Vrac IV

Table des matières

1 Théorème de Radon-Nikodym

Référence : Marc Briane, Gilles Pagès, *Théorie de l'intégration*

Théorème 1.1 : Théorème de RADON-NIKODYM

Soient μ et ν deux mesures σ-finies sur un espace mesurable $(X, \mathcal{A})$. Les assertions suivantes sont équivalentes :

(i) Pour tout $A \in \mathcal{A}$,

$$\mu(A) = 0 \Longrightarrow \nu(A) = 0$$

On dit que ν est *absolument continue* par rapport à μ : $\nu \ll \mu$;

(ii) Il existe une fonction $f \in L^1(\mu)$ positive μ-presque partout telle que

$$\forall A \in \mathcal{A}, \nu(A) = \int_A f \, \mathrm{d}\mu$$

Dans ce cas, la fonction f est unique μ-presque partout.

Démonstration :

[(ii) $\Longrightarrow$ (i)] Si $A \in \mathcal{A}$ vérifie $\mu(A) = 0$ alors

$$\nu(A) = \int_A f \, \mathrm{d}\mu = 0$$

[(i) $\Longrightarrow$ (ii)] On procède en quatre étapes :

1. Plaçons-nous d'abord dans le cas où $\nu(X), \mu(X) < +\infty$, et plus particulièrement ici $\nu \leqslant \mu$. Le théorème de Riesz nous fournit $f \in L^1(\mu)$ tel que $f \in [0,1]$ presque partout, et vérifiant **(ii)**.
2. Puis, dans le cas fini, on considère le cas général en remarquant que $\mu \leqslant \mu + \nu$.
3. On conclut sur l'existence de f dans le cas σ-fini en utilisant le théorème de Beppo Lévi.
4. On conclut en montrant l'unicité de f.

1. Supposons que $\nu \leqslant \mu$ *id est*

$$\forall A \in \mathcal{A}, \nu(A) \leqslant \mu(A)$$

Alors pour toute fonction g mesurable positive sur X :

$$\int_X g \, \mathrm{d}\nu \leqslant \int_X g \, \mathrm{d}\mu$$

De plus, ν et μ vérifient bien (i). On définit l'application

$$\Phi : \begin{pmatrix} L^2(\mu) & \longrightarrow & \mathbb{R} \\ g & \longmapsto & \int_X g \, \mathrm{d}\nu \end{pmatrix}$$

Alors Φ est une application linéaire bien définie, par l'inégalité sur les mesures. De plus, Φ est continue pour les normes usuelles : pour $g \in L^2(\mu)$, l'inégalité de Cauchy-Schwarz donne

$$|\Phi(g)| \leqslant \sqrt{\nu(X)} \cdot \|g\|_{L^2}$$

Par le théorème de représentation de Riesz, il existe alors une fonction $f \in L^2(\mu)$ tel que

$$\forall g \in L^2(\mu), \int_X g \, \mathrm{d}\nu = \int_X gf \, \mathrm{d}\mu$$

Montrons que $f \in [0,1]$ μ-presque partout. Supposons que $\mu(f < 0) > 0$. Puisque

$$\{f < 0\} = \bigcup_{n \in \mathbb{N}^*}^{\uparrow} \left\{ f \leqslant \frac{-1}{n} \right\}$$

il existe $n_0 > 0$ tel que $\mu\left(f \leqslant \frac{-1}{n_0}\right) < 0$. Alors

$$0 \leqslant \nu\left(f \leqslant \frac{-1}{n_0}\right) = \int_{\left\{f \leqslant \frac{-1}{n_0}\right\}} f \mathrm{d}\mu$$

Donc

$$0 \leqslant \nu(A) \leqslant \frac{-1}{n_0} \mu\left(f \leqslant \frac{-1}{n_0}\right) < 0$$

ce qui nous amène à une contradiction. Ainsi, $f \geqslant 0$ μ-presque partout. On montre de même que $f \leqslant 1$ μ-presque partout en utilisant

$$\{f > 1\} = \bigcup_{n \in \mathbb{N}^*}^{\uparrow} \left\{ f \geqslant 1 + \frac{1}{n} \right\}$$

Reste pour conclure à montrer que $f \in L^1(\mu)$. Puisque $\mu(X) < +\infty$ alors la fonction constante égale à 1 est dans $L^2(\mu)$ et donc

$$\Phi(1) = \int_X f \mathrm{d}\mu = \nu(X) < +\infty$$

Puisque $f \geqslant 0$, il suit que $f \in L^1(\mu)$. On a donc montré l'existence de f dans le cas $\mu(X), \nu(X)$ finis avec $\nu \leqslant \mu$ tel que $f \in L^1(\mu)$ tel que

$$\nu(A) = \Phi(\mathbf{1}_A) = \int_A f \, \mathrm{d}\mu$$

2. Restons dans le cadre $\mu(X), \nu(X)$ finis, mais dans le cas général quand au signe de leur différence. On a alors

$$\nu \leqslant \nu + \mu$$

Par **1.**, il existe alors $f \in L^1(\mu)$ telle que

$$\nu(A) = \int_A f \, \mathrm{d}(\mu + \nu)$$

Alors

$$\forall A \in \mathcal{A}, \int_A f \, \mathrm{d}\mu = \int_A (1-f) \, \mathrm{d}\nu$$

Notons $N = \{f = 1\}$. Montrons que N est μ-négligeable, et donc ν-négligeable par (i).

$$\mu(N) = \int_N \mathrm{d}\mu = \int_N f \, \mathrm{d}\mu$$

Ainsi,

$$\mu(N) = \int_N (1-f) \, \mathrm{d}\nu = 0$$

Cela nous permet de conclure car si $A \in \mathcal{A}$ alors

$$\nu(A) = \nu(A \cap N) + \nu\left(A \cap \complement_X N\right)$$

Le premier terme est nul. Quand au deuxième, le fait d'être sur le complémentaire de N permet de diviser par $1-f$:

$$\nu(A) = \int_A \frac{\mathbf{1}_{\complement_X N} \, (1-f)}{1-f} \, \mathrm{d}\nu$$

D'où

$$\nu(A) = \int_A \frac{\mathbf{1}_{\complement_X N} \, f}{1-f} \, \mathrm{d}\mu$$

Avec de plus $\frac{\mathbf{1}_{\complement_X N} \, f}{1-f} \in L^1(\mu)$ car

$$\int_X \frac{\mathbf{1}_{\complement_X N} \, f}{1-f} \, \mathrm{d}\mu = \int_{\complement_X N} \frac{f}{1-f} \, \mathrm{d}\mu$$

Donc

$$\int_X \frac{\mathbf{1}_{\complement_X N} \, f}{1-f} \, \mathrm{d}\mu = \int_{\complement_X N} \frac{1-f}{1-f} \, \mathrm{d}\nu = \nu(X) < \infty$$

Ce qui montre l'existence de f comme dans (ii) dans le cas où μ et ν sont finis.

3. Plaçons nous alors dans le cadre du théorème : supposons que μ et ν sont σ-finis. Alors

$$X = \bigsqcup_{k \in \mathbb{N}} A_k = \bigsqcup_{l \in \mathbb{N}} B_l$$

avec $\mu(A_k), \nu(B_k) < \infty$. On note alors

$$E_{k,l} \overset{\text{déf.}}{=} A_k \cap B_k$$

Alors $\mu(E_{k,l}) \leqslant \mu(A_k) < \infty$ et $\nu(E_{k,l}) < \infty$. Notons alors $(E_n)_n = (E_{k,l})_{k,l}$, possible car $\mathbb{N}$ est $\mathbb{N}^2$ sont en bijection. On pose de plus les deux suites de mesure suivantes :

$$\mu_n \overset{\text{déf.}}{=} \mu(\cdot \cap E_n)$$

et

$$\nu_n \overset{\text{déf.}}{=} \nu(\cdot \cap E_n)$$

Alors pour tout $n \in \mathbb{N}$, μ_n et ν_n sont finies. Par **2.**, il existe $f_n \in L^1(\mu)$ telle que

$$\forall A \in \mathcal{A}, \nu_n(A) = \int_A f_n \, \mathrm{d}\mu_n = \int_A f_n \mathbf{1}_{E_n} \, \mathrm{d}\mu$$

On conclut alors en posant

$$f \overset{\text{déf.}}{=} \sum_{n=0}^{+\infty} f_n \mathbf{1}_{E_n}$$

On peut alors calculer :

$$\int_A f \, \mathrm{d}\mu = \int_A \left(\sum_{n=0}^{+\infty} f_n \mathbf{1}_{E_n} \right) \mathrm{d}\mu$$

Par le théorème de Beppo Lévi :

$$\int_A f \, \mathrm{d}\mu = \sum_{n=0}^{+\infty} \int_A f_n \mathbf{1}_{E_n} \mathrm{d}\mu$$

Donc

$$\int_A f \, \mathrm{d}\mu = \sum_{n=0}^{+\infty} \nu_n(A) = \nu(A)$$

On a donc montré l'existence de f dans le cas σ-fini, et avons conclut sur l'équivalence.

4. Reste à montrer l'unicité μ-presque partout de f. Si f et $\tilde{f}$ vérifient (ii) alors

$$\nu(f < \tilde{f}) = \int_{\{f<\tilde{f}\}} f \, \mathrm{d}\mu = \int_{\{f<\tilde{f}\}} \tilde{f} \, \mathrm{d}\mu$$

Donc

$$\int_{\{f<\tilde{f}\}} (\tilde{f} - f) \, \mathrm{d}\mu = 0$$

Puisque sur le domaine d'intégration, la fonction intégrée ne s'annule pas, il suit que

$$\mu(f < \tilde{f}) = 0$$

Par symétrie, il suit que

$$\mu(f \neq \tilde{f}) = 0$$

D'où l'unicité μ-presque partout. □

2 Passage à la limite et moyennisation

Référence : Marc BRIANE, Gilles PAGÈS, *Théorie de l'intégration*
On utilise la propriété de densité suivante.

Proposition 2.1 : Densité des fonctions en escaliers dans L^p

On note λ_d la mesure de LEBESGUE sur $\mathbb{R}^d$. On dit que $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}$ est en *en escalier* s'il existe des pavés $(P_k)_{k\in[\![1,N]\!]}$ de la forme $\prod_{i=1}^d I_{i,k}$ où $I_{i,k}$ sont des intervalles tels que φ soit combinaison linéaire des $\mathbf{1}_{P_k}$. Alors l'ensemble des fonctions en escalier à support compact est dense dans $L^p(\mathbb{R}^d)$ pour la norme $\|\cdot\|_{L^p}$ associée, pour tout $p \in [1, +\infty[$ fini.

On va alors montrer la propriété de moyenne suivante.

Théorème 2.1

Pour $d \geqslant 1$, on note $Y \overset{\text{déf.}}{=} [0,1]^d$ et

$$L^p_{\#}(Y) = \left\{ f \in L^p_{\text{loc}}(\mathbb{R}^d), f \ Y\text{-périodique } \lambda_d - \text{pp} \right\}$$

où par définition, f est Y-périodique si pour tout vecteur de la base canonique e_i et $x \in \mathbb{R}^d$, $f(x+e_i) = f(x)$. Soient p, p' deux exposants conjugués finis. Alors pour tout $f \in L^p_{\#}(Y)$, et $g \in L^{p'}(\mathbb{R}^d)$ nulle presque partout en dehors d'un compact :

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f(nx)g(x)\ \mathrm{d}x = \left[\int_Y f(y)\ \mathrm{d}y\right]\left[\int_{\mathbb{R}^d} g(x)\ \mathrm{d}x\right]$$

Lemme 2.1 : Cas particulier pour $g = \mathbf{1}_Q$

Soit $Q = \prod_{i=1}^d [a_i, b_i]$ un pavé de $\mathbb{R}^d$. Alors pour tout $h \in L^1_{\#}(Y)$:

$$\lim_{n\to+\infty} \int_Q h(nx)\ \mathrm{d}x = \lambda_d(Q) \cdot \int_Y h(y)\ \mathrm{d}y$$

Démonstration du lemme : On considère pour $n > \max_{i\in[\![1,d]\!]} \frac{1}{|b_i - a_i|}$, les pavés suivants :

$$\forall \kappa \in \mathbb{Z}^d, Y_\kappa \overset{\text{déf.}}{=} \frac{1}{n}\kappa + \frac{1}{n}[0,1[^d$$

1. Notons I_n et J_n les ensembles suivants :

$$I_n \overset{\text{déf.}}{=} \left\{\kappa \in \mathbb{Z}^d, Y_\kappa \subset Q\right\}$$

et

$$J_n \overset{\text{déf.}}{=} \left\{\kappa \in \mathbb{Z}^d, Y_\kappa \cap \partial Q \neq \varnothing\right\}$$

L'idée est de calculer l'intégrale sur Q en utilisant la décomposition suivante :

$$Q = \left[\bigsqcup_{\kappa\in I_n} Y_\kappa\right] \sqcup \left[\bigsqcup_{\kappa\in J_n} Y_\kappa \cap Q\right]$$

Ainsi, notre calcul en fil rouge de ce lemme est le suivant :

$$\int_Q h(nx)\ \mathrm{d}x = \sum_{\kappa\in I_n} \int_{Y_\kappa} h(nx)\ \mathrm{d}x + \sum_{\kappa\in J_n} \int_{Y_\kappa\cap Q} h(nx)\ \mathrm{d}x$$

3. Pour la première intégrale,

$$\sum_{\kappa\in I_n} \int_{Y_\kappa} h(nx)\ \mathrm{d}x = \frac{\sharp I_n}{n^d} \int_Y h(y)\ \mathrm{d}y$$

On va voir que sur Y_κ, on peut exprimer l'intégrale de $h(nx)$ en fonction de celle de h sur Y. Par un changement de variables $y = nx$:

$$\int_{Y_\kappa} h(nx)\ \mathrm{d}x = \frac{1}{n^d} \int_{nY_\kappa} h(y)\ \mathrm{d}y$$

Or $nY_\kappa = [0,1[^d + \kappa$. Par Y-périodicité de h, et puisque l'intégrale sur un hyperplan est nulle :

$$\int_{Y_\kappa} h(nx)\ \mathrm{d}x = \frac{1}{n^d}\int_Y h(y)\ \mathrm{d}y$$

4. Pour la deuxième intégrale, on a pour $\kappa \in \mathbb{Z}^d$ fixé :

$$\left|\int_{Y_\kappa \cap Q} h(nx)\ \mathrm{dx}\right| \leqslant \int_{Y_\kappa} |h(nx)|\ \mathrm{d}x$$

Le même type de calcul qu'au **3.** donne :

$$\left|\int_{Y_\kappa \cap Q} h(nx)\ \mathrm{d}x\right| = \frac{1}{n^d}\int_Y |h(y)|\ \mathrm{d}y$$

Donc

$$\sum_{\kappa \in J_n}\int_{Y_\kappa \cap Q} h(nx)\ \mathrm{d}x = O\left(\frac{\sharp J_n}{n^d}\right)$$

5. Ainsi, à ce stade, on a :

$$\int_Q h(nx)\ \mathrm{d}x = \left(\frac{\sharp I_n}{n^d} + O\left(\frac{\sharp J_n}{n^d}\right)\right)\int_Y h(y)\ \mathrm{d}y$$

Un petit calcul asymptotique de ce qui se passe dans la parenthèse permettrait de conclure. Pour J_n, on définit les pavés approchant Q par :

$$\underline{Q}_n \overset{\text{déf.}}{=} \prod_{i=1}^d \left[a_i + \frac{1}{n}, b_i - \frac{1}{n}\right]$$

et

$$\overline{Q}_n \overset{\text{déf.}}{=} \prod_{i=1}^d \left[a_i - \frac{1}{n}, b_i + \frac{1}{n}\right]$$

Notre choix de n permet alors d'affirmer que si $\kappa \in J_n$ alors $Y_\kappa \subset \overline{Q}_n \backslash \underline{Q}_n$, si bien que

$$\lambda_d\left(\bigsqcup_{\kappa \in J_n} Y_\kappa\right) \leqslant \lambda_d\left(\overline{Q}_n \backslash \underline{Q}_n\right)$$

Et donc

$$\lambda_d\left(\bigsqcup_{\kappa \in J_n} Y_\kappa\right) = O\left(\frac{1}{n}\right)$$

Or, par définition d'une mesure

$$\lambda_d\left(\bigsqcup_{\kappa \in J_n} Y_\kappa\right) = \sum_{\kappa \in J_n} \lambda_d(Y_\kappa)$$

On a alors montré que

$$\frac{\sharp J_n}{n^d} = O\left(\frac{1}{n}\right)$$

D'autre part, pour I_n, la même décomposition de Q en fonction de I_n et J_n utilisée pour le calcul de notre intégrale donne en termes de mesures :

$$\lambda_d(Q) = \sum_{\kappa \in I_n} \lambda_d(Y_\kappa) + \sum_{\kappa \in J_n} \lambda_d(Y_\kappa \cap Q)$$

D'où

$$\lambda_d(Q) = \frac{\sharp I_n}{n^d} + O\left(\frac{1}{n}\right)$$

Par conséquent, on a alors

$$\int_Q h(nx)\ \mathrm{d}x = \lambda_d(Q)\int_Y h(y)\ \mathrm{d}y + O\left(\frac{1}{n}\right)$$

□

Ce cas particulier permet alors de montrer le cas général par densité.

Démonstration : Soit $g \in L^{p'}(\mathbb{R}^d)$, nulle en dehors d'un pavé Q. Soit $\varepsilon > 0$. Par la propriété de densité, il existe une application φ^ε en escalier telle que

$$\|g - \varphi^\varepsilon\|_{L^{p'}} \leqslant \varepsilon$$

Puisque $\mathbf{1}_Q\varphi^\varepsilon$ est elle encore en escalier, et vérifie la même inégalité, on suppose que φ^ε est nulle en dehors de Q.

1. Montrons que l'intégrale de l'énoncé est bien définie. Pour cela, constatons que par un changement de variables, pour tout $n \in \mathbb{N}$, $x \longmapsto f(nx)$ est un élément de $L^p_\sharp(Y) \subset L^p(Q)$. De plus, $g \in L^{p'}(Q)$. Puisque p et p' sont conjugués, il suit par inégalité de Hölder que

$$\forall n \in \mathbb{N}, x \longmapsto f(nx)g(x) \in L^1(Q)$$

2. On décompose alors la différence que l'on a étudié en faisant intervenir φ^ε sur lequel on pourra appliquer le lemme :

$$\begin{aligned}
& \left|\int_Q f(nx)g(x)\ \mathrm{d}x - \left(\int_Y f\right)\left(\int_Q g\right)\right| \\
\leqslant\ & \left|\int_Q f(nx)g(x)\ \mathrm{d}x - \int_Q f(nx)\varphi^\varepsilon(x)\ \mathrm{d}x\right| \\
+\ & \left|\int_Q f(nx)\varphi^\varepsilon(x)\ \mathrm{d}x - \left(\int_Y f\right)\left(\int_Q \varphi^\varepsilon\right)\right| \\
+\ & \left|\left(\int_Y f\right)\left(\int_Q g\right) - \left(\int_Y f\right)\left(\int_Q \varphi^\varepsilon\right)\right|
\end{aligned}$$

3. Pour le premier terme, l'inégalité de Hölder donne pour tout $n \in \mathbb{N}$:

$$\left|\int_Q f(nx)(g(x)-\varphi^\varepsilon(x))\,\mathrm{d}x\right| \leqslant \|f(\cdot n)\|_{L^p}\|g-\varphi^\varepsilon\|_{L^{p'}}$$

Or, si on note $h=|f|^p$, alors $h\in L^1_\#(Y)$ et par le lemme, $\|h(n\cdot)\|_{L^1}$ converge, donc est bornée. Or,

$$\|h(n\cdot)\|_{L^1}=\|f(n\cdot)\|_{L^p}^p$$

Donc $(\|f(n\cdot)\|_{L^p}^p)_{n\in\mathbb{N}}$ est bornée, par une constante $C_1>0$ et donc pour tout $n\in\mathbb{N}$:

$$\left|\int_Q f(nx)(g(x)-\varphi^\varepsilon(x))\,\mathrm{d}x\right| \leqslant C_1\varepsilon$$

4. Pour le dernier terme, c'est encore l'inégalité de Hölder qui nous fournit l'existence d'une constante $C_2>0$ (puisque Q est compact) telle que

$$\left|\left(\int_Y f\right)\left(\int_Q g-\varphi^\varepsilon\right)\right| \leqslant C_2\varepsilon$$

5. Enfin, pour le deuxième terme, le lemme appliqué à chaque pavé intervenant dans la décomposition de φ^ε donne alors pour n assez grand :

$$\left|\int_Q f(nx)\varphi^\varepsilon(x)\,\mathrm{d}x-\left(\int_Y f\right)\left(\int_Q \varphi^\varepsilon\right)\right| \leqslant \varepsilon$$

6. Il ne reste plus qu'à faire le bilan : pour n assez grand, on a l'existence d'une constante $C>0$ telle que

$$\left|\int_Q f(nx)g(x)\,\mathrm{d}x-\left(\int_Y f\right)\left(\int_Q g\right)\right| \leqslant C\varepsilon$$

D'où la limite annoncée dans le théorème. □

3 Théorème de Kolmogorov-Khintchine

Référence : Jean-Yves Ouvrard, *Probabilités 2*
Supposons acquise la loi forte des grands nombres dans ce cadre :

Proposition 3.1 : Loi forte des grands nombres

Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes L^2. On suppose que :

(i) La suite $(\mathbb{E}[X_n])_n$ converge vers $m\in\mathbb{R}$;

(ii) La série $\sum_j \frac{1}{j^2}\sigma^2_{X_j}$ est convergente.

Alors la moyenne empirique $\bar{X}_n \overset{\text{déf.}}{=} \frac{1}{n}\sum_{j=1}^n X_j$ converge vers m au sens L^2 et $\mathbb{P}$-presque sûrement :

$$\bar{X}_n \xrightarrow[n\to+\infty]{\mathbb{P}\text{-ps},L^2} m$$

Montrons l'équivalence suivante à l'aide de ce résultat.

Théorème 3.1 : Kolmogorov - Khintchine

Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées définies sur $(\Omega,\mathcal{F},\mathbb{P})$. Les assertions suivantes sont équivalentes :

(i) Il existe $c\in\mathbb{R}$ tel que

$$\bar{X}_n \overset{\text{déf.}}{=} \frac{1}{n}\sum_{j=1}^n X_j \xrightarrow[n\to+\infty]{\mathbb{P}\text{-ps}} c$$

(ii) La variable aléatoire X_1 est intégrable : $X_1\in L^1$.

Dans ce cas, on a alors $c=\mathbb{E}[X_1]$.

Montrons d'abord une première étape de ce théorème au sein d'un lemme.

Lemme 3.1 : Être L^1 et presque sûrement $o(n)$

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées . Alors les assertions suivantes sont équivalentes :

(i) $\left(\frac{X_n}{n}\right)_{n\in\mathbb{N}^*}$ converge $\mathbb{P}$-presque sûrement vers 0 ;

(ii) $X_1 \in L^1$

Démonstration du lemme : Le plan est le suivant.

1. On montre l'encadrement suivant, pour tout $\varepsilon > 0$:

$$\varepsilon \sum_{n=0}^{+\infty} \mathbb{P}(|X_1| > (n+1)\varepsilon) \leqslant \mathbb{E}[|X_1|] \leqslant \varepsilon \sum_{n=0}^{+\infty} \mathbb{P}(|X_1| > n\varepsilon)$$

2. Puis, on montre l'équivalence intermédiaire suivante à l'aide des lemmes de Borel-Cantelli

$$\mathbb{E}[|X_1|] < +\infty \iff \mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > n\}\right) = 0$$

3. On conclut sur l'équivalence du lemme, grâce à ces deux résultats conjugués une fois de plus avec les lemmes de Borel-Cantelli.

