfère ainsi la mesure de probabilité \mathbb{P} , définie sur Ω , en une mesure \mathbb{P}_X sur \mathbb{R}^k . Ce procédé fait disparaître la variable $w \in \Omega$; en pratique, **ce qui importe est la loi** de la variable aléatoire, conformément au point de vue moderne (initié par Lebesgue) sur l'intégration.

Remarque 1 : on utilise également le terme de « distribution » de X pour désigner \mathbb{P}_X , que nous éviterons ici bien que \mathbb{P}_X soit en particulier une distribution sur \mathbb{R}^k . Lorsque $k = 1$, on peut considérer la **fonction de répartition** $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par $F_X(t) = \mathbb{P}(\{w ; X(w) < t\})$. On vérifie alors que $(F_X)' = \mathbb{P}_X$, où la dérivée de F_X est prise au sens des distributions.

L'**espérance** de la variable aléatoire X est la quantité :

$$\mathbb{E}(X) = \int_{\mathbb{R}^k} t \, d\mathbb{P}_X(t) = \int_{\Omega} X(w) \, d\mathbb{P}(w)$$

où la deuxième égalité s'ensuit de (29) par un changement de variable. Cette quantité $\mathbb{E}(X)$ n'est autre que la **moyenne** de X ; elle n'est bien définie que lorsque X est \mathbb{P} -intégrable. Plus généralement ; on dit que X a un **moment d'ordre** $p \geq 1$ si la quantité :

$$\mathbb{E}(X^p) = \int_{\mathbb{R}^k} t^p \, d\mathbb{P}_X(t)$$

est finie ; X a un moment d'ordre p lorsque $X \in L_p(\Omega, \mathbb{P})$. Cette condition est automatiquement satisfaite lorsque Ω est un ensemble fini. Le cas $p = 2$, qui permet d'utiliser l'espace de Hilbert, est particulièrement important. Lorsque X a un moment d'ordre 2, on définit sa **variance** $\text{Var}(X)$ par :

$$\text{Var}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

et son **écart-type** σ_X est $\sigma_X = \sqrt{\text{Var}(X)}$.

Si l'image d'une variable aléatoire X est un nuage de points dans \mathbb{R}^k , $\text{Var}(X)$ est le moment d'inertie de ce nuage par rapport au centre de gravité. L'écart-type σ_X est une quantité homogène qui mesure la concentration de X autour de sa valeur moyenne.

Enfin, la **fonction caractéristique** $\varphi_X : \mathbb{R}^k \rightarrow \mathbb{C}$ de X est définie par :

$$\varphi_X(t) = \int_{\mathbb{R}^k} e^{it \cdot x} \, d\mathbb{P}_X(x) = \int_{\Omega} e^{it \cdot X(w)} \, d\mathbb{P}(w) \quad (30)$$

Donc, φ_X est essentiellement la transformée de Fourier de la loi \mathbb{P}_X . Il s'ensuit de l'inversion de Fourier (cf. théorème 2, § 2) que la loi de X est uniquement déterminée par φ_X .

Exemple 5 : on considère l'exemple **3c** du « pile ou face avec une pièce truquée », et la variable aléatoire $X = \{0, 1\}^n \rightarrow \mathbb{R}$ définie par $X(w) =$ nombre de « 1 » dans $w \in \{0, 1\}^n =$ nombre de « face » obtenu dans la suite de lancers w .

La loi \mathbb{P}_X de cette variable est notée $B(n, p)$ et est appelée **loi binomiale** de paramètres n et p . Elle est concentrée sur l'ensemble $\{0, 1, 2, \dots, n\}$ et pour k dans cet ensemble, on a :

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(\{X = k\}) = C_n^k p^k (1-p)^{n-k}$$

$$\text{Son espérance est } \mathbb{E}(X) = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = np$$

et sa variance est $\text{Var}(X) = np(1-p)$.

3.2 Indépendance

Nous introduisons maintenant un concept probabiliste fondamental.

Définition 7

Deux variables aléatoires X et Y définies sur le même espace de probabilité Ω sont dites **indépendantes** si pour tous les ouverts U et V , on a :

$$\mathbb{P}[X^{-1}(U) \cap Y^{-1}(V)] = \mathbb{P}[X^{-1}(U)] \cdot \mathbb{P}[Y^{-1}(V)] \quad (31)$$

Ci-dessus, on a noté :

$$\{w \in \Omega ; X(w) \in U\} = X^{-1}(U)$$

Un moment de réflexion montre que (31) exprime le fait que la connaissance de X ne donne aucune information sur Y , d'où l'indépendance « causale » de X et Y .

Si X et Y sont indépendantes, on a $\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y)$ et :

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \quad (32)$$

quand celles-ci sont définies. Notons que (32) provient de l'orthogonalité dans l'espace de Hilbert $L_2(\Omega, \mathbb{P})$ des variables centrées $(X - \mathbb{E}(X))$ et $(Y - \mathbb{E}(Y))$; c'est donc une version du théorème de Pythagore.