1. Par le théorème de Fubini, on dispose de l'égalité

$$\mathbb{E}[|X_1|] = \int_0^{+\infty} \mathbb{P}(|X_1| > x)\ \mathrm{d}x \in \mathbb{R}_+ \cup \{+\infty\}$$

Soit $\varepsilon > 0$. Par le théorème de Fubini-Tonelli :

$$\mathbb{E}[|X_1|] = \sum_{n=0}^{+\infty} \int_{n\varepsilon}^{(n+1)\varepsilon} \mathbb{P}(|X_1| > x)\ \mathrm{d}x$$

Ainsi, par décroissance de $x \longmapsto \mathbb{P}(|X_1| > x)$, on dispose de l'encadrement suivant, pour tout $\varepsilon > 0$:

$$\varepsilon \sum_{n=0}^{+\infty} \mathbb{P}(|X_1| > (n+1)\varepsilon) \leqslant \mathbb{E}[|X_1|] \leqslant \varepsilon \sum_{n=0}^{+\infty} \mathbb{P}(|X_1| > n\varepsilon)$$

2. Montrons que

$$\mathbb{E}[|X_1|] < +\infty \iff \mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > n\}\right) = 0$$

Supposons que $X \in L^1$. Alors pour $\varepsilon = 1$ de l'encadrement de **1.**, on a

$$\sum_{n\geqslant 1} \mathbb{P}(|X_1| > n) < +\infty$$

Puisque les $(X_n)_n$ sont identiquement distribuées, il suit que

$$\sum_{n\geqslant 1} \mathbb{P}(|X_n| > n) < +\infty$$

Ainsi, par le lemme de Borel-Cantelli 1, valable sans supposer d'indépendance :

$$\mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > n\}\right) = 0$$

Réciproquement, si

$$\mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > n\}\right) \neq 0$$

alors

$$\sum_{n\geqslant 1} \mathbb{P}(|X_n| > n) = +\infty$$

Le lemme de Borel-Cantelli 2, puisque les $(X_n)_n$ sont indépendantes, donne alors

$$\mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > n\}\right) = 1$$

et l'encadrement donne $\mathbb{E}[|X_1|] = +\infty$. On vient alors de montrer l'équivalence :

$$\mathbb{E}[|X_1|] < +\infty \iff \mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > n\}\right) = 0$$

3. Concluons alors en l'équivalence de l'énoncé. Si

$$\mathbb{P}\left(\frac{X_n}{n} \xrightarrow[n\to+\infty]{} 0\right) = 1$$

alors en particulier

$$\mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > n\}\right) = 0$$

donc $X_1 \in L^1$ par l'équivalence **2.**. Réciproquement, si $X_1 \in L^1$ alors par l'encadrement

$$\forall \varepsilon > 0, \sum_{n=0}^{+\infty} \mathbb{P}(|X_1| > (n+1)\varepsilon) < +\infty$$

Puisque les $(X_n)_n$ sont identiquement distribuées :

$$\forall \varepsilon > 0, \sum_{n=0}^{+\infty} \mathbb{P}(|X_n| > (n+1)\varepsilon) < +\infty$$

Par le lemme de Borel-Cantelli 1 :

$$\forall \varepsilon > 0, \mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > (n+1)\varepsilon\}\right) = 0$$

Donc, par dénombrabilité de $\mathbb{Q}$:

$$\mathbb{P}\left(\bigcup_{\varepsilon\in\mathbb{Q}_+^*}\limsup_{n\to+\infty}\{|X_n| > (n+1)\varepsilon\}\right) = 0$$

Puisque

$$\left\{\frac{X_n}{n} \xrightarrow[n\to+\infty]{} 0\right\} \subset \bigcap_{\varepsilon\in\mathbb{Q}_+^*}\liminf_{n\to+\infty}\left\{\frac{X_n}{n+1} \leqslant \varepsilon\right\}$$

Il suit que

$$\mathbb{P}\left(\left\{\frac{X_n}{n} \xrightarrow[n\to+\infty]{} 0\right\}\right) \geqslant 1$$

ce qui conclut en la convergence $\mathbb{P}$-presque sûre de $\left(\frac{X_n}{n}\right)_{n\in\mathbb{N}^*}$ vers 0. □

Montrons le théorème de Kolmogorov-Khintchine.

Démonstration : **[$\Longrightarrow$]** Si la moyenne empirique converge vers $c \in \mathbb{R}$ alors

$$\frac{X_n}{n} = \bar{X}_n - \frac{n-1}{n}\bar{X}_n$$

converge $\mathbb{P}$-presque sûrement vers 0. Par la caractérisation montrée dans le lemme, $X_1 \in L^1$.

[$\Longleftarrow$] 1. Supposons que $X_1 \in L^1$. Pour montrer la convergence presque sûre de la moyenne empirique, on est tenté d'appliquer la loi forte des grands nombres. Or, on a *a priori* aucune information quand au caractère L^2 des variables X_n. On va tronquer ces variables pour pourvoir montrer les différentes hypothèses. Constatons d'abord que par le lemme (et l'équivalence intermédiaire montrée au sein de ce lemme) :

$$\mathbb{P}\left(\limsup_{n\to+\infty}\{|X_n| > n\}\right) = 0$$

La troncature que l'on considère est alors la suivante :

$$\tilde{X}_n \overset{\text{déf.}}{=} \mathbf{1}_{\{|X_n|\leqslant n\}}X_n$$

Notons $\tilde{S}_n$ la somme des $\tilde{X}_n$, et montrons que

$$\mathbb{P}\left(\left(\frac{S_n}{n}\right)_n \text{ CV}\right) = \mathbb{P}\left(\left(\frac{\tilde{S}_n}{n}\right)_n \text{ CV}\right)$$

Pour cela, observons que

$$\mathbb{P}\left(\liminf_{n\to+\infty}\{X_n = \tilde{X}_n\}\right) = 1$$

Ainsi,

$$\mathbb{P}\left(\left(\frac{S_n}{n}\right)_n \text{ CV}\right) = \mathbb{P}\left(\left\{\left(\frac{S_n}{n}\right)_n \text{ CV}\right\} \cap \left\{\liminf_{n\to+\infty}\{X_n = \tilde{X}_n\}\right\}\right)$$

Par égalité asymptotique

$$\mathbb{P}\left(\left(\frac{S_n}{n}\right)_n \text{ CV}\right) = \mathbb{P}\left(\left\{\left(\frac{\tilde{S}_n}{n}\right)_n \text{ CV}\right\} \cap \left\{\liminf_{n\to+\infty}\{X_n = \tilde{X}_n\}\right\}\right)$$

Et donc

$$\mathbb{P}\left(\left(\frac{S_n}{n}\right)_n \text{ CV}\right) = \mathbb{P}\left(\left(\frac{\tilde{S}_n}{n}\right)_n \text{ CV}\right)$$

Ainsi, au lieu d'essayer d'appliquer les hypothèses de la loi des grands nombres à X_n, on peut se contenter de montrer ce résultat sur les $\tilde{X}_n$.

2. Puisque

$$\mathbb{E}\left[\tilde{X}_n\right] = \mathbb{E}\left[\tilde{X}_1\mathbf{1}_{\{|X_1|\leqslant n\}}\right]$$

Il suit, par théorème de convergence dominée, puisque $X_1 \in L^1$ que

$$\mathbb{E}\left[\tilde{X}_n\right] \xrightarrow[n\to+\infty]{} \mathbb{E}[X_1]$$

donc la suite $\left(\mathbb{E}\left[\tilde{X}_n\right]\right)$ est bien convergente.

3. Montrons que $\tilde{X}_n \in L^2$ et la convergence de la série $\sum_j \frac{\sigma^2_{\tilde{X}_j}}{j^2}$. D'une part, constatons que la troncature donne alors un caractère borné à $\tilde{X}_n^2$. Il suit que $\tilde{X}_n \in L^2$. Concluons sur la série. On dispose de l'encadrement suivant :

$$0 \leqslant \sigma^2_{\tilde{X}_n} \leqslant \mathbb{E}\left[\tilde{X}_n^2\right]$$

Ainsi, par théorème de Beppo Lévi :

$$0 \leqslant \sum_{n=1}^{+\infty}\frac{1}{n^2}\sigma^2_{\tilde{X}_n} \leqslant \mathbb{E}\left[\sum_{n=1}^{+\infty}X_1^2\mathbf{1}_{\{|X_1|\leqslant n\}}\right]$$

Attardons-nous au dernier terme, et écrivons

$$X_1^2 \mathbf{1}_{\{|X_1|\leqslant n\}} = X_1^2 \sum_{m=1}^{n} \mathbf{1}_{\{m-1<|X_1|\leqslant m\}}$$

L'intérêt d'avoir tronqué X_n réside alors dans cette somme qui est finie. On a alors, dans $\mathbb{R}_+ \cup \{+\infty\}$:

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{\{|X_1|\leqslant n\}} = \sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \sum_{m=1}^{n} \mathbf{1}_{\{m-1<|X_1|\leqslant m\}}$$

On intervertit, par Fubini-Tonelli la deuxième somme, et on garde de côté un $|X_1|$:

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{\{|X_1|\leqslant n\}} = \sum_{m=1}^{+\infty} |X_1| \sum_{n=m}^{+\infty} \frac{|X_1|}{n^2} \mathbf{1}_{\{m-1<|X_1|\leqslant m\}}$$

Alors, par l'indicatrice dans la somme intérieure :

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{\{|X_1|\leqslant n\}} \leqslant \sum_{m=1}^{+\infty} m|X_1| \mathbf{1}_{\{m-1<|X_1|\leqslant m\}} \sum_{n=m}^{+\infty} \frac{1}{n^2}$$

Or, on peut évaluer le reste de la série de Riemann par une comparaison avec une intégrale :

$$\sum_{n=m}^{+\infty} \frac{1}{n^2} = \frac{1}{m^2} + \sum_{n=m}^{+\infty} \frac{1}{(n+1)^2}$$

Donc

$$\sum_{n=m}^{+\infty} \frac{1}{n^2} \leqslant \frac{1}{m^2} + \sum_{n=m}^{+\infty} \int_n^{n+1} \frac{\mathrm{d}t}{t^2}$$

On utilise la relation de Chasles, et un calcul explicite donne alors

$$\sum_{n=m}^{+\infty} \frac{1}{n^2} \leqslant \frac{2}{m^2} + \frac{1}{m} \leqslant \frac{2}{m}$$

Ainsi, on a :

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{\{|X_1|\leqslant n\}} \leqslant 2 \sum_{m=1}^{+\infty} |X_1| \mathbf{1}_{\{m-1<|X_1|\leqslant m\}} = 2|X_1|$$

avec $X_1 \in L^1$. Ainsi,

$$\mathbb{E}\left[\sum_{n=1}^{+\infty} \frac{1}{n^2} X_1^2 \mathbf{1}_{\{|X_1|\leqslant n\}}\right] < +\infty$$

et donc on a montré que

$$\sum_{n=1}^{+\infty} \frac{\sigma^2_{\tilde{X}_n}}{n^2} < +\infty$$

Par la loi forte des grands nombres, on a alors

$$\frac{\tilde{S}_n}{n} \xrightarrow[n\to+\infty]{\mathbb{P}-\text{ps}} \mathbb{E}[X_1]$$

Et par **1.**, il suit que

$$\bar{X}_n \xrightarrow[n\to+\infty]{\mathbb{P}-\text{ps}} \mathbb{E}[X_1]$$

And that's a good place to stop. □

4 Théorème fondamental de la statistique

Référence : Jean-Yves Ouvrard, *Probabilités 2*
On suppose connu le théorème de Kolmogorov-Khintchine.

Théorème 4.1 : Théorème fondamental de la statistique

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F. On note F_n la *fonction de répartition empirique* :

$$\forall x \in \mathbb{R}, \forall \omega \in \Omega, F_n(x,\omega) \overset{\text{déf.}}{=} \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{\{X_j\leqslant x\}}(\omega)$$

Alors $\mathbb{P}$-presque sûrement, $F_n(\cdot,\omega)$ converge uniformément vers F sur $\mathbb{R}$:

$$\mathbb{P}\left(\lim_{n\to+\infty} \sup_{x\in\mathbb{R}} |F_n(\cdot,x) - F(x)| = 0\right) = 1$$

On utilise le lemme d'analyse réelle suivant.

Lemme 4.1 : Théorème de DINI 2 bis

Soient $(f_n)_n$ une suite de fonctions croissantes et f une autre fonction croissante définies sur $\mathbb{R}$. On suppose que ces fonctions sont à valeurs dans $[0,1]$.

(i) Soit D un sous-ensemble de $\mathbb{R}$ dénombrable et dense. Si $(f_n)_n$ converge simplement vers f sur D alors $(f_n)_n$ converge simplement vers f sur l'ensemble des points de continuité de f.

(ii) Supposons que les f_n et f sont des fonctions de répartition (*id est* ces fonctions sont de plus continues à droite et tendent vers 0 en $-\infty$ et 1 en $+\infty$). On note D la réunion de $\mathbb{Q}$ et des points de discontinuité de f. Si $(f_n)_n$ converge simplement vers f sur D et si

$$\forall x \in D, f_n(x^-) \xrightarrow[n\to+\infty]{} f(x^-)$$

alors $(f_n)_n$ converge uniformément vers f sur $\mathbb{R}$.

Remarque : Dans le cadre de fonction de répartition, $f(x^-)$ signifie simplement $\mathbb{P}(X < x)$.

Démonstration du lemme : **(i)** Soit $x \in \mathbb{R}$ un point de continuité de f. Alors il existe $\eta > 0$ tel que pour tout $x' \in \mathbb{R}$:

$$|x - x'| \leqslant \eta \Longrightarrow |f(x) - f(x')| \leqslant \varepsilon$$

Soient $y, y' \in D$ tels que

$$x - \eta < y < x < y' < x + \eta$$

Alors par convergence simple sur D :

$$f(y) = \lim_{n\to+\infty} f_n(y)$$

Par croissance des f_n :

$$f(y) \leqslant \liminf_{n\to+\infty} f_n(x) \leqslant \limsup_{n\to+\infty} f_n(x)$$

Par croissance, encore une fois :

$$f(y) \leqslant \limsup_{n\to+\infty} f_n(x) \leqslant \lim_{n\to+\infty} f_n(y')$$

D'où par convergence simple sur D :

$$f(y) \leqslant \limsup_{n\to+\infty} f_n(x) \leqslant f(y')$$

Ainsi, on a montré

$$\left|\limsup_{n\to+\infty} f_n(x) - \liminf_{n\to+\infty} f_n(x)\right| \leqslant |f(y) - f(y')|$$

Par le choix de y, y' et continuité de f, on a montré que pour tout $\varepsilon > 0$

$$\left|\limsup_{n\to+\infty} f_n(x) - \liminf_{n\to+\infty} f_n(x)\right| \leqslant \varepsilon$$

On a aussi montré les inégalités

$$f(y) - f(x) \leqslant \limsup_{n\to+\infty} f_n(x) - f(x) \leqslant f(y') - f(x)$$

Donc

$$\left|\limsup_{n\to+\infty} f_n(x) - f(x)\right| \leqslant \varepsilon$$

Bilan, on a donc bien

$$f(x) = \limsup_{n\to+\infty} f_n(x) = \liminf_{n\to+\infty} f_n(x)$$

donc $(f_n)_n$ converge bien simplement en tous points de continuité de f.

(ii) L'idée est ici de considérer une subidivision de l'espace d'arrivée $[0,1]$. Soit $k \in \mathbb{N}^*$. Pour $j \in [\![1, k]\!]$, on note

$$x_{j,k} \overset{\text{déf.}}{=} \sup\left\{x \in \mathbb{R}, f(x^-) \leqslant \frac{j}{k} \leqslant f(x)\right\}$$

et $x_{0,k} = -\infty$. Puisque f est une fonction de répartition, $x_{k,k} = +\infty$.

• Pour $k \in \mathbb{N}^*$, la suite finie de $\bar{\mathbb{R}}$ $(x_{j,k})_j$ est croissante. En effet, si $x \in \mathbb{R}$ vérifie

$$f(x^-) \leqslant \frac{j}{k} \leqslant f(x)$$

et $y \in \mathbb{R}$ vérifie

$$f(y^-) \leqslant \frac{j+1}{k} \leqslant f(y)$$

Alors

$$f(y) - f(x^-) \leqslant \frac{1}{k} \leqslant f(y^-) - f(x)$$

D'où $f(x) \leqslant f(y^-) \leqslant f(y)$. Ainsi, par croissance de f : $x \leqslant y \leqslant x_{j+1,k}$. D'où

$$x_{j,k} \leqslant x_{j+1,k}$$

• On dispose ainsi d'une décomposition de $\mathbb{R}$:

$$\mathbb{R} = \bigsqcup_{j=0}^{k-1}]x_{j,k}, x_{j+1,k}]$$

avec des intervalles éventuellement vides. On note alors

$$\Delta_n^1(k) \overset{\text{déf.}}{=} \max_{j\in[\![0,k-1]\!]} |f_n(x_{j,k}) - f(x_{j,k})|$$

et

$$\Delta_n^2(k) \overset{\text{déf.}}{=} \max_{j\in[\![0,k-1]\!]} \left|f_n(x_{j,k}^-) - f(x_{j,k}^-)\right|$$

Et on cherche à étudier

$$\Delta_n \overset{\text{déf.}}{=} \sup_{x\in\mathbb{R}} |f_n(x) - f(x)|$$

Alors

$$\Delta_n \leqslant \max\left\{\Delta_n^1(k), \Delta_n^2(k)\right\} + \frac{1}{k}$$

En effet, avec la convention $\sup_\varnothing = -\infty$, on a :

$$\Delta_n = \max_{j\in[\![1,k-1]\!]} \sup_{x\in]x_{j,k},x_{j+1,k}]} |f_n(x) - f(x)|$$

Pour $x \in]x_{j,k}, x_{j+1,k}]$, on a alors par croissance

$$f(x_{j,k}) \leqslant f(x) \leqslant f(x_{j+1,k}^-)$$

et

$$f_n(x_{j,k}) \leqslant f_n(x) \leqslant f_n(x_{j+1,k}^-)$$

Et par définition des $x_{j,k}$:

$$0 \leqslant f\left(x_{j+1,k}^-\right) - f\left(x_{j,k}\right) \leqslant \frac{1}{k}$$

Ainsi,

$$f_n(x) - f(x) \leqslant f_n\left(x_{j+1,k}^-\right) - f(x_{j,k})$$

Pour faire apparaître $\Delta_n^{1,2}(k)$ et $\frac{1}{k}$, on fait alors

$$\begin{aligned} f_n(x) - f(x) &\leqslant f_n\left(x_{j+1,k}^-\right) - f(x_{j+1,k}^-) \\ &+ f\left(x_{j+1,k}^-\right) - f(x_{j,k}) \end{aligned}$$

D'où

$$f_n(x) - f(x) \leqslant f_n\left(x_{j+1,k}^-\right) - f\left(x_{j+1,k}^-\right) + \frac{1}{k} \leqslant \Delta^2(k) + \frac{1}{k}$$

Même type de manipulation pour l'autre inégalité :

$$f_n(x) - f(x) \geqslant f_n\left(x_{j,k}\right) - f\left(x_{j+1,k}^-\right)$$

Ce qui donne cette fois-ci

$$f_n(x) - f(x) \geqslant f_n(x_{j,k}) - f(x_{j+1,k}^-) - \frac{1}{k} \geqslant -\Delta^1(k) - \frac{1}{k}$$

On a donc bien montré que

$$\Delta_n \leqslant \max\{\Delta_n^1(k), \Delta_n^2(k)\} + \frac{1}{k}$$

• Il ne reste plus qu'à montrer que chaque $\Delta_n^s(k)$ tend vers 0, c'est-à-dire montrer que

$$f_n(x_{j,k}) \xrightarrow[n\to+\infty]{} f(x_{j,k})$$

et

$$f_n\left(x_{j,k}^-\right) \xrightarrow[n\to+\infty]{} f\left(x_{j,k}^-\right)$$

Si $x_{j,k}$ est un point de continuité de f, c'est **(i)** qui permet de conclure sur ces deux limites. Sinon, $x_{j,k}$ est un point de discontinuité de f, donc appartient à D. Par hypothèse, $(f_n)_n$ converge simplement vers f sur D, donc encore une fois, ces limites sont vérifiées. Ainsi, pour tout $k \in \mathbb{N}^*$:

$$0 \leqslant \Delta_n \leqslant \max\{\Delta_n^1, \Delta_n^2\} + \frac{1}{k}$$

donne, pour tout $k \in \mathbb{N}^*$:

$$0 \leqslant \limsup_{n\to+\infty} \Delta_n \leqslant \frac{1}{k}$$

D'où

$$\lim_{n\to+\infty} \Delta_n = 0$$

Ainsi, $(f_n)_n$ converge bien uniformément vers f sur $\mathbb{R}$ tout entier. □

On peut alors montrer le théorème fondamental de la statistique.

Démonstration : **1.** Commençons par constater que pour tout $\omega \in \Omega$, la continuité à droite de F et $F(\cdot,\omega)$ donne

$$\sup_{x\in\mathbb{R}} |F_n(x,\cdot) - F(x)| = \sup_{x\in\mathbb{Q}} |F_n(x,\cdot) - F(x)|$$

avec $\mathbb{Q}$ dénombrable. Il suit donc que cette quantité définit bien une variable aléatoire.

2. Pour montrer une convergence presque sûre, on applique le théorème de Kolmogorov-Khintchine aux variables aléatoires $\left(\mathbf{1}_{X_j\leqslant x}\right)_{j\geqslant 1}$. Il s'agit là d'une suite de variables aléatoires indépendantes et identiquement distribuées inté-

grables, vérifiant de plus

$$\mathbb{E}\left[\mathbf{1}_{\{X_j\leqslant x\}}\right] = \mathbb{P}(X_j \leqslant x) = F(x)$$

Par le sens réciproque du théorème de Kolmogorov-Khintchine, il suit que

$$\forall x \in \mathbb{R}, F_n(x,\cdot) \xrightarrow[n\to+\infty]{\mathbb{P}-\text{ps}} F(x)$$

3. Or, de la même manière, avec les variables aléatoires $\left(\mathbf{1}_{\{X_j<x\}}\right)$, on obtient cette fois-ci

$$\forall x \in \mathbb{R}, F_n(x^-,\cdot) \xrightarrow[n\to+\infty]{\mathbb{P}-\text{ps}} F(x^-)$$

4. Si on se place dans l'esprit du lemme, on définit alors D la réunion de $\mathbb{Q}$ avec les points de discontinuité de F. Par les deux $\mathbb{P}$-presque sûres convergences, il existe pour tout $x \in D$ deux ensembles N_x^1 et N_x^2 négligeables tels que

$$\forall \omega \in \Omega\backslash N_x^1, F_n(x,\omega) \xrightarrow[n\to+\infty]{} F(x)$$

et

$$\forall \omega \in \Omega\backslash N_x^2, F_n(x^-,\omega) \xrightarrow[n\to+\infty]{} F(x^-)$$

Notons alors

$$N \overset{\text{déf.}}{=} \left[\bigcup_{x\in D} N_x^1\right] \cup \left[\bigcup_{x\in D} N_x^2\right]$$

Puisque D est dénombrable, N est négligeable. De plus, pour tout $\omega \in \Omega\backslash N$ et $x \in D$, $F_n(x,\omega)$ et $F_n(x^-,\omega)$ convergent vers $F(x)$ et $F(x^-)$. Par le lemme, il suit alors que pour tout $\omega \in \Omega\backslash N$,

$$\lim_{n\to+\infty} \sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)| = 0$$

C'est-à-dire que $F_n(\cdot,\omega)$ converge $\mathbb{P}$-presque partout uniformément vers F. □

5 Caractère métrique et complet de la convergence en probabilité

Référence : Jean-Yves Ouvrard, *Probabilités 2*

Proposition 5.1 : Distance sur L^0

On note L^0 l'ensemble des variables aléatoires $\mathbb{P}$-presque sûrement finies, définies presque partout. On définit les application suivantes, pour tout $X, Y \in L^0$:

$$d(X,Y) \overset{\text{déf.}}{=} \mathbb{E}\left[\frac{|X-Y|}{1+|X-Y|}\right]$$

et

$$\delta(X,Y) \overset{\text{déf.}}{=} \mathbb{E}\left[\min(1,|X-Y|)\right]$$

Alors :

1. d et δ définissent des distances sur L^0 ;
2. Ces distances sont équivalentes ;
3. La convergence pour l'une de ces deux distances est équivalente à la convergence en probabilité ;
4. L'espace métrique (L^0,d) est complet.

Démonstration : **1.** d et δ sont des distances. Pour cela, constatons la symétrie et l'homogénéité. Reste à montrer l'inégalité triangulaire. Pour d, il s'agit de la croissance de $x \longmapsto \frac{x}{1+x}$ qui permet d'écrire, lorsque $X, Y, Z \in L^0$:

$$d(X,Z) \leqslant \mathbb{E}\left[\frac{|X-Y|+|Y-Z|}{1+|X-Y|+|Y-Z|}\right]$$

d'où

$$d(X,Z) \leqslant d(X,Y) + d(Y,Z)$$

Pour δ, la positivité des objets manipulés permet de conclure par inégalité triangulaire pour la valeur absolue.

2. Constatons que pour tout $x \geqslant 0$:

$$\frac{1}{2}\min\{1,x\} \leqslant \frac{x}{1+x} \leqslant \min\{1,X\}$$

donc les distances d et δ sont bien équivalentes.

3. Soit $\varepsilon > 0$. D'une part, pour $X, Y \in L^0$:

$$\begin{aligned} d(X,Y) = & \quad \mathbb{E}\left[\frac{|X-Y|}{1+|X-Y|}\mathbf{1}_{\{|X-Y|\leqslant\varepsilon\}}\right] \\ & + \quad \mathbb{E}\left[\frac{|X-Y|}{1+|X-Y|}\mathbf{1}_{\{|X-Y|>\varepsilon\}}\right] \end{aligned}$$

Ainsi, par croissance de $x \longmapsto \frac{x}{1+x}$ et parce qu'elle est bornée par 1 :

$$d(X,Y) \leqslant \frac{\varepsilon}{1+\varepsilon}\mathbb{E}\left[\mathbf{1}_{\{|X-Y|\leqslant\varepsilon\}}\right] + \mathbb{E}\left[\mathbf{1}_{\{|X-Y|>\varepsilon\}}\right]$$

D'où

$$d(X,Y) \leqslant \varepsilon + \mathbb{P}(|X-Y| > \varepsilon)$$

De même, on a

$$d(X,Y) \geqslant \frac{\varepsilon}{1+\varepsilon}\mathbb{P}(|X-Y| > \varepsilon)$$

D'où, pour tout $\varepsilon > 0$,

$$\frac{\varepsilon}{1+\varepsilon}\mathbb{P}(|X-Y| > \varepsilon) \leqslant d(X,Y) \leqslant \varepsilon + \mathbb{P}(|X-Y| > \varepsilon)$$

Par conséquent, la convergence pour d est bien équivalente pour la convergence en probabilité.

4. Deux manières de la montrer : soit par la distance d, soit en utilisant la convergence en probabilité. Montrons ici que (L^0, d) est complet. Soit $(X_n)_n$ une suite de Cauchy pour d. Il existe une extraction ϕ telle que

$$\forall k \in \mathbb{N}, d\left(X_{\phi(k)}, X_{\phi(k+1)}\right) \leqslant \frac{1}{2^k}$$

Si bien que

$$\sum_{k=0}^{+\infty} d\left(X_{\phi(k)}, X_{\phi(k+1)}\right) < +\infty$$

Par théorème de Beppo Lévi, on a alors

$$\mathbb{E}\left[\sum_{k=0}^{+\infty} \frac{\left|X_{\phi(k+1)} - X_{\phi(k)}\right|}{1+\left|X_{\phi(k+1)} - X_{\phi(k)}\right|}\right] < +\infty$$

En particulier,

$$\sum_{k=0}^{+\infty} \frac{\left|X_{\phi(k+1)} - X_{\phi(k)}\right|}{1+\left|X_{\phi(k+1)} - X_{\phi(k)}\right|} < +\infty \ \mathbb{P}\text{-ps}$$

Ainsi,

$$\left|X_{\phi(k+1)} - X_{\phi(k)}\right| \xrightarrow[k\to+\infty]{\mathbb{P}-\text{ps}} 0$$

Enfin, rappelons que

$$\forall x \geqslant 0, \min\{1,x\} \leqslant \frac{2x}{1+x}$$

Ces trois derniers résultats permettent alors de conclure que

$$\sum_{k=0}^{+\infty} |X_{\phi(k+1)} - X_{\phi(k)}| < +\infty$$

donc $(X_{\phi(k))_k}$ converge $\mathbb{P}$-presque sûrement. Par théorème de convergence dominée (majoration par la constante 1), il suit que $(X_{\phi(k)})_k$ converge pour d. Ainsi, $(X_n)_n$ est une suite de Cauchy qui admet une sous-suite convergente, don converge. Il suit que (L^0, d) est complet.

4*. On peut aussi montrer ce résultat en termes de convergence en probabilité. Soit $(X_n)_n$ une suite de Cauchy pour la convergence en probabilités : pour tout $\varepsilon > 0$, pour tout $p \in \mathbb{N}$,

$$\mathbb{P}\left(|X_{n+p} - X_n| > \varepsilon\right) \xrightarrow[n\to+\infty]{} 0$$

On peut extraire une famille $(X_{\phi(k)})_k$ telle que

$$\forall k \in \mathbb{N}, \mathbb{P}\left(\left|X_{\phi(k+1)} - X_{\phi(k)}\right| > \frac{1}{2^k}\right) \leqslant \frac{1}{3^k}$$

Alors

$$\sum_{k=0}^{+\infty} \mathbb{P}\left(\left|X_{\phi(k+1)} - X_{\phi(k)}\right| > \frac{1}{2^k}\right) < +\infty$$

Par une application du lemme de Borel Cantelli 1, il suit que $(X_{\phi(k)})_k$ converge $\mathbb{P}$-presque sûrement. Notons X cette limite presque sûre, et montrons alors que $(X_n)_n$ converge en probabilité vers X. Pour cela, si $\varepsilon > 0$, on remarque que pour k assez grand :

$$\begin{aligned} \mathbb{P}(|X_n - X| > \varepsilon) \leqslant & \quad \mathbb{P}\left(\left|X_n - X_{\phi(k)}\right| > \tfrac{\varepsilon}{2}\right) \\ & + \quad \mathbb{P}\left(\left|X - X_{\phi(k)}\right| > \tfrac{\varepsilon}{2}\right) \end{aligned}$$

Donc $(X_n)_n$ converge bien en probabilité vers X. □

6 Inégalité de Hoeffding

Référence : Jean-Yves Ouvrard, *Probabilités 2*

Proposition 6.1 : Inégalité de Hoeffding

Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes telles que

* Les variables X_n sont presque sûrement bornées :

$$\forall n \in \mathbb{N}, \exists c_n > 0, |X_n| \leqslant c_n \ \mathbb{P}\text{-ps}$$

* Les variables X_n sont centrées : $\mathbb{E}[X_n] = 0$.

On note $S_n = \sum_{j=1}^n X_j$ la somme des variables aléatoires X_n. Alors

$$\forall \varepsilon > 0, \mathbb{P}(|S_n| > \varepsilon) \leqslant 2\exp\left(\frac{-\varepsilon^2}{2\sum_{j=1}^n c_j^2}\right)$$

Commençons par un petit lemme qui utilise la convexité de l'exponentielle.

Lemme 6.1 : Transformée de Laplace d'une VAR bornée ps

Soit X une variable aléatoire réelle $\mathbb{P}$-presque sûrement bornée par 1, d'espérance nulle. Alors

$$\forall t \in \mathbb{R}, \mathbb{E}\left[e^{tX}\right] \leqslant \exp\left(\frac{t^2}{2}\right)$$

Démonstration du lemme : Pour $t \in \mathbb{R}$ et $x \in [-1,1]$, on a l'égalité suivante :

$$tx = \frac{t}{2}(1-x) - \frac{t}{2}(1+x)$$

avec $\frac{1-x}{2}, \frac{1+x}{2} \in [0,1]$ de somme 1. Par convexité de l'exponentielle,

$$e^{tx} \leqslant \frac{1-x}{2}e^t + \frac{1+x}{2}e^{-t}$$

Puisque X est bornée $\mathbb{P}$-presque sûrement, $X \in L^1$ et on obtient

$$\mathbb{E}\left[e^{tX}\right] \leqslant \mathbb{E}\left[\frac{1-X}{2}\right]e^t + \mathbb{E}\left[\frac{1+X}{2}\right]e^{-t}$$

Puisque X est centrée, il suit que

$$\mathbb{E}\left[e^{tX}\right] \leqslant \cosh(t)$$

Or, en comparant les développements en séries entières de $\cosh$ et de $\exp\left(\frac{t^2}{2}\right)$, puisque $2^n n! \leqslant (2n)!$, il suit que

$$\mathbb{E}\left[e^{tX}\right] \leqslant \exp\left(\frac{t^2}{2}\right)$$

□

Muni de ce lemme, on peut conclure sur l'inégalité de Hoeffding.

Démonstration : 1. Pour $t' > 0$, le lemme donne, puisque $\frac{X_n}{c_n}$ est bornée par 1 presque sûrement :

$$\mathbb{E}\left[\exp\left(\frac{t'X_n}{c_n}\right)\right] \leqslant \exp\left(\frac{t'^2}{2}\right)$$

Ainsi, pour $t' = c_n t$:

$$\mathbb{E}\left[e^{tX_n}\right] \leqslant \exp\left(\frac{c_n^2 t^2}{2}\right)$$

Or,

$$\mathbb{E}\left[e^{tS_n}\right] = \mathbb{E}\left[\prod_{j=1}^n e^{tX_n}\right]$$

Par indépendance des $(X_n)_n$, on a alors

$$\mathbb{E}\left[e^{tS_n}\right] = \prod_{j=1}^{n} \mathbb{E}\left[e^{tX_n}\right]$$

Par positivité :

$$\mathbb{E}\left[e^{tS_n}\right] \leqslant \prod_{j=1}^{n} \exp\left(\frac{c_n^2 t^2}{2}\right)$$

Et donc

$$\mathbb{E}\left[e^{tS_n}\right] \leqslant \exp\left(\sum_{j=1}^{n} c_j^2 \frac{t^2}{2}\right)$$

2. Pour $\varepsilon > 0$ et $t > 0$, on a :

$$\mathbb{P}(S_n > \varepsilon) = \mathbb{P}\left(e^{tS_n} > e^{t\varepsilon}\right)$$

Par l'inégalité de Markov :

$$\mathbb{P}(S_n > \varepsilon) \leqslant e^{-t\varepsilon}\mathbb{E}\left[e^{tS_n}\right]$$

D'où l'inégalité

$$\mathbb{P}(S_n > \varepsilon) \leqslant \exp\left(\frac{t^2}{2}\sum_{j=1}^{n} c_j^2 - t\varepsilon\right)$$

3. On minimise en t ce polynôme de degré 2. Ce minimum est atteint en $\frac{\varepsilon}{\sum_{j=1}^{n} x_j^2}$ et donne finalement

$$\mathbb{P}(S_n > \varepsilon) \leqslant \exp\left(\frac{-\varepsilon^2}{2\sum_{j=1}^{n} c_j^2}\right)$$

4. On conclut en constatant que

$$\{|S_n| > \varepsilon\} = \{S_n > \varepsilon\} \sqcup \{-S_n > \varepsilon\}$$

où $-S_n$ est la somme des $-X_n$ centrées presque sûrement bornée. Ainsi, $-S_n$ vérifie l'inégalité **3.** et donc

$$\mathbb{P}(|S_n| > \varepsilon) \leqslant 2\exp\left(\frac{-\varepsilon^2}{2\sum_{j=1}^{n} c_j^2}\right)$$

D'où l'inégalité de Hoeffding. □

7 Théorème des trois séries de Kolmogorov

Référence : Jean-Yves Ouvrard, *Probabilités 2*

Théorème 7.1 : Théorème des trois séries de Kolmogorov

Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes. Pour $c > 0$, on note

$$Y_n \overset{\text{déf.}}{=} \mathbf{1}_{\{|X_n| \leqslant c\}} \cdot X_n$$

Alors les assertions suivantes sont équivalentes :

(i) La série $\sum_n X_n$ converge $\mathbb{P}$-presque sûrement ;

(ii) Les séries numériques $\sum_n \mathbb{E}[Y_n]$, $\sum_n \sigma^2_{X_n}$ et $\sum_n \mathbb{P}(|X_n| > c)$ convergent.

Pour arriver à nos fins, deux lemmes, l'un à propos du *symétrisé* d'une variable aléatoire, l'autre sur une condition suffisante de la convergence de $\sum_{n\in\mathbb{N}} \sigma^2_{X_n}$.

Lemme 7.1 : Symétrisé d'une variable aléatoire

Soit $X \in L^0$ une variable aléatoire. On définit le *symétrisé* de X comme étant la variable aléatoire X^{s} définie sur $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, \mathbb{P} \otimes \mathbb{P})$ par :

$$\forall (\omega, \omega') \in \Omega \times \Omega, X^{\mathsf{s}}(\omega, \omega') \overset{\text{déf.}}{=} X(\omega) - X(\omega')$$

Alors :

* X^{s} est une variable aléatoire sur $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, \mathbb{P} \otimes \mathbb{P})$;
* Pour $p \geqslant 1$, si $X \in L^p(\Omega)$ alors $X^{\mathsf{s}} \in L^p(\Omega^2)$;
* Si $X \in L^2$ alors $\mathbb{E}\left[X^{\mathsf{s}}\right] = 0$ et

$$\sigma^2_{X^s} = 2\sigma^2_X \ ;$$

* Si $(X_i)_{i\in I}$ est une famille quelconque de variables aléatoires indépendante alors la famille de ses symétrisés $(X_i^s)_{i\in I}$ est $\mathbb{P}\otimes\mathbb{P}$-indépendante.

Démonstration du lemme : On introduit la variable aléatoire à valeurs dans $\mathbb{R}^2$ suivante :

$$\forall \omega, \omega' \in \Omega, \hat{X}(\omega,\omega') \overset{\text{déf.}}{=} (X(\omega), X(\omega'))$$

On note $\hat{X}_1$ et $\hat{X}_2$ les lois marginales liées. Alors on constate que

$$X^s = \hat{X}_1 - \hat{X}_2 = f(\hat{X})$$

avec $f : (x,y) \longmapsto x-y$ mesurable. Vérifions alors chaque point, qui se fait de manière élémentaire, mais qui permet d'être au clair sur les notions de produit tensoriel de mesure ou de tribu.

1. Montrons que $\hat{X}$ est une variable aléatoire, pour en conclure que X^s en est une. Si A, B sont deux boréliens de $\mathbb{R}$ alors par définition de $\hat{X}$:

$$\{\hat{X} \in A\times B\} = \{X\in A\}\times\{X\in B\} \in \mathcal{A}\otimes\mathcal{A}$$

Donc $\hat{X}$ est bien une variable aléatoire sur Ω^2. Puisque f est mesurable, il suit que X^s l'est aussi. C'est aussi le cas des variables $\hat{X}_1$ et $\hat{X}_2$. Montrons qu'elles sont indépendantes. Observons que si A, B sont des boréliens de $\mathbb{R}$ alors

$$\{\hat{X}_1 \in A\}\times\{\hat{X}_2\in B\} = (\{X\in A\}\times\Omega)\times(\Omega\times\{X\in B\})$$

Ainsi, par définition de la mesure produit

$$\mathbb{P}\otimes\mathbb{P}\left(\left\{\hat{X}_1\in A\right\}\times\left\{\hat{X}_2\in B\right\}\right) = \mathbb{P}(X\in A)\mathbb{P}(X\in B)$$

Or,

$$\mathbb{P}(X\in A) = \mathbb{P}\otimes\mathbb{P}(\{X\in A\}\times\Omega)$$

Donc

$$\begin{gathered}\mathbb{P}\otimes\mathbb{P}\left(\left\{\hat{X}_1\in A\right\}\times\left\{\hat{X}_2\in B\right\}\right)\\ =\\ \mathbb{P}\otimes\mathbb{P}(\{X\in A\}\times\Omega)\cdot\mathbb{P}\otimes\mathbb{P}(\{X\in B\}\times\Omega)\end{gathered}$$

D'où

$$\begin{gathered}\mathbb{P}\otimes\mathbb{P}\left(\left\{\hat{X}_1\in A\right\}\times\left\{\hat{X}_2\in B\right\}\right)\\ =\\ \mathbb{P}\otimes\mathbb{P}\left(\hat{X}_1\in A\right)\cdot\mathbb{P}\otimes\mathbb{P}\left(\hat{X}_2\in B\right)\end{gathered}$$

Ce qui signifie exactement que $\hat{X}_1$ et $\hat{X}_2$ sont indépendantes.

2. Montrons que $X\in L^p$ si et seulement si $\hat{X}_1\in L^p(\Omega^2)$. Pour cela, le théorème de Fubini donne

$$\int_{\Omega\times\Omega}|X_1|^p\,\mathrm{d}\mathbb{P}\otimes\mathbb{P} = \int_\Omega\left[\int_\Omega\left|\hat{X}_1(\omega,\omega')\right|^p\,\mathrm{d}\mathbb{P}(\omega)\right]\,\mathrm{d}\mathbb{P}(\omega')$$

Par définition de $\hat{X}_1$:

$$\int_{\Omega\times\Omega}|X_1|^p\,\mathrm{d}\mathbb{P}\otimes\mathbb{P} = \int_\Omega\left[\int_\Omega|X(\omega)|^p\,\mathrm{d}\mathbb{P}(\omega)\right]\,\mathrm{d}\mathbb{P}(\omega')$$

L'intégrale par rapport à ω' ne contient alors qu'une constante, donc on obtient

$$\int_{\Omega\times\Omega}|X_1|^p\,\mathrm{d}\mathbb{P}\otimes\mathbb{P} = \int_\Omega|X(\omega)|^p\,\mathrm{d}\mathbb{P}(\omega)$$

Ce qui s'écrit en fait

$$\mathbb{E}\left[\left|\hat{X}_1\right|^p\right] = \mathbb{E}\left[|X|^p\right]$$

Donc X est intégrable si et seulement si $\hat{X}_1$ l'est, idem avec $\hat{X}_2$. Ainsi, si X est intégrable, alors X^s est différence de deux variables L^p, donc est dans L^p.

3. Constatons d'abord que

$$\mathbb{E}\left[X^s\right] = \mathbb{E}\left[\hat{X}_1\right] - \mathbb{E}\left[\hat{X}_2\right]$$

Par le calcul du **2.** :

$$\mathbb{E}\left[X^s\right] = \mathbb{E}\left[X\right] - \mathbb{E}\left[X\right] = 0$$

donc X^s est centrée. Puis, pour la variance, l'indépendance de X_1 et X_2 donne

$$\sigma^2_{X^s} = \sigma^2_{\hat{X}_1} + \sigma^2_{\hat{X}_2}$$

Donc, toujours par le même calcul qu'au **2.** :

$$\sigma^2_{X^s} = 2\sigma^2_X$$

4. Soit I fini. On se donne $(A_i)_{i\in I}$ et $(B_i)_{i\in I}$ deux familles de boréliens. Montrons que la famille $\left(\hat{X}_i\right)_{i\in I}$ est $\mathbb{P}\otimes\mathbb{P}$-indépendante.

$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}\left\{\hat{X}_i\in A_i\times B_i\right\}\right)$$
$$=$$
$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}\{X_i\in A_i\}\times\{X_i\in B_i\}\right)$$

Alors, par définition de la mesure produit

$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}\left\{\hat{X}_i\in A_i\times B_i\right\}\right)$$
$$=$$
$$\mathbb{P}\left(\bigcap_{i\in I}\{X_i\in A_i\}\right)\mathbb{P}\left(\bigcap_{i\in I}\{X_i\in B_i\}\right)$$

Par indépendance de la famille $(X_i)_{i\in I}$:

$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}\left\{\hat{X}_i\in A_i\times B_i\right\}\right)$$
$$=$$
$$\prod_{i\in I}\mathbb{P}\left(X_i\in A_i\right)\prod_{i\in I}\mathbb{P}\left(X_i\in B_i\right)$$

En regroupant, et en utilisant la mesure produit :

$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}\left\{\hat{X}_i\in A_i\times B_i\right\}\right)$$
$$=$$
$$\prod_{i\in I}\mathbb{P}\otimes\mathbb{P}\left(\{X_i\in A_i\}\times\{X_i\in B_i\}\right)$$

Soit, par définition de $\hat{X}_i$:

$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}\left\{\hat{X}_i\in A_i\times B_i\right\}\right)$$
$$=$$
$$\prod_{i\in I}\mathbb{P}\otimes\mathbb{P}\left(\hat{X}_i\in A_i\times B_i\right)$$

Ce qui signifie exactement que la famille $\left(\hat{X}_i\right)_{i\in I}$ est indépendante. Puisque $X_i^{\mathsf{s}}=f\left(\hat{X}_i\right)$, la famille $(X_i^{\mathsf{s}})_{i\in I}$ est indépendante. Pour le montrer, on a :

$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}X_i^{\mathsf{s}}\in A_i\right)$$
$$=$$
$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}\hat{X}_i\in f^{-1}(A_i)\right)$$

Avec f mesurable, donc $f^{-1}(A_i)$ est un borélien de $\mathbb{R}^2$. Par indépendance des $(\hat{X}_i)_{i\in I}$, on a alors

$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}X_i^{\mathsf{s}}\in A_i\right)$$
$$=$$
$$\prod_{i\in I}\mathbb{P}\otimes\mathbb{P}\left(\hat{X}_i\in f^{-1}(A_i)\right)$$

D'où

$$\mathbb{P}\otimes\mathbb{P}\left(\bigcap_{i\in I}X_i^{\mathsf{s}}\in A_i\right)$$
$$=$$
$$\prod_{i\in I}\mathbb{P}\otimes\mathbb{P}\left(X_i^{\mathsf{s}}\in A_i\right)$$

La famille $(X_i^{\mathsf{s}})_{i\in I}$ est donc bien indépendante. □

Le deuxième lemme est le suivant.

Lemme 7.2 : Condition suffisante de convergence de la série des variances

Soit $(X_n)_n$ une suite de variables aléatoires réelles centrées indépendantes. On note $S_n \overset{\text{déf.}}{=} \sum_{k=1}^n X_k$. On suppose alors :

* La suite $(X_n)_n$ est uniformément bornée $\mathbb{P}$-presque sûrement :

$$\exists C>0,\forall n\in\mathbb{N},|X_n|\leqslant C\quad\mathbb{P}\text{-ps};$$

* S_n vérifie :

$$\mathbb{P}\left(\sup_{n\in\mathbb{N}}|S_n|<+\infty\right)>0$$

Alors la série $\sum_n\sigma_{X_n}^2$ converge.