Le théorème suivant permet de comprendre les **sommes** de variables aléatoires indépendantes.

Théorème 5

Soit X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{R}^k , et $(X+Y)$ leur somme. Les fonctions caractéristiques vérifient :

$$\varphi_{X+Y} = \varphi_X \cdot \varphi_Y$$

Preuve ♦ Nous esquissons l'argument lorsque X et Y sont à valeurs dans \mathbb{Z} ; le cas général est analogue.

Pour tout $k \in \mathbb{Z}$, on a :

$$\{X+Y=k\} = \bigcup_{n \in \mathbb{Z}} (\{X=n\} \cap \{Y=k-n\})$$

et cette réunion est disjointe, donc :

$$\mathbb{P}(X+Y=k) = \sum_{n=-\infty}^{+\infty} \mathbb{P}(\{X=n\} \cap \{Y=k-n\})$$

Mais puisque X et Y sont indépendantes, on a :

$$\mathbb{P}(X+Y=k) = \sum_{n=-\infty}^{+\infty} \mathbb{P}(X=n) \cdot \mathbb{P}(Y=k-n) \quad (33)$$

et (33) signifie que $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$ (cf. article [AF 100 relation 46]) donc que **la loi d'une somme de deux variables indépendantes est la convolée des deux lois**. Puisque la transformation de Fourier échange la convolution et le produit ponctuel (cf. article [AF 100 relation 47] et paragraphe 2 de ce présent article), il s'ensuit que :

$$\varphi_{X+Y} = \varphi_X \cdot \varphi_Y \quad \blacklozenge$$

Les résultats ci-dessus s'étendent sans peine aux suites de variables aléatoires indépendantes deux à deux. On considère souvent des **suites de variables indépendantes et de même loi**, qui modélisent des répétitions indépendantes d'une même expérience.

Exemple 6 : la variable aléatoire X de l'exemple **5** est la somme de n variables indépendantes de même loi \mathbb{P}_{X_0} définie par $\mathbb{P}_{X_0}(\{1\}) = p$, $\mathbb{P}_{X_0}(\{0\}) = 1-p$. On a :

$$\varphi_{X_0}(t) = p e^{it} + (1-p) = 1 + p(e^{it} - 1)$$

et donc :

$$\varphi_X(t) = [1 + p(e^{it} - 1)]^n \quad (34)$$

3.3 Lois limites

Nous voudrions maintenant préciser la notion de « probable-ment juste, à une petite erreur près ». Notre espace Ω rassemble tous les « échantillons » $w \in \Omega$, et $X(w)$ est la mesure de l'échantillon. L'espérance $\mathbb{E}(X)$ est la vraie valeur, si \mathbb{P} modélise bien la prise au hasard de ces échantillons. Si on arrête les mesures après avoir sélectionné n échantillons et fait la moyenne des résultats obtenus, obtient-on une valeur proche de la vraie valeur – si on n'a pas été vraiment malchanceux dans ces n premiers tirages ?

La réponse (positive) à cette question est le contenu de la **loi faible des grands nombres** ci-dessous.

Théorème 6

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes de même loi. On suppose que $\mathbb{E}(X_1) = m$ et $\sigma_{X_1} = \sigma$ existent. Soit :

$$S_n = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

On a alors, pour tout $h > 0$:

$$\mathbb{P}[|S_n - m| > h] < \frac{\sigma^2}{nh^2} \quad (35)$$

Preuve ♦ On a $\mathbb{E}(S_n) = m$. D'après l'indépendance des (X_k) , on peut appliquer (32) qui implique :

$$\mathbb{E}[(S_n - m)^2] = \frac{\sigma^2}{n} \quad (36)$$

Pour toute variable aléatoire Z et $u > 0$, on a, en « plaçant un rectangle sous le graphe de $|Z|$ », l'inégalité de Bienaymé-Tchebichev :

$$\mathbb{P}(|Z| > u) \leq \frac{\mathbb{E}(|Z|)}{u} \quad (37)$$

En appliquant (37) à $Z = (S_n - m)^2$ et à $u = h^2$, on déduit (35) de (36). \blacklozenge

D'après (35), les moyennes S_n tendent vers la constante $m = \mathbb{E}(X_1)$ **en probabilité**, c'est-à-dire que pour tout $h > 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}[|S_n - m| > h] = 0$.

D'après la **loi forte** des grands nombres, on a même convergence presque partout et dans $L_1(\Omega, \mathbb{P})$. Notons au passage que la confusion entre différents modes de convergence est la source de bien des erreurs et « paradoxes ».

Le théorème 6 énonce que la suite $\frac{1}{n}(X_1 + X_2 + \dots + X_n)$ est proche de la constante m , pour n grand, avec une grande probabilité. Le théorème fondamental suivant, dit « **de la limite centrée** », donne des précisions quantitatives.

Théorème 7

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires à valeurs réelles satisfaisant les hypothèses du théorème 6.