Démonstration du lemme : 1. On dispose de la croissance des événements :

$$\left\{\sup_{n\in\mathbb{N}}|S_n|<+\infty\right\}=\bigcup_{l\in\mathbb{N}}^{\uparrow}\left\{\sup_{n\in\mathbb{N}}|S_n|\leqslant l\right\}$$

Par la deuxième hypothèse, il suit qu'il existe $l\in\mathbb{N}^*$ tel que

$$\mathbb{P}\left(\sup_{n\in\mathbb{N}}|S_n|\leqslant l\right)>0$$

On fixe $l\in\mathbb{N}^*$. On note

$$A\overset{\text{déf.}}{=}\left\{\sup_{n\in\mathbb{N}}|S_n|\leqslant l\right\}$$

et pour tout $p\in\mathbb{N}$:

$$A_p\overset{\text{déf.}}{=}\left\{\sup_{0\leqslant n\leqslant p}|S_n|\leqslant l\right\}$$

2. Montrons que

$$\mathbb{P}(A)\sigma^2_{X_{p+1}}\leqslant\mathbb{E}\left[\mathbf{1}_{A_p}S^2_{p+1}\right]-\mathbb{E}\left[\mathbf{1}_{A_p}S^2_p\right]$$

Pour cela, on développe $S^2_{p+1}=(S_p+X_{p+1})^2$:

$$\begin{gathered}\mathbb{E}\left[\mathbf{1}_{A_p}S^2_{p+1}\right]\\=\\\mathbb{E}\left[\mathbf{1}_{A_p}S^2_p\right]+2\mathbb{E}\left[\mathbf{1}_{A_p}S_pX_{p+1}\right]+\mathbb{E}\left[\mathbf{1}_{A_p}X^2_{p+1}\right]\end{gathered}$$

Par indépendance de la famille $(X_n)_n$, on a alors

$$\begin{gathered}\mathbb{E}\left[\mathbf{1}_{A_p}S^2_{p+1}\right]\\=\\\mathbb{E}\left[\mathbf{1}_{A_p}S^2_p\right]+2\mathbb{E}\left[\mathbf{1}_{A_p}S_p\right]\mathbb{E}\left[X_{p+1}\right]+\mathbb{E}\left[\mathbf{1}_{A_p}\right]\mathbb{E}\left[X^2_{p+1}\right]\end{gathered}$$

Et puisque les variables sont centrées :

$$\begin{gathered}\mathbb{E}\left[\mathbf{1}_{A_p}S^2_{p+1}\right]\\=\\\mathbb{E}\left[\mathbf{1}_{A_p}S^2_p\right]+\mathbb{P}(A_p)\mathbb{E}\left[X^2_{p+1}\right]\end{gathered}$$

Ainsi, par l'inclusion $A\subset A_p$, on a bien en réarrangeant et en utilisant le caractère centré de X_{p+1} :

$$\mathbb{P}(A)\sigma^2_{X_{p+1}}\leqslant\mathbb{E}\left[\mathbf{1}_{A_p}S^2_{p+1}\right]-\mathbb{E}\left[\mathbf{1}_{A_p}S^2_p\right]$$

3. On cherche à sommer cette inégalité, et aboutir à une somme télescopique. Constatons alors que par l'inclusion $A_{p+1}\subset A_p$:

$$\mathbf{1}_{A_p}=\mathbf{1}_{A_{p+1}}+\mathbf{1}_{A_p\setminus A_{p+1}}$$

Ainsi,

$$\begin{gathered}\mathbb{P}(A)\sigma^2_{X_{p+1}}\\\leqslant\\\mathbb{E}\left[\mathbf{1}_{A_{p+1}}S^2_{p+1}\right]+\mathbb{E}\left[\mathbf{1}_{A_p\setminus A_{p+1}}S^2_{p+1}\right]-\mathbb{E}\left[\mathbf{1}_{A_p}S^2_p\right]\end{gathered}$$

Or, $\mathbb{P}$-presque sûrement, le caractère uniformément borné de $(X_n)_n$ donne

$$|S_{p+1}|\leqslant|S_p|+C$$

Et donc

$$\begin{gathered}\mathbb{P}(A)\sigma^2_{X_{p+1}}\\\leqslant\\(l+C)^2\mathbb{P}\left(A_p\setminus A_{p+1}\right)+\mathbb{E}\left[\mathbf{1}_{A_{p+1}}S^2_{p+1}\right]-\mathbb{E}\left[\mathbf{1}_{A_p}S^2_p\right]\end{gathered}$$

Et on arrive à notre somme télescopique ici. On a alors

$$\begin{gathered}\mathbb{P}(A)\sum_{p=1}^{n-1}\sigma^2_{X_{p+1}}\\\leqslant\\(l+C)^2\sum_{p=1}^{n-1}\mathbb{P}\left(A_p\setminus A_{p+1}\right)+\mathbb{E}\left[\mathbf{1}_{A_n}S^2_n\right]-\mathbb{E}\left[\mathbf{1}_{A_1}S^2_1\right]\end{gathered}$$

D'où

$$\mathbb{P}(A)\sum_{p=1}^{n-1}\sigma^2_{X_{p+1}}\leqslant(l+C)^2+l^2$$

Donc la série $\sum_n\sigma^2_{X_n}$ est effectivement convergente, puisque $\mathbb{P}(A)>0$. □

Muni de ces lemmes, montrons le théorème des trois séries.

Démonstration : [$\Longrightarrow$] • Supposons que $\sum_{n\geqslant1}X_n$ converge $\mathbb{P}$-presque sûrement. Alors en particulier, $(X_n)_n$ converge $\mathbb{P}$-presque sûrement vers 0. Ainsi

$$\mathbb{P}\left(\liminf_{n\to+\infty}\{|X_n|\leqslant c\}\right)=1$$

donc

$$\mathbb{P}\left(\liminf_{n\to+\infty}\{|X_n|>c\}\right)=0$$

Par indépendance des $(X_n)_n$ et par contraposée du lemme de Borel-Cantelli, on obtient alors

$$\sum_{n=0}^{+\infty}\mathbb{P}(|X_n|>c)<+\infty$$

ce qui donne la convergence de la troisième série.

• De plus, par définition de la troncature Y_n, on a aussi

$$\mathbb{P}\left(\liminf_{n\to+\infty}\{X_n = Y_n\}\right) = 1$$

Ainsi, $\sum_{n\geqslant 1} Y_n$ est de même nature que $\sum_n X_n$, donc converge $\mathbb{P}$-presque sûrement.

• Pour les deux autres séries, utilisons le symétrisé Y_n^{s} de Y_n. Par le premier lemme, la famille $(Y_n^{\mathsf{s}})_{n\geqslant 1}$ est indépendante, et est centrée. Puisque $\sum_n Y_n$ converge presque sûrement, le théorème de Fubini assure alors la convergence $\mathbb{P}\otimes\mathbb{P}$-presque sûre de $\sum_n Y_n^{\mathsf{s}}$, donc

$$\mathbb{P}\left(\sup_{n\in\mathbb{N}}\left|\sum_{j=1}^{n} Y_j^{\mathsf{s}}\right| < +\infty\right)$$

De plus, puisque $(Y_n)_n$ est uniformément bornée par c, $(Y_n^{\mathsf{s}})_n$ est alors uniformément bornée par $2c$. Par le deuxième lemme, il suit que

$$\sum_{n=1}^{+\infty} \sigma^2_{Y_n^{\mathsf{s}}} < +\infty$$

Or,

$$\sigma^2_{Y_n^{\mathsf{s}}} = 2\sigma^2_{Y_n}$$

Donc la série $\sum_n \sigma^2_{Y_n}$ converge, d'où la convergence de la deuxième série.

• Notons enfin

$$\tilde{Y}_n \overset{\text{déf.}}{=} Y_n - \mathbb{E}\left[Y_n\right]$$

alors on a aussi la convergence de $\sum_n \sigma^2_{\tilde{Y}_n}$. En conséquence, pusique $\tilde{Y}_n$ est centrée, $\sum_n \tilde{Y}_n$ converge $\mathbb{P}$-presque sûrement. Puisque $\sum_n Y_n$ converge aussi $\mathbb{P}$-presque sûrement, il suit que $\sum_n \mathbb{E}[Y_n]$ converge, d'où la convergence de la première série. On a bien montré la convergence des trois séries.

[$\Longleftarrow$] Si les trois séries convergent, alors puisque

$$\sigma^2_{\tilde{Y}_n} = \sigma^2_{Y_n}$$

alors $\sum_{n\geqslant 1} \sigma^2_{\tilde{Y}_n}$ converge, et $\tilde{Y}_n$ est centrée donc la série $\sum_n \tilde{Y}_n$ converge $\mathbb{P}$-presque sûrement, donc $\sum_n Y_n$ converge $\mathbb{P}$-presque sûrement. De plus,

$$\sum_{n=0}^{+\infty} \mathbb{P}(X_n \neq Y_n) = \sum_{n=0}^{+\infty} \mathbb{P}(|X_n| > c) < +\infty$$

Par le premier lemme de Borel-Cantelli, on a alors

$$\mathbb{P}\left(\limsup_{n\to+\infty}\{X_n \neq Y_n\}\right) = 0$$

Donc

$$\mathbb{P}\left(\liminf_{n\to+\infty}\{X_n = Y_n\}\right) = 1$$

Or, la série $\sum_n Y_n$ converge $\mathbb{P}$-presque sûrement. Ainsi, on a bien montré que $\sum_n X_n$ converge $\mathbb{P}$-presque sûrement. □

8 Inversion de Fourier pour les probabilités

Référence : Jean-Yves Ouvrard, *Probabilités 2*

Définition 8.1

Soit μ une mesure de probabilité définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On note $\hat{\mu} : \mathbb{R} \longrightarrow \mathbb{C}$ sa *transformée de Fourier* définie par :

$$\forall t \in \mathbb{R}, \hat{\mu}(t) \overset{\text{déf.}}{=} \int_{\mathbb{R}} e^{\imath x t}\, \mathrm{d}\mu(x)$$

En particulier, si X est une variable aléatoire réelle, alors $\hat{\mathbb{P}}_X$ n'est rien d'autre que sa fonction caractéristique φ_X.

Théorème 8.1 : Inversion de Fourier pour les probabilités

Soit μ une mesure de probabilité sur $\mathbb{R}$. Soient $a < b$. Alors

$$\lim_{T\to+\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-\imath t a} - e^{-\imath t b}}{\imath t}\, \hat{\mu}(t)\, \mathrm{d}t = \frac{1}{2}\mu(\{a,b\}) + \mu(]a,b[).$$

Démonstration : **1.** Fixons $T > 0$. Commençons par constater que

$$\begin{pmatrix} [-T,T]\times\mathbb{R} & \longrightarrow & \mathbb{C} \\ (t,x) & \longmapsto & \frac{e^{-\imath ta}-e^{-\imath tb}}{\imath t}\, e^{\imath tx} \end{pmatrix}$$

est bornée par le théorème des accroissements finis. Ainsi, puisque

$$\int_{-T}^{T} \frac{e^{-\imath ta}-e^{-\imath tb}}{\imath t}\,\hat{\mu}(t)\,\mathrm{d}t = \int_{-T}^{T} \frac{e^{-\imath ta}-e^{-\imath tb}}{\imath t}\left[\int_{\mathbb{R}} e^{\imath tx}\,\mathrm{d}\mu(x)\right]\mathrm{d}t$$

Le théorème de Fubini donne alors

$$\int_{-T}^{T} \frac{e^{-\imath ta}-e^{-\imath tb}}{\imath t}\,\hat{\mu}(t)\,\mathrm{d}t = \int_{\mathbb{R}}\left[\int_{-T}^{T} \frac{e^{-\imath ta}-e^{-\imath tb}}{\imath t}\, e^{\imath tx}\,\mathrm{d}t\right]\mathrm{d}\mu(x)$$

Notons alors $I_T(x)$ l'intégrale intérieure, avec un facteur 2π, et étudions sa limite à x fixé lorsque $[T \to +\infty]$:

$$I_T(x) \overset{\text{déf.}}{=} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{\imath t(x-a)}-e^{\imath t(x-b)}}{\imath t}\,\mathrm{d}t$$

2. Pour étudier la limite, on va exprimer $I_T(x)$ en fonction de l'intégrale de $\frac{\sin x}{x}$, dont on suppose connue sa valeur sur $[0,+\infty[$, à savoir $\frac{\pi}{2}$.

• Supposons que $x \notin \{a,b\}$. On peut alors découper l'intégrale en deux et faire les changements de variables $u = t(x-a)$ et $u = t(x-b)$. Cela donne

$$I_T(x) = \frac{1}{2\pi}\int_{-T(x-a)}^{T(x-a)} \frac{e^{\imath u}}{\imath u}\,\mathrm{d}u - \frac{1}{2\pi}\int_{-T(x-b)}^{T(x-b)} \frac{e^{\imath u}}{\imath u}\,\mathrm{d}u$$

Or, l'application $u \longmapsto \frac{\cos u}{u}$ est impaire, donc il ne reste plus que

$$I_T(x) = \frac{1}{2\pi}\int_{-T(x-a)}^{T(x-a)} \frac{\sin u}{u}\,\mathrm{d}u - \frac{1}{2\pi}\int_{-T(x-b)}^{T(x-b)} \frac{\sin u}{u}\,\mathrm{d}u$$

Selon la position de x par rapport à a, b, on peut alors déterminer la limite de $I_T(x)$. Si $x < a(< b)$ alors $x-a < 0$ et $x-b < 0$. Ainsi,

$$I_T(x) \xrightarrow[T\to+\infty]{} \frac{1}{2\pi}(-\pi+\pi) = 0$$

De même si $(a <)b < x$. Si $x \in]a,b[$ alors les deux intégrales convergent vers la même limite en valeur absolue, mais opposées. Ce qui donne :

$$I_T(x) \xrightarrow[T\to+\infty]{} \frac{1}{2\pi}(\pi+\pi) = 1$$

• Supposons que $x = a$. Dans ce cas,

$$I_T(a) = \frac{1}{2\pi}\int_{-T}^{T} \frac{1-e^{\imath t(b-a)}}{\imath t}\,\mathrm{d}t$$

Alors, on découpe selon partie réelle et imaginaire, puis on fait le changement de variables $u = T(b-a)$, qui va inverser le signe de la deuxième intégrale.

$$I_T(a) = \frac{1}{2\pi}\int_{-T(b-a)}^{T(b-a)} \frac{1-\cos u}{\imath u}\,\mathrm{d}u + \frac{1}{2\pi}\int_{-T(b-a)}^{T(b-a)} \frac{\sin u}{\imath u}\,\mathrm{d}u$$

La première intégrale est nulle par imparité de $u \longmapsto \frac{1-\cos u}{u}$. Quabd à la deuxième, sa limite est bien connue : il suit que

$$I_T(a) \xrightarrow[T\to+\infty]{} \frac{1}{2}$$

De même pour $I_T(b)$, ce qui correspond à faire le changement de variables t donne $-t$. On a alors déterminé la limite de $I_T(x)$, que l'on résume en termes d'indicatrices :

$$\forall x \in \mathbb{R}, I_T(x) \xrightarrow[T\to+\infty]{} \frac{1}{2}\mathbf{1}_{\{a,b\}}(x) + \mathbf{1}_{]a,b[}(x)$$

3. Pour revenir à notre problème de départ, on a alors montré la convergence simple de $I_T(x)$ vers une fonction bornée donc intégrable par rapport à la mesure de probabilité μ. De plus, par les calculs explicites de $I_T(x)$ en fonction de l'intégrale de $\frac{\sin u}{u}$, on peut majorer $I_T(x)$ par des constantes indépendantes de T. En effet, puisque $\alpha \longmapsto \int_0^\alpha \frac{\sin u}{u}\,\mathrm{d}u$ est continue et tend vers $\frac{\pi}{2}$ en $+\infty$, il s'agit là d'une fonction bornée, disons par $M > 0$. Ainsi, $|I_T(x)| \leqslant 2M$, pour tout $x \in \mathbb{R}$ et $T > 0$. Par théorème de convergence dominée, il suit que notre intégrale

$$I_\varphi(T) \overset{\text{déf.}}{=} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-\imath ta}-e^{-\imath tb}}{\imath t}\,\hat{\mu}(t)\,\mathrm{d}t$$

vérifie

$$I_\varphi(T) \xrightarrow[T\to+\infty]{} \int_{\mathbb{R}}\left(\frac{1}{2}\mathbf{1}_{\{a,b\}}(x) + \mathbf{1}_{]a,b[}(x)\right)\mathrm{d}\mu(x)$$

D'où la formule d'inversion de Fourier pour les probas annoncée.

□

Calculons par un argument d'analyse complexe l'intégrale de $\frac{\sin u}{u}$. (Calcul classique, l'une des références est Patrice Tauvel, *Exercices d'analyse complexe*, exercice 19 du chapitre *Calculs d'in-*

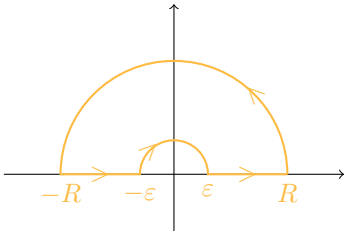
tégrales).

Lemme 8.1 : Calcul d'une intégrale

On a

$$\int_0^{+\infty} \frac{\sin u}{u}\,\mathrm{d}u = \frac{\pi}{2}$$

Démonstration du lemme : Il s'agit d'appliquer le théorème des résidus à la fonction méromorphe $g : z \longmapsto \frac{e^{\imath z}}{z}$ sur le contour suivant :



g admet un pôle simple en 0, dont le résidu en ce point vaut 1. Si $\Gamma_{R,\varepsilon}$ désigne le contour, le théorème des résidus fournit alors l'égalité suivante, en fixant l'orientation induite par le grand demi-cercle de rayon R (qui ne se voit pas dans le signe du résidu puisque l'indice en 0 de ce contour est nul) :

$$\oint_{\Gamma_{R,\varepsilon}} \frac{e^{\imath z}}{z}\,\mathrm{d}z = 0$$

Notons γ_R le contour du demi-cercle de rayon R, γ_ε celui de rayon ε. On a le découpage suivant :

$$\oint_{\Gamma_{R,\varepsilon}} = \int_{\gamma_R} + \int_{\gamma_\varepsilon} + \int_{[-R,-\varepsilon]} + \int_{[\varepsilon,R]}$$

Pour la première intégrale, la paramétrisation

$$\forall \theta \in [0,\pi] \gamma_R(\theta) = Re^{\imath\theta}$$

donne

$$\int_{\gamma_R} = \int_0^{\pi} \frac{\exp\left(\imath R e^{\imath\theta}\right)}{Re^{\imath\theta}} \cdot \imath R e^{\imath\theta}\,\mathrm{d}\theta$$

On peut simplifier, et obtenir

$$\int_{\gamma_R} = \imath \int_0^{\pi} e^{\imath R\cos\theta} e^{-R\sin\theta}\,\mathrm{d}\theta$$

Par théorème de convergence dominée (on intègre sur un compact, tout le monde est continue sur ce compact), il suit que

$$\int_{\gamma_R} \frac{e^{\imath z}}{z}\,\mathrm{d}z \xrightarrow[R\to+\infty]{} 0$$

De même pour le petit cercle, dont la paramétrisation change à cause de l'orientation, qui donne

$$\int_{\gamma_\varepsilon} = -\imath \int_0^{\pi} e^{\imath\varepsilon\cos\theta} e^{-\varepsilon\sin\theta}\,\mathrm{d}\theta$$

Par théorème de convergence dominée, on a cette fois-ci :

$$\int_{\gamma_\varepsilon} \frac{e^{\imath z}}{z}\,\mathrm{d}z \xrightarrow[\varepsilon\to 0^+]{} -\pi$$

Examinons les intégrales sur les segments. Un petit changement de variables donne

$$\int_{[-R,-\varepsilon]} + \int_{[\varepsilon,R]} = \int_{\varepsilon}^{R} \frac{2\imath\sin u}{u}\,\mathrm{d}u$$

Il suit qu'un passage à la limite successif $[R \to +\infty]$ et $[\varepsilon \to 0]$ donne

$$2\imath \int_0^{+\infty} \frac{\sin u}{u}\,\mathrm{d}u = \imath\pi$$

d'où la valeur de l'intégrale annoncée. □

9 Caractérisation des variables aléatoires gaussiennes

Référence : Jean-Yves Ouvrard, *Probabilités 2*

Proposition 9.1 : Une caractérisation de la gaussienne avec les mesures empiriques

Soient μ une loi de probabilité et $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées L^2 de loi μ. On note

$$M_n \overset{\text{déf.}}{=} \frac{1}{n}\sum_{k=1}^{n} X_k \text{ et } \Sigma_n \overset{\text{déf.}}{=} \frac{1}{n}\sum_{k=1}^{n} X_k^2 - M_n^2.$$

Alors les assertions suivantes sont équivalentes :

(i) $\mu = \mathcal{N}(m, \sigma^2)$ avec $m = \mathbb{E}[X_1]$ et $\sigma^2 = \sigma_{X_1}^2$;

(ii) M_n et Σ_n sont indépendantes.

Démonstration : **[$\Longrightarrow$]** Supposons que μ soit la loi gaussien de paramètre (m, σ^2). Si $X_1, \cdots, X_n$ sont donnés, on note

$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

Alors $\vec{X}$ suit la loi $\mathcal{N}(m\vec{\mathbf{1}}, \sigma^2 I_n)$. Soit C une matrice orthogonale de la forme :

$$C \overset{\text{déf.}}{=} \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ * & * & & * \\ \vdots & & & \vdots \\ * & * & \cdots & * \end{pmatrix} \in \mathcal{O}_n(\mathbb{R})$$

Alors $[C\vec{X}]_1$, la première composante de $C\vec{X}$ vaut :

$$\left[C\vec{X}\right]_1 = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k = \sqrt{n} M_n$$

Ainsi, on peut calculer Σ_n en fonction de C :

$$\Sigma_n = \frac{1}{n}\left\|\vec{X}\right\|^2 - \frac{1}{n}\left[C\vec{X}\right]_1^2$$

Puisque C est orthogonale, on a alors

$$\Sigma_n = \frac{1}{n}\left\|C\vec{X}\right\|^2 - \frac{1}{n}\left[C\vec{X}\right]_1^2$$

Ainsi, par définition de la norme (euclidienne) :

$$\Sigma_n = \frac{1}{n}\sum_{k=2}^{n}\left[C\vec{X}\right]_k^2$$

Donc Σ_n ne contient pas le terme $\left[C\vec{X}\right]_1$. Enfin, si $\vec{X}$ suit une loi normale alors

$$C\vec{X} \sim \mathcal{N}\left(mC\vec{\mathbf{1}}, \sigma^2 C I_n {}^t C\right) = \mathcal{N}\left(mC\vec{\mathbf{1}}, \sigma^2 I_n\right)$$

En conséquence, en regardant la covariance, les variables $\left[C\vec{X}\right]_i$ et $\left[C\vec{X}\right]_j$ sont indépendantes, dès lors que $i \neq j$. Par conséquent, M_n et Σ_n sont bien indépendants.

Remarquons au passage (cela ne sert à rien dans notre démonstration) que pour $m = 0$ et $\sigma^2 = 1$, M_n suit $\mathcal{N}\left(0, \frac{1}{n}\right)$ et Σ_n suit la loi du chi 2 à $n-1$ paramètres : $\chi^2(n-1)$.

[$\Longleftarrow$] Supposons M_n et Σ_n indépendantes. Supposons dans un premier temps que $m = 0$. On note φ la transformée de Fourier de μ, c'est la fonction caractéristique de X_1. On note $S_n = nM_n$ la somme des n premiers X_k, et $V_n = n\Sigma_n$. Puisque Σ_n et M_n sont indépendantes, c'est alors le cas de S_n et de V_n. L'idée est d'utiliser le théorème d'injectivité de la transformée de Fourier pour les probabilités, c'est-à-dire montrer que φ est la fonction caractéristique d'une variable aléatoire gaussienne. Pour cela, on aboutit à une équation différentielle en calculant $\mathbb{E}\left[V_n e^{iuS_n}\right]$ de deux manières différentes.

1. Calculons $\mathbb{E}[V_n]$. Par définition :

$$V_n = \sum_{k=1}^{n} X_k^2 - \frac{S_n^2}{n}$$

Ainsi,

$$\mathbb{E}[V_n] = n\sigma^2 - \frac{1}{n}\mathbb{E}\left[S_n^2\right]$$

Or, les $(X_i)_i$ sont supposées centrées $(m = 0)$, et sont indépendantes. Ainsi,

$$\mathbb{E}[V_n] = n\sigma^2 - \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[X_k^2\right] = (n-1)\sigma^2$$

2. Continuons, et utilisons l'indépendance pour calculer $\mathbb{E}\left[V_n e^{iuS_n}\right]$, pour $u \in \mathbb{R}$:

$$\mathbb{E}\left[V_n e^{iuS_n}\right] = \mathbb{E}[V_n]\varphi_{S_n}(u)$$

Ainsi, par indépendance des X_n :

$$\mathbb{E}\left[V_n e^{iuS_n}\right] = \mathbb{E}[V_n](\varphi(u))^n$$

3. Or, on peut aussi calculer cette même espérance en fonction de φ' et de φ''. Au lieu d'utiliser l'indépendance entre V_n et S_n, on utilise la définition de V_n :

$$\mathbb{E}\left[V_n e^{iuS_n}\right] = \sum_{k=1}^{n}\mathbb{E}\left[X_k^2 e^{iuS_n}\right] - \frac{1}{n}\mathbb{E}\left[S_n^2 e^{iuS_n}\right]$$

On fait apparaître les X_k dans le premier terme pour exploiter leur indépendance :

$$\mathbb{E}\left[V_n e^{\imath u S_n}\right] = \sum_{k=1}^{n} \mathbb{E}\left[X_k^2 \prod_{j=1}^{n} e^{\imath u X_j}\right] - \frac{1}{n}\mathbb{E}\left[S_n^2 e^{\imath u S_n}\right]$$

Alors, par indépendance :

$$\begin{aligned}\mathbb{E}\left[V_n e^{\imath u S_n}\right] = & \sum_{k=1}^{n}\left(\mathbb{E}\left[X_k^2 e^{\imath u X_k}\right]\prod_{j\neq k}\mathbb{E}\left[e^{\imath u X_j}\right]\right) \\ - & \frac{1}{n}\mathbb{E}\left[S_n^2 e^{\imath u S_n}\right]\end{aligned}$$

Puisque X_k et S_n sont L^2, on dispose des égalités sur les fonctions caractéristiques :

$$\mathbb{E}\left[X_k^2 e^{\imath u X_k}\right] = -\varphi''(u)$$

et

$$\mathbb{E}\left[S_n^2 e^{\imath u S_n}\right] = -(\varphi^n)''(u)$$

Par conséquent, en dérivant deux fois :

$$\mathbb{E}\left[S_n^2 e^{\imath u S_n}\right] = -n\left(\varphi''(u)\varphi^{n-1}(u) + \varphi'(u)^2\varphi^{n-2}(u)\right)$$

Il suit l'expression suivante :

$$\begin{aligned}\mathbb{E}\left[V_n e^{\imath u S_n}\right] = & -\sum_{k=1}^{n}\varphi''(u)\varphi^{n-1}(u) \\ & + \varphi''(u)\varphi^{n-1}(u) \\ & + (n-1)\varphi'(u)^2\varphi^{n-2}(u)\end{aligned}$$

4. Par **2.** et **3.**, on obtient l'équation différentielle suivante :

$$\begin{aligned}(n-1)\sigma^2\varphi(u)^n = & -(n-1)\varphi''(u)\varphi^{n-1}(u) \\ & + (n-1)\varphi'(u)^2\varphi^{n-2}(u)\end{aligned}$$

Ainsi, sur $\{\varphi \neq 0\}$, on obtient la simplification suivante :

$$\sigma^2 = \frac{\varphi'(u)^2}{\varphi(u)^2} - \frac{\varphi''(u)}{\varphi(u)} = \frac{-\mathrm{d}}{\mathrm{d}u}\left[\frac{\varphi'}{\varphi}\right]$$

D'où, sur $\{\varphi \neq 0\}$, puisque $\varphi(0) = 1$ et $\varphi'(0) = 0$:

$$\varphi(u) = \exp\left(\frac{-\sigma^2 u^2}{2}\right)$$

5. Il ne nous reste plus qu'à montrer que φ ne s'annule jamais. Puisque $\varphi(0) = 1$, on a $0 \in \varphi^{-1}(\mathbb{R}\backslash\{0\})$. Puisque φ est continue, $\varphi^{-1}(\mathbb{R}\backslash\{0\})$ est ouvert, contenant 0. Il existe alors $a_1 > 0$ telle que

$$[-a_1, a_1] \subset \varphi^{-1}(\mathbb{R}\backslash\{0\})$$

On peut alors résoudre l'équation différentielle, toujours avec $\varphi(0) = 0$ et $\varphi'(0) = 0$. On a alors

$$\forall u \in [-a_1, a_1], \varphi(u) \overset{(*)}{=} \exp\left(\frac{\sigma^2 u^2}{2}\right)$$

Considérons alors

$$a \overset{\text{déf.}}{=} \sup\{\ \alpha \geqslant a_1, [-a_1, \alpha] \subset \{\varphi \neq 0\}\ \}$$

Supposons par l'absurde que $a < +\infty$. Alors on aurait l'inclusion suivante (éventuellement avec $[-a_1, a-\varepsilon] = \varnothing$) :

$$\forall \varepsilon > 0, [-a_1, a-\varepsilon] \subset \{\varphi \neq 0\}$$

Par conséquent, φ serait de la forme $(*)$ sur $[-a_1, a-\varepsilon]$, donc sur $[-a_1, a]$ par continuité de φ et du second membre. En particulier, $\varphi(a) \neq 0$. Par continuité de φ, il existerait alors $b > a$ tel que

$$[-a_1, a] \subsetneq [-a_1, b] \subset \{\varphi \neq 0\}$$

Ce qui contredit la maximalité de a. Donc, φ vérifie $(*)$ sur $[-a_1, +\infty[$. De plus, sur $[-a_1, +\infty[$:

$$\varphi(-u) = \overline{\varphi(u)} = \varphi(u)$$

Il suit que φ est définie sur $\mathbb{R}$, et vérifie $(*)$. Ainsi,

$$\forall u \in \mathbb{R}, \varphi(u) = \widehat{\mathcal{N}(0, \sigma^2)}(u)$$

Par théorème d'injectivité, il suit alors que

$$\mu = \mathcal{N}(0, \sigma^2)$$

6. Pour conclure, il ne reste plus qu'à montrer que ce résultat est vrai pour tout $m \in \mathbb{R}$. On note alors $\tilde{X}_n = X_n - m$. Alors $\tilde{S}_n = S_n - nm$ et $\tilde{V}_n = V_n$, donc $\tilde{S}_n$ et $\tilde{V}_n$ sont indépendantes. Par ce qu'on a montré, on a alors

$$\mathbb{P}_{\tilde{X}_1} = \mathcal{N}(0, \sigma^2)$$

Ainsi,

$$\mathbb{P}_{X_1} = \mathcal{N}(m, \sigma^2)$$

Ce qui caractérise bien la loi normale. □

10 Théorème de Hadamard Lévy

Références :

— *Analyse pour l'agrégation*, Claude ZUILY et Hervé QUÉFFELEC ;

— *Analyse pour l'agrégation de mathématiques : 40 développements*, Julien et Laurent BERNIS

Théorème 10.1 : Théorème de HADAMARD-LÉVY

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ une fonction de classe C^2. Les assertions suivantes sont équivalentes :

(i) Pour tout compact K de $\mathbb{R}^n$, $f^{-1}(K)$ est compact (on dit que f est *propre*) et pour tout $x \in \mathbb{R}^n$, $\mathrm{d}f(x) \in \mathrm{GL}_n(\mathbb{R})$;

(ii) $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ est un C^1-difféomorphisme.

Remarque : f est propre si et seulement si $|f(x)| \xrightarrow[|x|\to+\infty]{} +\infty$.

Les deux références en question font intervenir les mêmes arguments. Mais, dans ZQ, la démonstration est complète mais très longue, tandis que les BERNIS admettent un théorème de CAUCHY à paramètre qui permettent de pouvoir probablement caser ce développements dans les fameuses 15 minutes. Énonçons le théorème admis, et démontrons le théorème d'HADAMARD-LÉVY avec et sans ce théorème. L'énoncé est adapté de l'ouvrage de Sylvie BENZONI-GAVAGE.

Théorème 10.2 : de CAUCHY-LIPSCHITZ à paramètre, cas C^1

Soient I un intervalle ouvert de $\mathbb{R}$, U un ouvert de $\mathbb{R}^n$ et Λ un autre ouvert de $\mathbb{R}^n$. Si

$$f : I \times U \times \Lambda \longrightarrow \mathbb{R}^n$$

est de classe C^1 alors les solutions $x(t, \lambda)$ de

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t, \lambda) = f(t, x(t), \lambda)$$

sont de classe C^1 sur $I \times \Lambda$.

Rédigeons la preuve du théorème de HADAMARD-LÉVY, où on va mettre en évidence l'endroit exact où le théorème de CAUCHY à paramètre est utilisé, et comment ZQ résout cette difficulté. La stratégie est d'utiliser ces deux lemmes qui apparaissent quand la stratégie de démonstration à l'aide d'équations différentielles est fixée.

Lemme 10.1 : CNS pour être un difféomorphisme

Soit $f \in \mathcal{C}^2(\mathbb{R}^n)$ tel que $f(0) = 0$, propre et telle que

$$\forall x \in \mathbb{R}^n, \mathrm{d}f(x) \in \mathrm{GL}_n(\mathbb{R})$$

Alors, s'il existe $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ telle que

(i) Pour tout $y \in \mathbb{R}^n$, $f(g(y)) = y$;

(ii) L'application g est surjective ;

alors $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ est C^1-difféomorphisme.

Le théorème d'inversion local donne le caractère C^1 et les hypothèses sur g la bijectivité. Le deuxième lemme est plus technique, et il ne sera pas question de l'aborder dans un développement, si ce n'est dans une question du jury.

Lemme 10.2 : Calcul d'une différentielle

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ de classe C^2 telle que

$$\forall x \in \mathbb{R}^n, \mathrm{d}f(x) \in \mathrm{GL}_n(\mathbb{R})$$

On définit pour tout $y \in \mathbb{R}^n$:

$$F_y : x \longrightarrow [\mathrm{d}f(x)]^{-1} \cdot y$$

Alors F_y est de classe C^1 et vérifie pour tout $x \in \mathbb{R}^n$ et $h \in \mathbb{R}^n$:

$$\mathrm{d}F_y(x) \cdot h = -[\mathrm{d}f(x)]^{-1} \cdot \mathrm{d}^2 f(x) \cdot \left(h, [\mathrm{d}f(x)]^{-1} \cdot y\right)$$

Démontrons enfin le théorème.

Démonstration : [(ii) $\Longrightarrow$ (i)] Si f est un C^1-difféomorphisme alors f^{-1} est continue, donc pour tout compact K de $\mathbb{R}^n$, $f^{-1}(K)$ est compact, donc f est propre. De plus, par définition d'une application réciproque, pour tout $x \in \mathbb{R}^n$:

$$\mathrm{d}\left(f \circ f^{-1}\right)(x) = \mathrm{id} = \mathrm{d}\left(f^{-1} \circ f\right)(x)$$

D'où par le théorème des fonctions composées pour l'égalité de droite :

$$\mathrm{d}f^{-1}\left(f(x)\right) \cdot \mathrm{d}f(x) = \mathrm{id}$$

et l'autre égalité avec $y = f^{-1}(x)$ donne en fait que pour tout $y \in \mathbb{R}^n$ (f est une bijection) :

$$\mathrm{d}f(y) \cdot \mathrm{d}f^{-1}(f(y)) = \mathrm{id}$$

Par conséquent, on a montré que $\mathrm{d}f(x)$ est donc bien inversible à droite et à gauche (même si en dimension finie, cette dernière inversibilité est équivalente à la première) pour tout $x \in \mathbb{R}^n$.

[(i) $\Longrightarrow$ (ii)] C'est cette implication qui demande de la créativité. Commençons déjà par dire que quitte à considérer $\mathrm{f} = f - f(0)$, on peut supposer que $f(0) = 0$ sans perdre le caractère bijectif. L'idée est la suivante : on cherche à déterminer une application g comme dans le lemme **1.**. Pour cela, on définit $x : \mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ vérifiant

$$\forall t \in \mathbb{R}, \forall y \in \mathbb{R}^n, f(x(t,y)) \overset{(*)}{=} ty$$

La fonction $g : y \longmapsto x(1,y)$ conclura alors. Voilà le plan de démonstration.

1. En dérivant cette égalité, on aboutit à un système différentiel vérifié par x, avec de plus $x(0) = 0$ qui admet une solution maximale sur $[0, T^*[$.

2. Montrons que $T^* = +\infty$ pour pouvoir définir sereinement $g(y) = x(1,y)$.

3. Montrons que g est continue. (C'est là que le théorème de Cauchy à paramètre pointe le bout de son nez).

4. Il ne reste plus qu'à montrer que g vérifie bien les deux points du lemme. La première est vraie par construction, et la surjectivité se montre par **connexité** de $\mathbb{R}^n$.

1. Afin de définir x, on dérive l'égalité $(*)$ par rapport à t :

$$\frac{\partial}{\partial t} f(x(t,y)) = y$$

Par le théorème des fonctions composées,

$$\mathrm{d}f(x(t,y)) \cdot \frac{\partial x}{\partial t}(t,y) = y$$

Notons $\dot{x} = \frac{\partial x}{\partial t}$. Puisque $\mathrm{d}f(x)$ est inversible, pour tout $x \in \mathbb{R}^n$, x vérifie alors

$$\dot{x}(t,y) = [\mathrm{d}f(x(t,y))]^{-1} \cdot y$$

Et puisque $f(0) = 0$, et que $\mathrm{d}f(0)$ est inversible, le théorème d'inversion local donne un caractère bijectif à f au voisinage de 0. On définit alors $x(0,y) = 0$. Par conséquent, pour tout $y \in \mathbb{R}^n$, on définit $x(\cdot, y)$ comme étant une (la) solution de

$$\begin{cases} \dot{x}(t) &= [df(x(t))]^{-1} \cdot y \\ x(0) &= 0 \end{cases}$$

Ce qui s'écrit avec les notations du lemme **2** :

$$\begin{cases} \dot{x}(t) &= F_y(x(t)) \\ x(0) &= 0 \end{cases}$$

Alors pour tout $y \in \mathbb{R}^n$, F_y est C^1 sont localement lipschitzienne, le théorème de Cauchy-Lipschitz fournit une solution maximale $x(\cdot, y)$ sur un intervalle ouvert contenant 0. Notons $I = [0, T^*[$ l'intervalle maximal du côté positif.

2. Sur I, on peut alors vérifier que ce système est équivalent à

$$\forall t \in I, f(x(t,y)) = ty$$

Pour définir g, on souhaite prendre $t = 1$, donc montrer que $1 \in I$. On montre que $T^* = +\infty$ pour cela, à l'aide

du théorème de sortie de tout compact. Ce théorème nous fournit une alternative : soit $T^* = +\infty$, soit la solution explose au voisinage de T^*. Par l'absurde, si $T^* < +\infty$ alors pour tout $y \in \mathbb{R}^n$

$$|x(t,y)| \xrightarrow[t\to T^*]{} +\infty$$

Or,

$$\forall t \in I, |f(x(t,y))| \leqslant T^*|y| < +\infty$$

et donc $x(t,y) \in f^{-1}\left(\bar{B}(0,T^*y]\right)$ est dans l'image réciproque d'un compact par une application propre, donc appartient à un compact, donc est majoré par une constante finie. D'où la contradiction et donc $T^* = +\infty$. On peut alors enfin définir

$$\forall y \in \mathbb{R}^n, g(y) \overset{\text{déf.}}{=} x(1,y)$$

3. Puisque $(y,x) \longmapsto F_y(x)$ est de classe C^1, le théorème de Cauchy à paramètres conclut alors sur le caractère continu de g.

3*. Montrons "à la main" que g est continue. Soit $y_0 \in \mathbb{R}^n$. Si $|y - y_0| \leqslant 1$ alors $|y| \leqslant |y_0| + 1$. Considérons alors

$$K_0 \overset{\text{déf.}}{=} f^{-1}\left(B(0, 1 + |y_0|)\right)$$

qui est un compact de $\mathbb{R}^n$ puisque f est propre. Ainsi, il existe une boule B_0 centrée en 0 telle que $K_0 \subset B_0$. Cette boule est convexe. Par définition, pour tout $y \in B(y_0, 1)$, et $t \in [0,1]$, $x(t,y) \in K_0$ donc pour tout $\lambda \in [0,1]$:

$$\lambda x(t,y_0) + (1-\lambda)x(t,y) \in B_0$$

De plus, par les équations différentielles vérifiées par $x(\cdot, y)$ et $x(\cdot, y_0)$, la mise sous forme intégrale donne

$$\begin{aligned} x(t,y_0) - x(t,y) = & \int_0^t [\mathrm{d}f(x(s,y_0))]^{-1} \cdot y_0 \ \mathrm{d}s \\ + & \int_0^t [\mathrm{d}f(x(s,y))]^{-1} \cdot y \ \mathrm{d}s \end{aligned}$$

On fait apparaître la différence des y, et celle des différentielles pour avoir :

$$\begin{aligned} & x(t,y_0) - x(t,y) \\ = & \int_0^t [\mathrm{d}f(x(s,y_0))]^{-1} \cdot (y_0 - y) \ \mathrm{d}s \\ + & \int_0^t \left([\mathrm{d}f(x(s,y))]^{-1} - [\mathrm{d}f(x(s,y_0))]^{-1}\right) \cdot y \ \mathrm{d}s \end{aligned}$$

On note A la première intégrale et B la deuxième. L'objectif est de pouvoir appliquer le lemme de Grönwall, donc de déterminer des inégalités du type accroissements finis dans chaque intégrale, pour $t \in [0,1]$ et $|y - y_0| \leqslant 1$. Pour A, c'est littéralement l'inégalité des accroissements finis :

$$|A| \leqslant \sup_{z \in B_0} \left\|[\mathrm{d}f(z)]^{-1}\right\| \cdot |y - y_0|$$

On note $M_0 \overset{\text{déf.}}{=} \sup_{z\in B_0} \left\|[\mathrm{d}f(z)]^{-1}\right\|$. Pour B, on utilise la convexité de B_0 que l'on a évoqué en début de partie. Par définition de F_y, on a :

$$B = \int_0^t \left(F_y(x(s,y)) - F_y(x(s,y_0))\right) \ \mathrm{d}s$$

La convexité de B_0 permet alors d'écrire :

$$\begin{aligned} B = & \int_0^t \left[\int_0^1 \mathrm{d}F_y\Big\{x(s,y_0) + \lambda\big(x(s,y) - x(s,y_0)\big)\Big\} \right. \\ & \left. \cdot\Big(x(s,y_0) - x(s,y)\Big) \ \mathrm{d}\lambda \right] \mathrm{d}s \end{aligned}$$

D'où

$$|B| \leqslant \int_0^t \sup_{z\in B_0} \|\mathrm{d}F(z)\| \cdot |x(s,y_0) - x(s,y)| \ \mathrm{d}s$$

On note C_0 la constante $\sup_{z\in B_0} \|\mathrm{d}F(z)\|$. On a alors montré que

$$\begin{aligned} |x(t,y_0) - x(t,y)| \ \leqslant \ & M_0|y - y_0| \\ + \ & C_0 \int_0^t |x(s,y_0) - x(s,y)| \ \mathrm{d}s \end{aligned}$$

On peut alors utiliser le lemme de Grönwall pour aboutir alors à

$$|x(t,y_0) - x(t,y)| \leqslant M_0|y - y_0|e^{C_0 t}$$

Et donc pour tout $y \in B(y_0, 1)$:

$$|g(y) - g(y_0)| \leqslant M_0 e^{C_0}|y - y_0|$$

Donc g est en effet continue.

4. Montrons alors que g vérifie ben les deux hypothèses du lemme **1.**. Par définition de $x(t,y)$, on a bien

$$\forall y \in \mathbb{R}^n, f(g(y)) = y$$

5. Le point délicat est la surjectivité de g. Montrons pour cela que $g(\mathbb{R}^n)$ est ouvert et fermé dans $\mathbb{R}^n$. Puisqu'il est non vide, il suivra par connexité que g est surjective.

• Montrons que $g(\mathbb{R}^n)$ est fermé. On note $(x_k)_k$ une suite de $g(\mathbb{R}^n)$ convergeant vers $x \in \mathbb{R}^n$, et $x_k = g(y_k)$. Alors par la relation avec f et par continuité :

$$f(g(y_k)) = y_k \xrightarrow[k\to+\infty]{} f(x)$$

Notons $y = f(x)$. Alors par continuité de g (et c'est ici que **3.** est essentiel !) :

$$g(y_k) = x_k \xrightarrow[k\to+\infty]{} g(y)$$

D'où $x = g(y)$, donc $x \in g(\mathbb{R}^n)$. D'où le caractère fermé de $g(\mathbb{R}^n)$.

• Montrons que $g(\mathbb{R}^n)$ est ouvert. Soit $x_0 = g(y_0) \in g(\mathbb{R}^n)$. Alors $y_0 = f(x_0)$. Puisque $\mathrm{d}f(x_0)$ est inversible, le théorème d'inversion locale nous fournit deux ouverts U_0 voisinage de x_0 et V_0 voisinage de y_0 tels que $f_{|_{U_0}} : U_0 \longrightarrow V_0$ soit un C^1 difféomorphisme. De plus, $g(y_0) = x_0$. Par continuité de g, il existe W_0 un voisinage de y_0 tel que $g(W_0) \subset U_0$. Quitte à considérer $W_0 \cap V_0$, on suppose que $W_0 \subset V_0$. Par conséquent, $f^{-1}(W_0)$ est un voisinage ouvert de x_0. Si x est dans ce voisinage alors $f(x) \overset{\text{déf.}}{=} y \in W_0$ vérifie $f^{-1}(y) \in U_0$ et $g(y) \in U_0$. Alors

$$\begin{cases} f(f^{-1}(y)) &= y \\ f(g(y)) &= y \end{cases}$$

Puisque f est bijective sur U_0, il suit que

$$g(y) = f^{-1}(y) = x$$

donc la voisinage de x_0 vérifie $f^{-1}(W_0) \subset g(\mathbb{R}^n)$, ce qui montre le caractère ouvert de $g(\mathbb{R}^n)$. On a donc bien montré par connexité que g est surjective. L'application g vérifie donc bien les deux hypothèses du lemme **1.**. Par conséquent, on a bien montré que f est un C^1-difféomorphisme de $\mathbb{R}^n$. □

11 Théorème de Sophie Germain

Référence : Serge Francinou, Hervé Gianella, Serge Nicolas, *Oraux X/ENS, Algèbre 1*

On étudie ce qu'on appelle *le premier cas* du grand théorème de Fermat, théorème qui est à l'origine de toute la théorie des idéaux.

Théorème 11.1 : Théorème de Sophie Germain

Soit $p \geqslant 3$ un nombre premier tel que $q \overset{\text{déf.}}{=} 2p+1$ soit lui aussi premier. Alors l'équation d'inconnues $(x, y, z) \in \mathbb{Z}^3$

$$\begin{cases} x^p + y^p + z^p &= 0 \\ xyz &\not\equiv 0 \ [p] \end{cases}$$

n'admet aucune solution.

Remarque : Le cas $p = 2$ est traité dans le tome I du vrac, comme le cas $n = 4$. L'idée centrale est la suivante

$$x^p + y^p = \prod_{i=0}^{p-1} (x + \zeta_i y)$$

avec ζ_i les racines p-ème de l'unité.

Démonstration : Le plan est le suivant : on raisonne par l'absurde en supposant qu'il existe un triplet (x, y, z) solution.