Pour tout $x > 0$, on a :

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left[\left|S_n - m\right| > \frac{\sigma x}{\sqrt{n}}\right] = \sqrt{\frac{2}{\pi}} \int_x^{+\infty} e^{-\frac{t^2}{2}} dt$$

Preuve ♦ On suppose par translation que $m = 0$. On pose :

$$Z_n = \frac{1}{\sigma\sqrt{n}} \left(\sum_{i=1}^n X_i \right)$$

Un changement de variable montre que :

$$\varphi_{\frac{X}{\sigma\sqrt{n}}}(t) = \varphi_X\left(\frac{t}{\sigma\sqrt{n}}\right)$$

et d'après l'indépendance des (X_i) , on peut utiliser le théorème 5 qui montre que :

$$\varphi_{Z_n}(t) = \left[\varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n \quad (38)$$

Un développement limité de l'exponentielle dans la définition (30) de φ_{X_1} montre que :

$$\varphi_{X_1}(u) = 1 - \frac{u^2 \sigma^2}{2} + \varepsilon(u^2)$$

et (38) montre alors que :

$$\varphi_{Z_n}(t) = \left[1 - \frac{t^2}{2n} + \varepsilon\left(\frac{t^2}{2n}\right) \right]^n$$

et par conséquent :

$$\lim_{n \rightarrow +\infty} \varphi_{Z_n}(t) = e^{-\frac{t^2}{2}} \quad (39)$$

Mais d'après le lemme 2, la fonction $f(t) = e^{-\frac{t^2}{2}}$ est sa propre transformée de Fourier, à normalisation près.

Le résultat s'ensuit alors de (39) et du théorème 2 d'inversion de Fourier. ♦

Une variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ est dite **gaussienne** si sa loi \mathbb{P}_X est une loi de Gauss, encore appelée loi **normale**, c'est-à-dire si pour tout $a < b$,

$$\mathbb{P}_X([a, b]) = \mathbb{P}(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(t-m)^2}{2\sigma^2}} dt$$

où $m = \mathbb{E}(X)$ et $\sigma = \sqrt{\text{Var}(X)}$. Lorsque $m = 0$ et $\sigma = 1$, on parle de loi **normale réduite**. Le théorème 7 explique l'importance de la

loi normale. Ainsi, une suite de valeurs observées pour une grandeur physique donnée se répartit en général de façon gaussienne autour de la valeur réelle : les « grandes erreurs » sont rares, et le théorème 7 nous dit cela de manière quantitative. De plus, **les grandeurs qui résultent d'un grand nombre de petits facteurs indépendants se répartissent autour de leur moyenne suivant une loi gaussienne** ; c'est en général le cas des données biologiques.

Rappelons que la variable aléatoire $X = X_{n,p}$ de l'exemple 5, de loi binomiale $B(n, p)$, est la somme de n variables indépendantes. Il est donc possible de lui appliquer le théorème 7 ; de fait, la loi de la variable :

$$\frac{X_{n,p} - np}{\sqrt{np(1-p)}}$$

tend vers une loi normale réduite quand n tend vers $+\infty$ (on dit alors que l'on a **convergence en loi**). Mais on peut considérer une autre forme de convergence, où n tend vers $+\infty$ mais où $p_n = \frac{\lambda}{n}$ tend vers 0, avec $np_n = \lambda > 0$.

On modélise, dans ce cas, la répétition d'un grand nombre d'expériences identiques qui ont toutes une faible chance de succès. D'après (34), si X_n a la loi $B\left(n, \frac{\lambda}{n}\right)$, sa fonction caractéristique est :

$$\varphi_{X_n}(t) = \left(1 + \frac{\lambda}{n}(e^{it} - 1)\right)^n$$

On a donc :

$$\lim_{n \rightarrow +\infty} \varphi_{X_n}(t) = e^{\lambda(e^{it} - 1)} \quad (40)$$

Identifions la fonction limite dans (40) comme une fonction caractéristique. Si la loi de X est la **loi de Poisson** \mathcal{P}_λ , concentrée sur \mathbb{N} et telle que :

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

pour tout $k \in \mathbb{N}$, on a d'après (30) :

$$\begin{aligned} \varphi_X(t) &= \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{itk} \\ &= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} \\ &= e^{\lambda(e^{it} - 1)} \end{aligned}$$

Il s'ensuit donc de (40) que la loi de (X_n) converge vers la loi de Poisson \mathcal{P}_λ de paramètre λ . Cette loi de Poisson est parfois appelée « **loi des événements rares** » ; elle modélise la probabilité pour que, dans un intervalle de temps donné, un événement qui se produit très rarement se produise k fois dans une population très importante. On peut ainsi modéliser les accidents graves ou les anomalies. Un exemple physique est donné par le nombre de particules alpha (c'est-à-dire de noyaux d'hélium) émises dans un intervalle de temps de longueur t par un gramme de radium ; dans ce cas, on a approximativement $n = 10^{22}$ et $p = 10^{-12} t$, où t est mesuré en secondes.

La théorie des probabilités recèle de très nombreuses applications aux sciences naturelles, aux statistiques et aux mathématiques financières. Mais cela nous éloignerait petit à petit de l'analyse fonctionnelle.