1. On réduit d'abord le problème en supposant que x, y, z sont premiers entre eux, ce qui implique dans notre cas qu'ils soient premiers entre eux dans leur ensemble ;

2. On étudie les puissances p-ème modulo $2p+1 = q$: ils ne valent que ± 1 ou 0. Ainsi, q divise une et une seule des inconnues ;

3. Par un argument de factorisation du même type que de la remarque, on montre qu'il existe a, b, c entiers tels que $x + y = c^p$, $x + z = b^p$ et $y + z = a^p$;

4. L'étude des puissances déterminées en **3.** conjuguée à celle des puissances p-ème en **2.** conclut en une contradiction.

Démontrons alors le théorème.

1. Si d désigne le PGCD de x, y, z alors

$$\begin{cases} \left(\frac{x}{d}\right)^p + \left(\frac{y}{d}\right)^p + \left(\frac{z}{d}\right)^p &= 0 \\ \frac{xyz}{d^3} &\not\equiv 0 \ [p] \end{cases}$$

On peut alors supposer que x, y, z soient premiers entre eux. Montrons qu'ils sont alors premiers deux à deux. Si $x \wedge y > 1$, on se donne p_0 premier qui divise x et y. Ainsi, p_0 divise $x^p + y^p$, donc z^p, donc z. Donc p_0 divise x, y et z, qui sont premiers entre eux, d'où une contradiction. Donc x, y, z sont donc bien deux à deux premiers entre eux.

2. Soit $m \in \mathbb{Z}$. Déjà, si m est divisible par $2p+1 = q$ alors $m^p \equiv 0 \ [q]$. Supposons que m ne soit pas divisible par q. Par le petit théorème de Fermat,

$$m^{2p} \equiv 1 \ [q]$$

Puisque q est premier, il suit que $m^p \equiv \pm 1 \ [q]$. Montrons alors que l'une des inconnues est divisible par q. Si aucune de ces inconnues n'est divisible par q alors x^p, y^p, z^p sont congrus à ± 1 modulo q et donc leur somme est congrue à ± 1 ou à ± 3 modulo q, ce qui contredit le fait que $x^p + y^p + z^p = 0$ et $q \geqslant 5$. Supposons que l'inconnue divisible par q est x. Puisque x est premier avec y et z alors q ne divise pas y et z, donc q divise x, et seulement x parmi les inconnues.

3. Montrons qu'il existe a et α des entiers tels que $y + z = a^p$ et

$$\sum_{k=0}^{p-1} (-z)^{p-1-k} y^k = \alpha^p$$

Pour cela, on constate la factorisation suivante :

$$(-x)^p = (y+z) \sum_{k=0}^{p-1} y^k (-z)^{p-1-k}$$

Et il nous suffit alors de montrer que $y + z$ et $\sum_{k=0}^{p-1} y^k (-z)^{p-1-k}$ sont premiers entre eux. Encore une fois, on procède par l'absurde : considérons p_1 un premier qui divise ces deux nombres. Alors p_1^2 divise x^p, donc p_1 divise x. De plus, puisque

$$y + z \equiv 0 \ [q]$$

alors

$$0 \equiv \sum_{k=0}^{p-1} y^k (-z)^{p-1-k} \equiv p y^{p-1} \ [p_1]$$

Or, par hypothèse, y et p sont premiers entre eux, donc le lemme de Gauss nous donne l'alternative suivante : soit p_1 divise p, soit p_1 divise y^{p-1}. Si p_1 divise p, les nombres en présence étant premiers, on a alors $p = p_1$, et donc $p_1 = p$ divise x, ce qui contredit le fait que p ne divise pas xyz. Quand à la deuxième alternative, si p_1 divise y^{p-1} alors p_1 divise x et y, ce qui contredit le caractère premier de x et y. Il suit donc que $y + z$ et $\sum_{k=0}^{p-1} y^k (-z)^{p-1-k}$ sont premiers entre eux. Par conséquent, il existe a et α entiers tels que $y + z = a^p$ et

$$\sum_{k=0}^{p-1} (-z)^{p-1-k} y^k = \alpha^p$$

Par symétrie de x, y, z, on a aussi l'existence de b et c entiers tels que $b^p = x + z$ et $c^p = x + y$.

4. Concluons. Montrons que α^p contredit le point **2.**. On a :

$$\alpha^p = \sum_{k=0}^{p-1} y^k (-z)^{p-1-k}$$

Or,

$$y \equiv y + x \ [q] \equiv c^p \ [q]$$

Or, q ne divise pas y, donc ne divise pas c. Par la caractérisation des carrés de **2.**, il suit que

$$y \equiv \pm 1 \ [q]$$

De même, $z \equiv z + x \ [q]$ donne cette fois-ci que $z \equiv \pm 1 \ [q]$. De plus, modulo q,

$$b^p + c^p = 2x + a^p \equiv a^p \ [q]$$

De plus,

$$b^p + c^p = y + z \in \{\pm 2, 0\} \ [q]$$

Si q ne divise pas a, alors $a^p \equiv \pm 1 \ [q]$, mais cela contredit

$$a^p \in \{\pm 2, 0\} \ [q]$$

donc q divise a : $a \equiv 0 \ [q]$. D'où

$$y + z \equiv 0 \ [q]$$

Par conséquent, puisque $p - 1$ est paire :

$$\alpha^p \equiv p y^{p-1} \ [q] \equiv p \ [q]$$

Mais $q = 2p + 1$, donc p n'est pas congru à 1 ou -1 ou 0 modulo p, donc α^p non plus. Cela contredit ouvertement la description **2.**. Cela conclut le théorème de Sophie Germain. □

12 Variante du théorème des deux carrés

Référence : Moi-même, une fois n'est pas coutûme

Théorème 12.1 : Variante du théorème des deux carrés

Considérons l'ensemble suivant :

$$\Xi \overset{\text{déf.}}{=} \left\{ n \in \mathbb{N}, \exists a, b \in \mathbb{Z}, n = a^2 + 2b^2 \right\}$$

Alors les assertions suivantes sont équivalentes, pour $n \in \mathbb{N}$:

(i) $n \in \Xi$;

(ii) Pour tout premier p tel que p soit congru à $-3, -1$ ou à 5 modulo 16, la valuation p-adique de n est paire : $\nu_p(n) \in 2\mathbb{N}$.

Il suffit en fait de reprendre la démonstration du tome III pour les deux carrés en étudiant cette fois le fait que -2 soit un carré ou non modulo p. C'est l'objet du lemme suivant, démontré dans *Exercices d'algèbre pour l'agrégation*, de nos très chers Serge et Hervé.

Lemme 12.1 : Quadraticité de -2 modulo p

Soit $p \geqslant 3$ un nombre premier. Les assertions suivantes sont équivalentes :

(i) -2 n'est pas un carré modulo p ;

(ii) $p \equiv -3, -1, 5\ [16]$.

Démonstration du lemme : **1.** Rappelons que $x \in \mathbb{Z}$ est un carré modulo p si et seulement si

$$x^{\frac{p-1}{2}} \equiv 1\ [p]$$

Le lecteur apeuré de ne plus se souvenir d'une preuve exacte de ce fait aura son bonheur dans le tome III, dans le section liée au théorème des deux carrés.

2. Montrons que 2 vérifie

$$2^{\frac{p-1}{2}} \equiv (-1)^{\frac{p^2-1}{8}}\ [p].$$

Constatons au passage que le fait que $\frac{p^2-1}{8}$ soit entier n'est pas évident. Pour cela, on dit que $\alpha \in \mathbb{Z}$ est de résidu minimal α' modulo p si $\alpha \equiv \alpha'\ [p]$ et $\alpha' \in \left]\frac{-p}{2}, \frac{p}{2}\right]$. On considère alors μ le nombre de résidus minimaux strictement négatifs de l'ensemble $\{2, 4, \cdots, p-1\}$. On note

$$\lambda \overset{\text{déf.}}{=} \frac{p-1}{2} - \mu.$$

Si A désigne l'ensemble des résidus minimaux de $\{2, 4, \cdots, p-1\}$ alors on dispose de la partition suivante :

$$A = \bigsqcup_{i=1}^{\lambda} \{r_i\} \sqcup \bigsqcup_{j=1}^{\mu} \{-r'_j\}$$

où les r_i sont positifs ou nuls et les r'_j sont strictement positifs. Montrons qu'ils sont deux à deux distincts. Les $(r_i)_i$ sont deux à deux distincts car ils ne sont pas congrus modulo p. Idem pour les $(r'_j)_j$. Enfin, si $r_i = r'_j$, il existe a, b dans $\left\{1, 2, \cdots, \frac{p-1}{2}\right\}$ tels que

$$2a \equiv r_i \equiv r_j \equiv -2b\ [p]$$

Alors

$$2(a+b) \equiv 0\ [p]$$

Puisque p est premier impair, il suit que $a + b \equiv 0\ [p]$, ce qui contredit le fait que $0 < a + b < p$. Donc la partition annoncée de A en est bien une. Par conséquent, A est inclus dans $\left\{1, 2, \cdots, \frac{p-1}{2}\right\}$, et est de cardinal $\frac{p-1}{2}$, donc est égal à cet ensemble. Ainsi,

$$\prod_{i=1}^{\lambda} r_i \prod_{j=1}^{\mu} r_j = \left(\frac{p-1}{2}\right)!$$

Or, par construction :

$$\prod_{k=1}^{p-1} (2k) \equiv \prod_{i=1}^{\lambda} r_i \prod_{j=1}^{\mu} (-r_j)\ [p]$$

D'où

$$2^{\frac{p-1}{2}} \left(\frac{p-1}{2}\right)! \equiv (-1)^{\mu} \left(\frac{p-1}{2}\right)!\ [p]$$

Puisque p ne divise pas $\left(\frac{p-1}{2}\right)!$, il suit que

$$2^{\frac{p-1}{2}} \equiv (-1)^{\mu}\ [p]$$

Pour conclure, il ne reste plus qu'à trouver un bon reste modulo 2 de μ pour conclure. Si $p \equiv 1\ [p]$ alors $\lambda = \frac{p-1}{4}$ et

$$\mu = \frac{p-1}{4} \overset{\text{déf.}}{=} \mu_1.$$

Si $p \equiv 3\ [4]$ alors $\lambda = \left\lfloor \frac{p-1}{4} \right\rfloor = \frac{p-3}{4}$ et

$$\mu = \frac{p+1}{4} \overset{\text{déf.}}{=} \mu_2$$

Or, $2\mu_1\mu_2 = \frac{p^2-1}{8} \in \mathbb{N}$ a même parité que μ_1 lorsque $p \equiv 1\ [4]$ et que μ_2 si $p \equiv 3\ [4]$, donc est de même parité que μ. On a alors la conclusion souhaitée :

$$2^{\frac{p-1}{2}} \equiv (-1)^{\frac{p^2-1}{8}}\ [p]$$

3. Ce point **2.** étant connu (c'est un résultat considéré comme classique malgré sa démonstration non élémentaire), concluons ce lemme. Utilisons la caractérisation de **1.**.

$$(-2)^{\frac{p-1}{2}} \equiv (-1)^{\frac{(p-1)(p+5)}{8}} \ [p]$$

Une étude de $(x-1)(x+5)$ lorsque $x \in {}^{\mathbb{Z}}/_{16\mathbb{Z}}$ permet de conclure que -2 est un carré modulo p si et seulement si

$$p \equiv -7, -5, 1, 3 \ [16]$$

donc n'est pas un carré si et seulement si p est congrue à toutes ses valeurs possibles autre que celles-ci modulo 16, d'où le résultat. □

Une autre vérification à faire, c'est de vérifier que cette fois-ci, c'est non pas $\mathbb{Z}[\mathbf{i}]$ mais $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$ qui est en jeu et et lui aussi euclidien, par le même argument de division sur $\mathbb{C}$, puis par le fait qu'une demi-diagonale du rectangle de côtés 1 et $\sqrt{2}$ soit de longueur strictement inférieure à 1.

Lemme 12.2 : Un anneau de GAUSS

L'anneau $\left(\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right], +, \cdot\right)$ est un anneau euclidien de fonction euclidienne sa *norme*, donnée par :

$$\forall z \in \mathbb{Z}\left[\mathbf{i}\sqrt{2}\right], N(z) \overset{\text{déf.}}{=} z\bar{z} = a^2 + 2b^2$$

De plus, $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]^\times = \{\pm 1\}$.

Démonstration du lemme : Le schéma de démonstration est le suivant :

- N est une fonction multiplicative : $N(zz') = N(z)N(z')$.
- On détermine les éléments inversibles de $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$ grâce au caractère multiplicatif de N en se ramenant sur $\mathbb{N}$: $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]^\times = \{\pm 1\}$.
- On fait la division dans $\mathbb{C}$ de z par $\omega \neq 0$. Il existe un élément $q \in \mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$ tel que $\left|\frac{z}{\omega} - q\right| \leqslant \frac{\sqrt{3}}{2} < 1$. On en déduit que $z = \omega q + r$ est une division euclidienne avec $r = z - q\omega$ de norme inférieure ou égal à celle de ω. □

Maintenant que ces propriétés sont acquises, il ne reste plus qu'à copier-coller la démonstration du théorème des deux carrés en changeant des petits détails et en faisant gaffe aux erreurs d'alignement dans la compilation du code LaTeX(ah oui, et en changeant les $\imath$ et $\mathbf{i}$, histoire de ne plus être en harmonie avec les notations introduites).

Démonstration : **1.** Observons d'abord que Ξ est l'image de N, la norme sur $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$. En conséquence, Ξ est stable par multiplication.

2. Traitons le cas des nombres premiers. Pour cela, montrons que les propriétés suivantes sont équivalentes, pour p un nombre premier :

(i) p est irréductible dans $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$;
(ii) $p \equiv -3, -1, 5 \ [16]$;
(iii) p n'est **pas** dans Ξ.

(i) $\iff$ **(ii)** Utilisons des isomorphismes classiques entre anneaux quotients. Notons pour tout anneau A que la notation ${}^{A}/_{\langle x \rangle}$ sous-entend que $\langle x \rangle$ est un idéal de A, notation qui sera donc indépendante de A malgré sa dépendance en pratique. Par définition de $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$, avec $\mathbf{i}\sqrt{2}$ la classe de X :

$${}^{\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]}/_{\langle p \rangle} \simeq {}^{\left({}^{\mathbb{Z}[X]}/_{\langle X^2+2 \rangle}\right)}/_{\langle p \rangle}$$

On « réduit au même idéal » :

$${}^{\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]}/_{\langle p \rangle} \simeq {}^{\mathbb{Z}[X]}/_{\langle p, X^2+2 \rangle}$$

Et on introduit cette fois p au numérateur pour avoir finalement

$${}^{\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]}/_{\langle p \rangle} \simeq {}^{\mathbb{F}_p[X]}/_{\langle X^2+2 \rangle}$$

Par conséquent, p est irréductible dans $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$ si et seulement si ${}^{\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]}/_{\langle p \rangle}$ est intègre si et seulement si ${}^{\mathbb{F}_p[X]}/_{\langle X^2+2 \rangle}$ est intègre si et seulement si X^2+2 est irréductible sur ${}^{\mathbb{Z}}/_{p\mathbb{Z}}$ si et seulement si -2 n'admet aucun carré modulo p si et seulement $p \equiv -3, -1, 5 \ [16]$ par le lemme **1.**.

(i) $\implies$ **(iii)** Par contraposée, si $p \in \Xi$ alors $p = \left(a + \mathbf{i}\sqrt{2}b\right)\left(a - \mathbf{i}\sqrt{2}b\right)$. Puisque p est premier, $a + \mathbf{i}\sqrt{2}b$ et $a - \mathbf{i}\sqrt{2}b$ ne sont pas dans les inversibles de $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$, donc p n'est pas irréductible dans $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$.

(iii) $\implies$ **(i)** Par contraposée, si

$$p = \left(a + \mathbf{i}\sqrt{2}b\right)\left(c + \mathbf{i}\sqrt{2}d\right)$$

avec les deux facteurs n'appartenant pas à $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]^\times$. Alors

$$p^2 = N\left(a + \mathbf{i}\sqrt{2}b\right) N\left(c + \mathbf{i}\sqrt{2}d\right)$$

avec chacune de ces normes supérieures ou égales à 2. Par caractère premier de p, il suit que $p = N\left(a + \mathbf{i}\sqrt{2}b\right)$, donc $p \in \Xi$.

3. Concluons sur le théorème grâce à la décomposition en facteurs premiers.

• Si $\nu_p(n) \in 2\mathbb{N}$ pour les nombres premiers congrus à $-3, -1$ ou à 5 modulo 16 alors n s'écrit comme :

$$n = 2^{\nu_p(n)} \cdot \prod_{\substack{p\in\mathbb{P}\\ p\equiv 1\ [16]\\ p\equiv 3\ [16]\\ p\equiv -7\ [16]\\ p\equiv -5\ [16]}} p^{\nu_p(n)} \cdot \prod_{\substack{p\in\mathbb{P}\\ p\equiv -3\ [16]\\ p\equiv -1\ [16]\\ p\equiv 5\ [16]}} p^{\nu_p(n)}$$

Puisque $2 = 0^2 + 2 \cdot 1^2$, $2 \in \Xi$, donc le premier facteur appartient à Ξ par stabilité par produit de Ξ. Le deuxième facteur appartient à Σ par **2.**. Quand au troisième :

$$\prod_{\substack{p\in\mathbb{P}\\ p\equiv -3,-1,5\ [16]}} p^{\nu_p(n)} = \left(\prod_{\substack{p\in\mathbb{P}\\ p\equiv -3,-1,5\ [16]}} p^{\frac{\nu_p(n)}{2}}\right)^2$$

donc le deuxième facteur appartient aussi à Ξ. Par stabilité multiplicative de Ξ, montrée en **1.**, il suit que $n \in \Xi$.

• Montrons par récurrence forte sur $n \geqslant 2$ la propriété suivante :

$$n \in \Xi \Longrightarrow \begin{array}{c}\forall p \in \mathbb{P}\\ p \equiv -3, -1, 5\ [16]\end{array}, \nu_p(n) \in 2\mathbb{N}$$

Pour $n = 2$, n est premier et ne comporte pas de nombre premier dans sa décomposition en facteurs premiers congrus à -3, -1 ou 5 modulo 16. Par propriété sur l'ensemble vide, il suit que la propriété est vraie au rang $n = 2$.

Soit $n \geqslant 3$. Supposons que pour tout $m \leqslant n-1$, $m \in \Xi$ implique que pour les premiers congrus à 3 modulo 4, $\nu_p(m)$ est pair. Montrons ce résultat pour n. Supposons que $n \in \Xi$. Alors

$$n = a^2 + 2b^2$$

Si p premier ne divise pas n, il est alors vrai que $\nu_p(n) = 0$ est pair. Supposons que p divise n avec p congru à $-3, -1$ ou 5 modulo 16. Alors p divise dans $\mathbb{Z}\left[\mathbf{i}\sqrt{2}\right]$ l'élément $a + \mathbf{i}\sqrt{2}b$ (ou $a - \mathbf{i}\sqrt{2}b$, mais $\bar{p} = p$). Ainsi, p divise a et b dans $\mathbb{Z}$. Il suit alors que

$$n = p^2\left(\left(\frac{a}{p}\right)^2 + 2 \cdot \left(\frac{b}{p}\right)^2\right)$$

Ainsi, $\frac{n}{p^2} \in \Xi$, avec $\frac{n}{p^2} < n$. Par hypothèse de récurrence, $\nu_p\left(\frac{n}{p^2}\right)$ est pair. Il suit alors que

$$\nu_p(n) = 2 + \nu_p\left(\frac{n}{p^2}\right) \in 2\mathbb{N}$$

Ce qui conclut la récurrence. □

13 Problème de Laplace sur le disque

Références :

— *Partial Diffenrential Equation*, Steven Krantz ;

— *Real and Complex Analysis*, Walter Rudin

Théorème 13.1 : Problème de Laplace sur le cercle

Soient $D = B(0,1) \subset \mathbb{C}$, et f continue sur ∂D. Alors l'équation

$$\left\{\begin{array}{rcl} \Delta u & = & 0 \quad \text{sur } D \\ u_{|\partial D} & = & f \end{array}\right.$$

admet une unique solution $u \in \mathcal{C}^0(\bar{D}) \cap \mathcal{C}^2(D)$. Elle est donnée par

$$u\left(re^{\mathbf{i}\theta}\right) = \left\{\begin{array}{ll} \displaystyle\int_0^{2\pi} f\left(e^{\mathbf{i}(\theta-t)}\right) P_r(t)\ \mathrm{d}t & \text{pour} \quad 0 \leqslant r < 1 \\ f\left(e^{\mathbf{i}\theta}\right) & \text{si} \quad r = 1 \end{array}\right.$$

où P_r est le *noyau de Poisson* :

$$\forall t \in \mathbb{R}, P_r(t) \overset{\text{déf.}}{=} \frac{1}{2\pi} \cdot \frac{1-r^2}{1-2r\cos(t)+r^2}$$

Démonstration :

1. Avant de démontrer proprement le théorème, donnons l'intuition derrière. Supposons que f soit une fonction e_n trigonométrique. Alors une solution au problème est donnée par :

$$\forall n \in \mathbb{Z}, u_n(z) = \begin{cases} z^n & \text{si} \quad n \in \mathbb{N} \\ \bar{z}^n & \text{si} \quad n \leqslant 0 \end{cases}$$

Cette fonction est bien harmonique : c'est immédiat quand on sait que $\Delta = 4\frac{\partial}{\partial z}\frac{\partial}{\partial \bar{z}}$. La fonction u_n se réécrit en coordonnées polaires plus simplement en :

$$\forall n \in \mathbb{Z}, u_n\left(re^{\mathrm{i}\theta}\right) = r^{|n|}e^{\mathrm{i}n\theta}$$

De plus, l'équation est linéaire. Ainsi, si f est un polynôme trigonométrique, la solution sera combinaison linéaire des u_n. Supposons ici que f est développable en série de Fourier (les hypothèses du théorème ne nous permettent pas d'affirmer que cela est le cas). Alors on cherche une solution u sous la forme, pour $r \in [0,1[$:

$$u\left(re^{\mathrm{i}\theta}\right) = \sum_{n\in\mathbb{Z}} c_n(f) r^{|n|} e^{\mathrm{i}n\theta}$$

2. • Cette somme est bien définie : le lemme de Riemann-Lebesgue assure que $c_n(f)$ tend vers 0, puisqu'on a supposé f continue, et $r < 1$. Montrons que u est une solution du problème. Le problème est ici de montrer la continuité au bord, puisque cette somme n'a aucune raison de converger pour $r = 1$. On va mettre u sous une forme intégrale, qui se rapprochera d'un produit de convolution avec un noyau, que l'on nommera *noyau de Poisson*. Par définition de $c_n(f)$:

$$u\left(re^{\mathrm{i}\theta}\right) = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} r^{|n|} e^{\mathrm{i}n(\theta-t)} f\left(e^{\mathrm{i}t}\right)\,\mathrm{d}t$$

Or,

$$\sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} \left| r^{|n|} e^{\mathrm{i}n(\theta-t)} f\left(e^{\mathrm{i}t}\right)\right|\,\mathrm{d}t \leqslant \frac{2\|f\|_\infty}{1-r} < +\infty$$

Par le théorème de Fubini,

$$u\left(re^{\mathrm{i}\theta}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[\sum_{n\in\mathbb{Z}} r^{|n|} e^{\mathrm{i}n(\theta-t)}\right] f\left(e^{\mathrm{i}t}\right)\,\mathrm{d}t$$

On note alors

$$P_r(t) \overset{\text{déf.}}{=} \sum_{n\in\mathbb{Z}} r^{|n|} e^{\mathrm{i}nt}$$

le *noyau de Poisson*. Exprimons P_r sous forme explicite. Deux calculs de sommes géométriques permettent de conclure sur la forme annoncée de P_r :

$$P_r(t) = \frac{1-r^2}{1-2r\cos t + r^2}$$

Remarquons au passage que

$$P_r(t) = \Re\left[\frac{1+re^{\mathrm{i}t}}{1-re^{\mathrm{i}t}}\right]$$

Par conséquent, u s'exprime aussi sous forme intégrale sur D par :

$$u\left(re^{\mathrm{i}\theta}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\left(e^{\mathrm{i}t}\right) P_r(\theta-t)\,\mathrm{d}t$$

• Vérifions alors effectivement que u est la solution au problème. Pour l'harmonicité, deux options : on peut contempler la formulation en série de u et utiliser un théorème d'interversion entre dérivées complexes et somme, ou utiliser la petite remarque astucieuse sur P_r pour aboutir à :

$$u\left(re^{\mathrm{i}\theta}\right) = \Re\left[\frac{1}{2\pi}\int_0^{2\pi} \frac{1+re^{\mathrm{i}(\theta-t)}}{1-re^{\mathrm{i}(\theta-t)}}\, f\left(e^{\mathrm{i}t}\right)\,\mathrm{d}t\right]$$

Soit encore

$$u(z) = \Re\left[\frac{1}{2\pi}\int_0^{2\pi} \frac{e^{\mathrm{i}t}+z}{e^{\mathrm{i}t}-z}\, f\left(e^{\mathrm{i}t}\right)\,\mathrm{d}t\right]$$

La majoration pour $|z| = r$ donne alors

$$\left|\frac{e^{\mathrm{i}t}+z}{e^{\mathrm{i}t}-z}\right| \leqslant \frac{1+r}{1-r}$$

permet alors de conclure au caractère holomorphe de l'intégrale fonction de z, et donc au caractère harmonique de u sur D.

• Montrons que u est continue sur $\bar{D}$. u l'est sur D, montrons qu'elle l'est au voisinage de ∂D. On le fait en deux temps : on montre d'abord une continuité radiale pour conclure sur la continuité.

▷ Pour la continuité radiale, on se donne $e^{\imath\theta} \in \partial D$. Montrons que

$$\Delta_r \overset{\text{déf.}}{=} \left|u\left(re^{\mathrm{i}\theta}\right) - f\left(e^{\mathrm{i}\theta}\right)\right| \xrightarrow[r\to 1^-]{} 0$$

Soit $\varepsilon > 0$. Par théorème de Heine, f est uniformément continue : on note $\delta > 0$ le module de continuité associé. Constatons que

$$\int_{-\pi}^{\pi} |P_r(t)|\,\mathrm{d}t = \int_{-\pi}^{\pi} P_r(t)\,\mathrm{d}t = 2\pi$$

P_r est en effet positive par son écriture explicite, et d'intégrale 1 par son expression en série. Ainsi,

$$\Delta_r = \frac{1}{2\pi}\left|\int_{-\pi}^{\pi}\left[f\left(e^{\mathbf{i}(\theta-t)}\right)-f\left(e^{\mathbf{i}(\theta)}\right)\right]P_r(t)\,\mathrm{d}t\right|$$

D'où, en faisant intervenir δ :

$$\Delta_r \leqslant \frac{\varepsilon}{2\pi}\int_{-\delta}^{\delta} P_r(t)\,\mathrm{d}t + \frac{1}{2\pi}\int_{\delta}^{2\pi} P_r(t)\,\mathrm{d}t$$

Occupons-nous de la deuxième intégrale. Soit $t \in [\delta, 2\pi]$. Alors en complétant un carré :

$$P_r(t) = \frac{1-r^2}{(1-r\cos t)^2 + r^2\sin^2 t}$$

D'où

$$P_r(t) \leqslant \frac{1-r^2}{(1-r\cos t)^2}$$

Or,

$$\cos t \leqslant 1 - \frac{t^2}{2}$$

et $t \longmapsto \frac{1}{(1-x)^2}$ est croissante sur $[0,1[$. D'où

$$P_r(t) \leqslant \frac{1-r^2}{\left(1-r+\frac{r\delta^2}{2}\right)^2}$$

D'où pour $t \in [\delta, 2\pi]$:

$$P_r(t) \leqslant \frac{4\left(1-r^2\right)}{r^2\delta^4}$$

Ainsi,

$$\Delta_r \leqslant \varepsilon + \frac{8\pi(1-r^2)}{r^2\delta^4}$$

Le deuxième terme est plus petit que ε pour r assez proche de 1, disons sur $[r_0, 1[$. Par conséquent, pour tout $r \in [r_0, 1[$:

$$\Delta_r \leqslant 2\varepsilon$$

D'où

$$\Delta_r \xrightarrow[r\to 1^-]{} 0$$

▷ Concluons. Si $e^{\mathbf{i}\theta} \in D$, et $z = re^{\mathbf{i}t} \in D$ alors

$$\begin{aligned}\left|u\left(re^{\mathbf{i}t}\right) - f\left(e^{\mathbf{i}\theta}\right)\right| \leqslant & \left|u\left(re^{\mathbf{i}t}\right) - f\left(e^{\mathbf{i}t}\right)\right| \\ + & \left|f\left(e^{\mathbf{i}t}\right) - f\left(e^{\mathbf{i}\theta}\right)\right|\end{aligned}$$

Pour $|\theta - t| < \delta$, l'uniforme continuité de f majore le deuxième terme. Pour $r \in [r_0, 1[$, la continuité radiale majore le premier. D'où

$$\left|u\left(re^{\mathbf{i}t}\right) - f\left(e^{\mathbf{i}\theta}\right)\right| \leqslant 2\varepsilon$$

La continuité de u est alors montrée pour $\left|z - e^{\mathbf{i}\theta}\right| < \min\{\delta, 1-r_0\}$. On a donc bien une solution à notre problème !

3. Montrons l'unicité de la solution. Soient u_1 et u_2 deux solutions. Alors $u = u_1 - u_2$ vérifie le problème de Dirichlet avec des conditions au bord nulles. Supposons que u soit non nul : plus précisément qu'il existe $z_0 \in D$ tel que $u(z_0) > 0$. Soit $0 < \varepsilon < u(z_0)$. On définit alors

$$g(z) = u(z) + \varepsilon|z|^2$$

D'une part,

$$g(z_0) \geqslant u(z_0) > \varepsilon$$

Or, $g_{|\partial D} = \varepsilon$, donc g admet son maximum sur $\bar{D}$ au sein de l'ouvert D, en $z_1 \in D$. En particulier,

$$\frac{\partial^2 g}{\partial x^2}(z_1), \frac{\partial^2 g}{\partial y^2}(z_1) \leqslant 0$$

Ce qui contredit

$$\Delta g(z_1) = 4\varepsilon$$

D'où $u \leqslant 0$. Par symétrie, il suit que $u = 0$. D'où l'unicité de la solution. □

14 Théorème de Rademacher

Référence : Aucune, le développement provient d'une leçon de cette année...

Proposition 14.1 : Théorème de RADEMACHER :

Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$. Les applications suivantes sont équivalentes :

(i) f est lipschitzienne ;

(ii) Il existe une fonction $g \in L^\infty(\mathbb{R})$ telle que

$$\forall x, y \in \mathbb{R}, f(x) - f(y) = \int_y^x g(t)\ \mathrm{d}t$$

On admet le lemme suivant :

Lemme 14.1 : Formes linéaires continues sur L^1

Si $T \in L^1(\mathbb{R})'$ alors

$$\exists g \in L^\infty(\mathbb{R}), \forall \varphi \in L^1(\mathbb{R}), \langle T, \varphi\rangle = \int_{\mathbb{R}} g(t)\varphi(t)\ \mathrm{d}t$$

Démonstration : [(ii) $\Longrightarrow$ (i)] Si f vérifie cette égalité, alors pour tout $x, y \in \mathbb{R}$:

$$|f(x) - f(y)| \leqslant \|g\|_{L^\infty}|x - y|$$

donc f est bien lipschitzienne.

[(i) $\Longrightarrow$ (ii)] Supposons que f soit M lipschitzienne. L'idée est de considérer la dérivée de f au sens des distributions : on note

$$\xi : \begin{pmatrix} \mathcal{D}(\mathbb{R}) & \longrightarrow & \mathbb{R} \\ \varphi & \longmapsto & -\displaystyle\int_{\mathbb{R}} \varphi'(x)f(x)\ \mathrm{d}x \end{pmatrix}$$

1. ξ est une forme linéaire continue sur $\mathcal{D}(\mathbb{R})$ pour la norme $\|\cdot\|_{L^1}$, donc se prolonge sur L^1 continument ;
2. Si $g \in L^\infty(\mathbb{R})$ désigne l'application associée à T, on considère $h : x \longmapsto \int_0^x g$, et on montre que $h' = g$ au sens des distributions ;
3. Puisque $f' = h'$ au sens des distributions, il suit que $f = h + c$ au sens des fonctions continues, d'où l'égalité.

1. Montrons qu'il existe une constante $M > 0$ telle que

$$\forall \varphi \in \mathcal{D}(\mathbb{R}), |\langle \xi, \varphi\rangle| \leqslant M\|\varphi\|_{L^1}$$

Pour cela, puisqu'un calcul direct ne nous aide pas, on définit une suite qui approche la dérivée de φ :

$$\psi_n(x) \overset{\text{déf.}}{=} n\left[\varphi\left(x + \frac{1}{n}\right) - \varphi(x)\right]$$

Alors ψ_n converge simplement vers φ'. De plus, ψ_n est de support inclus dans $\mathrm{Supp}(\varphi) + \left[\frac{-1}{n}, \frac{1}{n}\right]$. Ainsi, pour tout $x \in \mathbb{R}$ et $n \in \mathbb{N}$:

$$|\psi_n(x)f(x)| \leqslant 2\|\varphi\|_{L^1}\|f\|_{L^\infty}\mathbf{1}_{\mathrm{Supp}(\varphi)+[-1,1]}$$

Par théorème de convergence dominée :

$$\int_{\mathbb{R}} \psi_n(x)f(x)\ \mathrm{d}x \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}} \varphi'(x)f(x)\ \mathrm{d}x$$

On exprime autrement cette suite d'intégrales. On fait des petits changements de variables :

$$\int_{\mathbb{R}} \psi_n(x)f(x)\ \mathrm{d}x = n\int_{\mathbb{R}} \left[f\left(x + \frac{1}{n}\right) - f(x)\right]\varphi(x)\ \mathrm{d}x$$

D'où

$$\left|\int_{\mathbb{R}} \psi_n(x)f(x)\ \mathrm{d}x\right| \leqslant M\|\varphi\|_{L^1}$$

D'où la continuité de ξ pour la norme L^1. De plus, $\mathcal{D}(\mathbb{R})$ est dense dans L^1 pour sa norme associée. Par théorème de prolongement des applications linéaires continues, il suit qu'il existe $T \in L^1(\mathbb{R})'$ qui prolonge ξ. Par le lemme, il existe $g \in L^\infty$ telle que

$$\forall \varphi \in L^1, \langle T, \varphi\rangle = \int_{\mathbb{R}} g\varphi$$

En restriction à $\mathcal{D}(\mathbb{R})$, il suit que $g = f'$ au sens des distributions.

2. Montrons que $h' = g$ dans $\mathcal{D}'(\mathbb{R})$, en utilisant le théorème de Fubini. Soit $\varphi \in \mathcal{D}(\mathbb{R})$. Par définition :

$$\langle h', \varphi\rangle = -\int_{\mathbb{R}} h(x)\varphi'(x)\ \mathrm{d}x$$

Puis, par définition de de h :

$$\langle h', \varphi\rangle = -\int_{\mathbb{R}} \left[\int_0^x g(t)\ \mathrm{d}t\right]\varphi'(x)\ \mathrm{d}x$$

On découpe en deux intégrales, selon le signe de x :

$$\begin{aligned}\langle h', \varphi\rangle = & -\int_0^{+\infty}\left[\int_0^x g(t)\ \mathrm{d}t\right]\varphi'(x)\ \mathrm{d}x \\ & + \int_{-\infty}^0\left[\int_x^0 g(t)\ \mathrm{d}t\right]\varphi'(x)\ \mathrm{d}x\end{aligned}$$

On souhaite appliquer le théorème de Fubini, pour cela on introduit $R > 0$ tel que le support de φ' soit inclus dans $[-R, R]$. Alors

$$\int_0^{+\infty}\int_0^x \left|g(t)\ \varphi'(x)\ \mathbf{1}_{[0,x](t)}\right| \mathrm{d}t\,\mathrm{d}x \leqslant R\|g\|_{L^\infty}\|\varphi'\|_{L^1} < \infty$$

De même pour la deuxième intégrale. Par Fubini, il suit que

$$\begin{aligned}\langle h', \varphi\rangle = &-\int_0^{+\infty}\left[\int_t^{+\infty}\varphi'(x)\ \mathrm{d}x\right] g(t)\ \mathrm{d}t\\ &+\int_{-\infty}^{0}\left[\int_{-\infty}^{t}\varphi'(x)\ \mathrm{d}x\right] g(t)\ \mathrm{d}t\end{aligned}$$

En intégrant, il suit alors :

$$\langle h', \varphi\rangle = \int_{\mathbb{R}}\varphi(x)g(x)\ \mathrm{d}x = \langle g, \varphi\rangle$$

D'où $h' = g$ au sens des distributions.

3. Il suit que $f' = h'$ au sens des distributions. Par propriété sur cet espace, il suit qu'il existe $c \in \mathbb{R}$ telle que $f = h + c$ au sens des distributions. Or, f, h sont continues, donc cette égalité est vraie au sens des fonctions continues, d'où finalement

$$\forall x, y \in \mathbb{R}, f(x) - f(y) = \int_y^x g(t)\ \mathrm{d}t$$

□

15 Théorème de Markov-Kakutani

Référence : Analyse mathématique, une maîtrise de l'implicite, Frédéric Testard.

Théorème 15.1 : Théorème de Markov-Kakutani

Soit E un espace euclidien de dimension finie. Soit G un sous-groupe compact de $\mathrm{GL}(E)$. Considérons enfin K un compact de E stable par G :

$$\forall x \in K, \forall g \in G, g(x) \in K$$

Alors G admet un point fixe par K :

$$\exists x \in K, \forall g \in G, g(x) = x$$

Lemme 15.1 : Compact convexe stable par G

Soit C un compact convexe non vide stable par G qui contient au moins deux éléments. Alors il existe $\mathcal{A} \subset C$ tel que $\mathcal{A}$ soit stable par G, et tel que le diamètre $\delta(A)$ de $\mathcal{A}$ vérifie :

$$\delta(\mathcal{A}) < \delta(C)$$

De plus, il existe $a \in \mathcal{A}$ tel que

$$\delta(\mathcal{A}) = \sup_{g,g' \in G}\|g(a) - g'(a)\|$$

Démonstration du lemme : Par caractère compact de G, le théorème d'Ascoli nous donne l'équicontinuité de G : pour tout $\varepsilon > 0$, il existe une constante $\alpha_\varepsilon > 0$ telle que

$$\forall x, y \in C, \forall g \in G, \|x - y\| \leqslant \alpha_\varepsilon \Longrightarrow \|g(x) - g(y)\| \leqslant \varepsilon$$

Il existe alors $x_1, \cdots, x_N \in C$ tels que

$$C \subset \bigcup_{k=1}^{N} B(x_k, \alpha_\varepsilon)$$

Le plan de démonstration est le suivant :

1. Pour tout $x \in C$, et $g \in G$, il existe un indice $i \in [\![1, N]\!]$ tel que

$$\|x - g(x_i)\| \leqslant \varepsilon;$$

2. Si m désigne l'isobarycentre des $(x_i)_i$ alors

$$\forall x \in C, \forall g \in G, \|x - g(m)\| \leqslant \frac{\varepsilon}{N} + \frac{N-1}{N}\delta(C);$$

3. On conclut en définissant $\mathcal{A} = \{g(m), g \in G\}$, pour $\varepsilon = \frac{\delta(C)}{2}$.

Définissons alors $\varepsilon = \frac{\delta(C)}{2} > 0$.

1. Observons que si $x \in C$ et $g \in G$, alors $g^{-1}(x) \in C$. Il existe alors un indice i, dépendant de g et de x tel que

$$\left\|x_i - g^{-1}(x)\right\| \leqslant \alpha_\varepsilon$$

Par équicontinuité, appliqué à g, il suit que

$$\|x - g(x_i)\| \leqslant \varepsilon$$

2. Soit m l'isobarycentre des x_i. Soit $x \in C$ et $g \in G$. Il existe un indice i_0 tel que

$$\|x - g(x_{i_0})\| \leqslant \varepsilon$$

Alors

$$\|x - g(m)\| = \left\|x - \frac{1}{N}\sum_{i=1}^{N} g(x_i)\right\|$$

Ce qui se réécrit par inégalité triangulaire :

$$\|x - g(m)\| \leqslant \frac{1}{N}\sum_{i=1}^{N} \|x - g(x_i)\|$$

On peut alors conclure sur le théorème.

Démonstration : Par compacité de G, le théorème de d'Ascoli nous fournit l'équicontinuité de G : si $\varepsilon > 0$, il existe $\alpha_\varepsilon > 0$ tel que

$$\forall x, y \in K, \forall g \in G, \|x - y\| \leqslant \alpha_\varepsilon \Longrightarrow \|g(x) - g(y)\| \leqslant \varepsilon$$

Le plan de démonstration est le suivant.

1. Soit D l'application définie sur K comme étant le diamètre de l'orbite sous G de $x \in K$:

$$\forall x \in K, D(x) \overset{\text{déf.}}{=} \sup_{g,g' \in G} \|g(x) - g'(x)\|$$

Alors il existe $c_0 \in K$ tel que $D(c_0)$ est le minimum de D sur K.

2. Si le compact K est réduit à un singleton, alors l'élément de K est effectivement point fixe pour G ;

3. Sinon, supposons par l'absurde que c_0 n'est pas point fixe pour G. On définit C la fermeture de l'enveloppe convexe de l'orbite de c_0 sous G. Alors C est un compact convexe non vide qui possède au moins deux éléments et qui vérifie $\delta(C) = D(c_0)$;

4. Le lemme nous permet d'aboutir à une contradiction.

1. Pour montrer que D admet un minimum, on montre que D est continue sur le compact K. Soient $\varepsilon > 0$ et $x, y \in K$ tels que $\|x - y\| \leqslant \alpha_\varepsilon$. Alors, pour tout $g, g' \in G$:

En isolant le terme $i = i_0$, et par définition du diamètre, il suit que

$$\|x - g(m)\| \leqslant \frac{\varepsilon}{N} + \frac{N-1}{N}\delta(C)$$

3. Ainsi, pour tout $g, g' \in G$, il suit que

$$\|g'(m) - g(m)\| \leqslant \frac{\varepsilon}{N} + \frac{N-1}{N}\delta(C)$$

Cette majoration étant indépendante de g, il suit que

$$D(m) \leqslant \frac{\varepsilon}{N} + \frac{N-1}{N}\delta(C) \leqslant \frac{N-\frac{1}{2}}{N}\delta(C)$$

où on a noté

$$\forall x \in C, D(x) \overset{\text{déf.}}{=} \sup_{g,g' \in G} \|g(x) - g'(x)\|$$

D'où $D(m) < \delta(C)$. De plus, si on définit $\mathcal{A} = \{g(m), g \in G\}$ alors par définition, $D(m) = \delta(A)$. On a bien montré que $D(m) = \delta(A) < \delta(C)$. □

$$\begin{aligned}\|g(x) - g'(x)\| \leqslant & \quad \|g(x) - g(y)\| \\ & + \|g(y) - g'(y)\| \\ & + \|g'(y) - g'(x)\|\end{aligned}$$

D'où

$$\|g(x) - g'(x)\| \leqslant 2\varepsilon + D(y)$$

D'où

$$D(x) - D(y) \leqslant 2\varepsilon$$

Par symétrie de la démonstration, il suit que pour tout $\|x - y\| \leqslant \alpha_\varepsilon$:

$$|D(x) - D(y)| \leqslant 2\varepsilon$$

Donc D est (uniformément) continue sur K donc D admet un minimum en $c_0 \in K$:

$$D(c_0) = \min_{x \in K} D(x)$$

2. Si $K = \{c_0\}$, alors la stabilité de K par G implique directement que c_0 est point fixe pour G.

3. Supposons que K contient plus de deux éléments, et que c_0 n'est pas point fixe pour G. On définit

$$F \overset{\text{déf.}}{=} \overline{\{g(c_0), g \in G\}}$$

et $C \overset{\text{déf.}}{=} \text{Conv}(F)$. Le théorème d'Ascoli appliqué à G permet de donner le caractère bornée de F (relative compacité ponctuelle). F possède au moins deux éléments.

• Montrons que C est compact. Pour cela, on considère

$$\mathcal{E} \overset{\text{déf.}}{=} \left\{ (t_1, \cdots, t_{n+1}) \in [0,1]^{n+1}, \sum_{k=1}^{n+1} t_k = 1 \right\}$$

et l'application

$$f : \mathcal{E} \times F^{n+1} \longrightarrow E$$

définie par

$$f\left((t_i)_i, (A_i)_i\right) = \sum_{i=1}^{n+1} t_k A_k$$

Alors f est continue et par théorème de Carathéodory,

$$C = f\left(\mathcal{E} \times F^{n+1}\right)$$

donc C est compact.

• Déterminons le diamètre de C. Constatons d'abord que la diamètre ne voit pas l'adhérence, donc plus précisément :

$$\delta(F) = D(c_0)$$

Montrons alors que $\delta(F) = \delta(C)$. L'inclusion $F \subset C$ nous fournit l'inégalité directe. Montrons l'autre inégalité. Pour cela, on se donne $M \in C$. Par Carathéodory, il existe $(t_1, \cdots, t_{n+1}) \in \mathcal{E}$ et $A_1, \cdots, A_{n+1} \in F$ tels que

$$M = \sum_{k=1}^{n+1} t_k A_k$$

Ainsi, si $N \in F$ alors

$$\|M - N\| \leqslant \sum_{k=1}^{n+1} t_k \|A_k - N\| \leqslant \delta(F)$$

Et si $L \in C$, alors $L = \sum_{k=1}^{n+1} s_k B_j$ et

$$\|M - L\| \leqslant \sum_{k=1}^{n+1} t_k \|M - B_k\| \leqslant \delta(F)$$

d'où $\delta(C) \leqslant \delta(F)$. D'où $\delta(C) = D(c_0)$.

4. Par conséquent, C est un compact convexe qui contient au moins deux éléments, et qui est stable par G. Par le lemme, il existe $a \in C$ tel que $D(a) < \delta(C) = D(c_0)$, ce qui contredit le caractère minimal de $D(c_0)$. □

16 Irréductibilité des polynômes cyclotomiques

Référence : Daniel Perrin, *Cours d'algèbre*

Théorème 16.1

Soit

$$\Phi_n \overset{\text{déf.}}{=} \prod_{\zeta \in \mu_n^*(\mathbb{C})} (X - \zeta)$$

Alors $\Phi_n \in \mathbb{Z}[X]$ est irréductible dans $\mathbb{Q}$. En conséquence, le polynôme minimal de $\zeta \in \mu_n^*(\mathbb{C})$ est exactement Φ_n, et l'extension cyclotomique $\mathbb{Q}(\zeta)/\mathbb{Q}$ vérifie :

$$[\mathbb{Q}(\zeta) : \mathbb{Q}] = \varphi(n)$$

Démonstration : Le plan de démonstration est le suivant :

1. Pour p premier avec n, si $\zeta \in \mu_n^*(\mathbb{C})$, c'est aussi le cas de ζ^p. On considère f et g les polynômes minimaux de ζ et ζ^p pour $\mathbb{Q}$. Alors $f, g \in \mathbb{Z}[X]$ et f et g divisent Φ_n.
2. Par l'absurde, on a $f = g$, en réduisant les deux polynômes sur $\mathbb{F}_p[X]$ et en montrant que dans ce cas, Φ_n admet une racine double dans un corps de décomposition.
3. Il suit que $f = \Phi_n$, car f contient toutes les racines de Φ_n, donc Φ_n est irréductible dans $\mathbb{Q}[X]$.

1. On a le fait suivant : $\mathbb{Z}[X]$ est factoriel. On suppose aussi acquis le fait que $\Phi_n \in \mathbb{Z}[X]$. On peut alors décomposer Φ_n en produit de facteurs irréductibles :

$$\Phi_n = \prod_{i=1}^{m} f_i^{r_i}$$

où les $f_i \in \mathbb{Z}[X]$ sont irréductibles. Or,

$$\Phi_n(\zeta) = \Phi_n\left(\zeta^p\right) = 0$$

Il existe donc deux indices i, j tels que $f_i(\zeta) = 0$ et $f_j(\zeta^p) = 0$. Ainsi, f_i divise f, qui est lui-même irréductible et unitaire. D'où $f = f_i$, et de même $g = f_j$. Donc f et g divisent Φ_n, et sont des polynômes de $\mathbb{Z}[X]$.

2. Supposons que $f \neq g$. Alors par irréductbilité de f et g, il suit que fg divise Φ_n. De plus,

$$g(\zeta^p) = 0$$

donc $g(X^p)$ est annulateur pour ζ, donc f divise $g(X^p)$, dans $\mathbb{Q}[X]$, par définition d'être minimal dans $\mathbb{Q}[X]$. Il existe $h \in \mathbb{Q}[X]$ tel que

$$g(X^p) = f(X)h(X)$$

De plus, il existe $b \in \mathbb{Z}$ tel que $bh \overset{\text{déf.}}{=} \tilde{h} \in \mathbb{Z}[X]$. Ainsi,

$$bg(X^p) = f(X)\tilde{h}(X)$$

Puisque $b \neq 0$ (sans quoi $h = 0$), le lemme de Gauss conclut que f divise $g(X^p)$ dans $\mathbb{Z}[X]$. On considère alors que $h \in \mathbb{Z}[X]$. Projetons cette égalité dans $\mathbb{F}_p$, on note $\bar{P} \in \mathbb{F}_p[X]$ la projection de $P \in \mathbb{Z}[X]$. Par morphisme de Frobenius,

$$\bar{g}(X^p) = \bar{g}(X)^p$$

Ainsi,

$$\bar{g}(X)^p = \bar{f}(X)\bar{h}(X)$$

Si φ est un diviseur irréductible de $\bar{f}$ alors φ divise aussi $\bar{g}(X)^p$ donc $\bar{g}$. Ainsi, φ^2 divise $\bar{f}\bar{g}$, donc divise $\bar{\Phi}_n$. Par conséquent, $\bar{\Phi}_n$ admet une racine double dans une extension de $\mathbb{F}_p$. Or, si $\zeta \in \mathbb{F}_p$ est racine de $\bar{\Phi}_n$ alors ζ est racine de $X^n - 1$, qui est premier avec son polynôme dérivé donc n'admet pas de facteur carré, sachant que p ne divise pas n, donc ζ ne peut être racine double de $\bar{\Phi}_n$. D'où la contradiction recherchée, d'où $f = g$.

3. On note $\zeta \overset{\text{déf.}}{=} e^{\frac{2i\pi}{n}}$. Montrons que f, le polynôme minimal de ζ, contient toutes les racines de Φ_n. Pour tout $\zeta' \in \mu_n^*(\mathbb{C})$, il existe $m \in \mathbb{Z}$ premier avec n tel que $\zeta' = \zeta^m$. Décomposons m en produit de facteurs premiers :

$$m = \prod_{i=1}^{r} p_i^{\alpha_i}$$

où chaque p_i ne divise pas n, donc $p_i^{\alpha_i}$ ne divise pas non plus n. Par le point **(2)**, il suit que

$$f\left(\zeta^{p_1^{\alpha_1}}\right) = 0$$

Et de même, puisque le polynôme minimal de $\zeta^{p_1^{\alpha}}$ est le même que celui de $\left(\zeta^{p_1^{\alpha}}\right)^{p_2^{\alpha_2}} = \zeta^{p_1^{\alpha_1} p_2^{\alpha_2}}$, il suit par récurrence que

$$f\left(\zeta^m\right) = f(\zeta') = 0$$

Ainsi, f contient toutes les racines de Φ_n, donc Φ_n divise f, et lui-même divise Φ_n par **1.**. Ces deux polynômes étant unitaires, il suit que $f = \Phi_n$, d'où le caractère irréductible de Φ_n sur $\mathbb{Q}[X]$. □

17 Nombre moyen de cycles

Références :

— *Exercices de mathématiques MP-MP**, Thierry Dugardin et Marc Rezzouk ;

— *Mathématiques tout-en-un MP-MP**, Claude Deschamps et co.

Proposition 17.1 : Nombre moyen de cycles

Le nombre moyen de cycles dans la décomposition d'une permutation de $\mathfrak{S}_n$ prise aléatoirement (pour une loi uniforme) est $\sum_{k=1}^{n} \frac{1}{k}$. La variance associée est équivalente à $\ln(n)$ lorsque $[n \to +\infty]$.

Démonstration : La définition suivante permet d'éclairer cette proposition : si $\sigma \in \mathfrak{S}_n$, on dit que $\sigma(i)$, pour $i \in [\![1, n]\!]$ est un maximum provisoire si

$$\sigma(i) = \max\{\sigma(1), \cdots, \sigma(i)\}$$

Dans $(\mathfrak{S}_n, \mathcal{P}(\mathfrak{S}_n), \mathbb{P})$, où $\mathbb{P}$ est la probabilité uniforme, on définit $X_n(\sigma)$ comme étant le nombre de maximums relatifs de σ. On cherche la loi de X_n.

1. On introduit pour cela la variable aléatoire Z_k : $\mathfrak{S}_n \longrightarrow \{0, 1\}$, indicatrice de l'événement $\sigma(k)$ est un maximum provisoire. Alors les variables aléatoires $(Z_1, \cdots, Z_n)$ sont indépendantes ;
2. Par construction, X_n est la somme des Z_k : il est alors aisé de connaître sa fonction génératrice ;

3. On peut déterminer certaines probabilités $\mathbb{P}(X_n = 1), \mathbb{P}(X_n = 2)$ ou encore $\mathbb{P}(X_n = n)$, par curiosité. On peut surtout déterminer l'espérance et la variance de X_n, grâce à une dérivée logarithmique ;
4. On se ramène à notre problème en introduisant la bijection suivante : si $\sigma = \gamma_1 \circ \cdots \circ \gamma_r$ est décomposé en cycle (on écrit les 1-cycles), on réécrit un cycle $\gamma = (i_1, \cdots, i_s)$ en plaçant l'élément maximal présent dans le cycle en tête, et on ordonne les cycles ainsi écrits *par ordre croissant de l'élément maximal*. L'application Φ en question est alors la permutation concaténée. Alors Φ est bien une bijection, et $\Phi^{-1}(\{X_n = k\})$ est l'ensemble des permutations possédant k cycles, en comptant les 1-cycles.

1. Pour montrer que les $(Z_k)_k$ sont indépendants, puisqu'il s'agit de variables de Bernouilli, il nous suffit de montrer que les événements $(\{Z_k = 1\})_k$ sont indépendants. Soient $1 \leqslant \phi(1) < \cdots < \phi(k) \leqslant n$ des entiers. On procède par dénombrement. L'objectif est de déterminer le cardinal de

$$\{\sigma \in \mathfrak{S}_n, Z_{\phi(1)}(\sigma) = \cdots = Z_{\phi(k)} = 1\}$$

On cherche alors à caractériser les permutations σ telles que ses maximums relatifs soient exactement les $\sigma(\phi(j))$. C'est là qu'on cherche à faire du dénombrement : sortez vos mains.

• Par définition, $\phi(k)$ réalise le maximum de σ, et est strictement supérieur à au moins $\phi(k) - 1$ éléments. Pour définir une permutation σ qui possède un maximum provisoire en $\phi(k)$, on a alors $n - \phi(k)$ possibilités de placer $\phi(k)$ (les premières places sont prises par les éléments plus petits). Il s'agit alors de déterminer le nombre de parties à $n - \phi(k)$ éléments dans un ensemble à n éléments : c'est $\binom{n}{n-\phi(k)} = \binom{n}{\phi(k)}$.

• Supposons $\phi(k)$ fixé. On cherche alors à définir une permutation σ dont son avant-dernier maximum provisoire est en $\phi(k-1)$. On ne peut le placer qu'après au moins $\phi(k-1)$ positions, pour laisser la place à ses copains. Mais il doit aussi être placé avant $\phi(k)$. On cherche alors le nombre de parties à $\phi(k) - \phi(k-1) - 1$ éléments dans un ensemble à $\phi(k) - 1$ éléments. Il y a alors $\binom{\phi(k)-1}{\phi(k-1)}$ choix de placement. Quand à la valeur que peut prendre $\sigma(\phi(k-1))$, elle ne peut pas valoir $\sigma(\phi(k)) = n$ et doit être supérieure ou égale à $\sigma(\phi(k-1))$, d'où $(\phi(k) - \phi(k-1) - 1)!$ choix de valeurs (σ est bijective). D'où finalement

$$\binom{\phi(k)-1}{\phi(k-1)} \cdot (\phi(k) - \phi(k-1) - 1)! = \frac{(\phi(k)-1)!}{\phi(k-1)!}$$

choix possibles de permutations en se fixant $\phi(k)$.

• Supposons par récurrence avoir construits les permutations en se fixant tous les indices $\phi(2), \cdots, \phi(k)$. Il ne nous reste plus qu'à se fixer l'indice $\phi(1)$ pour déterminer la liste des permutations dont on connaît l'emplacement des maximums provisoires. De la même manière, on aura en fait $\frac{(\phi(2)-1)!}{\phi(1)!}$ choix possibles. Pour connaître entièrement les permutations, il ne nous reste plus qu'à déterminer les coefficients en dehors de $[\![\phi(1), \phi(k)]\!]$: cela nous fait $(\phi(1)-1)!(n-\phi(k))!$ choix restants.

• Tout ce dénombrement permet alors de conclure que

$$\mathbb{P}\left(\bigcap_{j=1}^k Z_{\phi(j)} = 1\right) = \frac{(\phi(1)-1)!(n-\phi(k))!}{n!} \cdot \binom{n}{\phi(k)} \prod_{j=2}^k \frac{(\phi(j)-1)!}{\phi(j-1)!}$$

Ce qui se simplifie sympathiquement en :

$$\mathbb{P}\left(\bigcap_{j=1}^k Z_{\phi(j)} = 1\right) = \prod_{j=1}^k \frac{1}{\phi(j)}$$

Or, le cas $k = 1$ donne alors que $\mathbb{P}(Z_i = 1) = \frac{1}{i}$, d'où

$$\mathbb{P}\left(\bigcap_{j=1}^k Z_{\phi(j)} = 1\right) = \prod_{j=1}^k \mathbb{P}(Z_{\phi(j)} = 1)$$

Les variables $(Z_k)_k$ sont donc bien indépendantes.

2. Or, $X_k(\sigma)$, la variables qui compte le nombre de maximums provisoires, n'est rien d'autre que la somme des Z_k. Par indépendance, il suit que

$$\forall t \in [-1, 1], G_{X_n}(t) = \prod_{k=1}^n G_{Z_k}(t)$$

Avec

$$G_{Z_k}(t) = \mathbb{P}(Z_k = 0) + t\mathbb{P}(Z_k = 1)$$

D'où

$$G_{Z_k}(t) = 1 + \frac{t-1}{k}$$

D'où

$$G_{X_n}(t) = \prod_{k=1}^n \left(1 + \frac{t-1}{k}\right) = \frac{1}{n!} \prod_{k=0}^{n-1} (t+k)$$

3. Il est alors possible de déterminer l'espérance de X_n, en dérivant G_{X_n}, que l'on note g_n dans la suite. Néanmoins, la méthode bourrine peut être optimisée (par définition d'être bourrin) grâce à un peu de de dérivée logarithmique. En effet, constatons que l'on a, là où c'est bien défini :

$$\frac{g_n'(t)}{g_n(t)} = \sum_{k=0}^{n-1} \frac{1}{t+k}$$

D'où, en prenant $t = 1$ (on le peut, pas de dénominateur qui s'annule) :

$$\mathbb{E}[X_n] = \sum_{k=1}^n \frac{1}{k} \underset{n \to +\infty}{\sim} \ln n$$

On fait de même avec la variance, en dérivant l'expression "logarithmique" :

$$\frac{g_n''(t)}{g_n(t)} - \frac{g_n'(t)^2}{g_n(t)^2} = \sum_{k=0}^{n-1} \frac{-1}{(t+k)^2}$$

D'où, puisque

$$\mathrm{Var}(X_n) = g_n''(1) - \mathbb{E}[X_n]^2 + \mathbb{E}[X_n]$$

On obtient :

$$\mathrm{Var}(X_n) = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k^2} \underset{n\to+\infty}{\sim} \ln n$$

4. Concluons. Constatons d'abord que Φ est bien définie par unicité de la décomposition en cycle, à l'ordre près. Montrons que Φ admet un inverse Ψ. Pour cela, on le définit explicitement de la sorte : si $\rho \in \mathfrak{S}_n$, et qu'on dispose de la liste $\{\rho(1), \rho(i_2), \cdots, \rho(i_k)\}$ de ses maximums provisoires, on définit

$$\Psi(\rho) \overset{\text{déf.}}{=} \begin{matrix} (\rho(1) & \cdots & \rho(i_2)-1) \\ (\rho(i_2) & \cdots & \rho(i_3)-1) \\ \vdots & & \\ (\rho(i_k) & \cdots & \rho(n)) \end{matrix}$$

Par construction de Φ, on a alors $\Psi \circ \Phi = \mathrm{id}_{\mathfrak{S}_n}$. Puisqu'il s'agit d'applications entre espaces de cardinaux finis et égaux, il suit que $\Psi = \Phi^{-1}$. De plus, toujours par construction de Φ :

$$\Phi^{-1}(\{X_n = k\}) = \Psi(\{X_n = k\})$$

Et donc, par construction de Ψ, si σ possède k maximums provisoires, cela signifie que $\Psi(\sigma)$ possède k cycles, en comptant ceux de longueur 1. Puisque Ψ est une bijection, il suit alors que $\mathbb{E}[\Psi(X_n)] = \mathbb{E}[X_n]$. Il y a donc bien en moyenne $\sum_{k=1}^{n} \frac{1}{k}$ cycles dans la décomposition d'une permutation de $\mathfrak{S}_n$. □

Petits exemples : Pour $n = 3$, on a

σ	$(1\ 2\ 3)$	$(1\ 3\ 2)$	$(1\ 2)$	$(1\ 3)$	$(2\ 3)$	id
$X_3(\sigma)$	2	1	2	1	2	3

D'où $\mathbb{E}[X_3] = \frac{11}{6}$ et $\mathrm{Var}(X_3) = \frac{17}{36}$.
Explicitons les applications Φ et Ψ. Pour

$$\sigma = (8\ \ 7\ \ 6\ \ 4\ \ 2\ \ 1\ \ 5\ \ 3) \in \mathfrak{S}_8$$

On décompose σ en cycles, en faisant apparaître les 1-cycles :

$$\sigma = (1\ \ 8\ \ 3\ \ 6)(2\ \ 7\ \ 5)(4)$$

Puis, on écrit chaque cycle avec comme premier coefficient le plus grand possible :

$$\sigma = (8\ \ 3\ \ 6\ \ 1)(7\ \ 5\ \ 2)(4)$$

On permute les cycles de sorte que chaque premier coefficient se présente comme une suite croissante.

$$\sigma = (4)(7\ \ 5\ \ 2)(8\ \ 3\ \ 6\ \ 1)$$

On définit alors

$$\Phi(\sigma) = (4\ \ 7\ \ 5\ \ 2\ \ 8\ \ 3\ \ 6\ \ 1)$$

Explicitons Ψ. Si

$$\rho = (4\ \ 3\ \ 5\ \ 1\ \ 2\ \ 7\ \ 6) \in \mathfrak{S}_7$$

Les maximums provisoires de ρ sont, dans l'ordre $\{4, 5, 7\}$. On définit alors

$$\Psi(\rho) = (4\ \ 3)(5\ \ 1\ \ 2)(7\ \ 6)$$

Il est alors plus clair de voir que $\Psi \circ \Phi = \mathrm{id}_{\mathfrak{S}_n}$.

18 Simplicité du groupe spécial orthogonal dans l'espace

Référence : Serge Francinou, Hervé Gianella et Serge Nicolas, *Oraux X/ENS, Algèbre 3*

Théorème 18.1 : Simplicité de $\mathrm{SO}_3(\mathbb{R})$

Soit $\mathrm{SO}_3(\mathbb{R})$ l'ensemble des matrices orthogonales de déterminant 1. Alors $\mathrm{SO}_3(\mathbb{R})$ est un groupe simple : ses seuls sous-groupes distingués sont triviaux.

La voie principale pour démontrer algébriquement ce théorème est d'utiliser la notion de connexité par arcs. On supposera connu le fait que $\mathrm{SO}_3(\mathbb{R})$ est engendré par les retournements. On supposera aussi connu le fait que l'action de $\mathrm{SO}_3(\mathbb{R})$ sur les droites est transitive.

Lemme 18.1 : Condition suffisante d'égalité à $\mathrm{SO}_3(\mathbb{R})$

Soit $G_0 \lhd \mathrm{SO}_3(\mathbb{R})$ connexe par arcs et non réduit à $\{I_3\}$. Alors $G_0 = \mathrm{SO}_3(\mathbb{R})$.

Démonstration du lemme : La plan est le suivant :

- Montrons que G_0 contient alors une rotation d'angle $\frac{\pi}{2}$, à l'aide d'un chemin reliant I_3 à $g \in G_0$, où $g \neq I_3$ et du théorème des valeurs intermédiaires.
- Puisque G_0 est distingué dans $\mathrm{SO}_3(\mathbb{R})$ et que l'action de $\mathrm{SO}_3(\mathbb{R})$ sur les droites est transitive, G_0 possède alors tous les retournements, et donc tous les éléments de $\mathrm{SO}_3(\mathbb{R})$, puisque qu'ils engendrent ce dernier.

1. Soit $g \in G_0$ différent de I_3. Quitte à considérer g^{-1}, on peut supposer que l'angle θ_0 associé à g est dans $[0, \pi[$.

• Supposons dans un premier temps que $\theta_0 \in \left[\frac{\pi}{2}, \pi\right[$. Alors en particulier, $\cos\theta_0 \leqslant 0$. De plus, l'application

$$\varphi : \begin{pmatrix} \mathrm{SO}_3(\mathbb{R}) & \longrightarrow & \mathbb{R} \\ R & \longmapsto & \frac{\mathrm{Tr}(R)-1}{2} \end{pmatrix}$$

est continue, et vérifie en particulier $\varphi(g) = \cos\theta_0 \leqslant 0$. Enfin, G_0 est connexe par arcs, donc il existe un chemin γ reliant I_3 à g. Alors $\varphi\circ\gamma$ est une application définie sur $[0,1]$ à valeurs réelles qui est continue, et qui vérifie $\varphi\circ\gamma(0) = 1$ et $\varphi\circ\gamma(1) \leqslant 0$. Par le théorème des valeurs intermédiaires, il existe alors $s \in [0,1]$ tel que

$$\varphi\circ\gamma(s) = 0$$

On note $\gamma(s) = r \in G$. Alors r est associé à un angle θ_1 qui vérifie $\cos\theta_1 = 0$. Donc, $\theta_1 \equiv \frac{\pi}{2}\ [\pi]$. Il suit que r^2 est un retournement qui appartient à G.

• Si $\theta_0 \in \left]0, \frac{\pi}{2}\right[$, alors il existe un rang $N \in \mathbb{N}$ tel que $g^N \in G_0$ soit d'angle $N\theta_0 \in \left]\frac{\pi}{2}, \pi\right[$. Ainsi, G_0 contient là aussi un retournement.

2. Notons R le retournement en question. Puisque G_0 est distingué dans $\mathrm{SO}_3(\mathbb{R})$,

$$\forall g \in \mathrm{SO}_3(\mathbb{R}), gRg^{-1} \in G_0$$

est lui-même un retournement, d'axe $g(\Delta)$, où Δ est l'axe de R. Puisque l'action de $\mathrm{SO}_3(\mathbb{R})$ sur les droites est transitive (elle n'admet qu'une seule orbite : pour tout couple de droites, on dispose d'une rotation qui transforme la première en la deuxième), il suit alors que G_0 contient tous les retournements, qui engendrent $\mathrm{SO}_3(\mathbb{R})$. Il suit que $G_0 = \mathrm{SO}_3(\mathbb{R})$. □

Démontrons alors la simplicité de $\mathrm{SO}_3(\mathbb{R})$.

Démonstration : Soit G un sous-groupe distingué de $\mathrm{SO}_3(\mathbb{R})$, non réduit à $\{I_3\}$. L'idée est d'appliquer le lemme non pas directement à G, dont on ne sait pas s'il est connexe par arcs, mais à G_0, la composante connexe par arcs de I_3 dans G :

$$G_0 \overset{\text{déf.}}{=} \left\{ g \in G, \exists \gamma \in \mathcal{C}^0([0,1], G), \begin{array}{l} \gamma(0) = I_3 \\ \gamma(1) = g \end{array} \right\}$$

Le plan est le suivant :

- G_0 est un sous-groupe de G ;
- Ainsi, G_0 est bien distingué dans $\mathrm{SO}_3(\mathbb{R})$;
- Pour conclure avec le lemme, il reste à montrer que G_0 contient un autre élément que I_3. Par l'absurde, si $G_0 = \{I_3\}$, alors toutes les composantes connexes de $\mathrm{SO}_3(\mathbb{R})$ sont des singletons, ce qui implique que $G = \{I_3\}$, ce qui contredit notre hypothèse de départ sur G.

1. Soient $A, B \in G_0$. Il existe deux chemins γ_A et γ_B reliant I_3 à A et B. Alors le chemin

$$t \longmapsto \gamma_A(t)\gamma_B(t)^{-1}$$

est à valeur dans G, est continue et relie I_n à AB^{-1}. Ainsi, $AB^{-1} \in G_0$, donc G_0 est bien un sous-groupe de G.

2. Si $g \in \mathrm{SO}_3(\mathbb{R})$, et $A \in G_0$, de chemin γ_A reliant I_3 à A alors

$$t \longmapsto g\gamma_A(t)g^{-1}$$

relie I_3 à gAg^{-1}, et est à valeurs dans G car $A \in G_0 \subset G$, qui est distingué dans $\mathrm{SO}_3(\mathbb{R})$. Donc $G_0 \lhd \mathrm{SO}_3(\mathbb{R})$.

3. Supposons que $G_0 = \{I_3\}$.

• Soient $A, B \in G$ dans la même composante connexe par arcs dans G. Notons γ le chemin joignant A à B. Alors

$$t \longmapsto \gamma(t)B^{-1}$$

relie AB^{-1} à I_3, et est à valeurs dans G. D'où $AB^{-1} \in G_0 = \{I_3\}$, d'où $A = B$. Les composantes connexes par arcs de G seraient alors des singletons.

• Or, si $R = R(\theta)$ désigne la rotation d'angle θ, et si

$$\gamma_R : t \longmapsto \begin{pmatrix} \cos\theta t & -\sin\theta t & 0 \\ \sin\theta t & \cos\theta t & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

alors γ_R relie R à I_3, dans $\mathrm{SO}_3(\mathbb{R})$. Soit $A \in G$. Alors le chemin

$$t \longmapsto \gamma_R(t)A\gamma_R(t)^{-1}$$

relie RAR^{-1} à A, et est à valeurs dans G, car $G \lhd \mathrm{SO}_3(\mathbb{R})$. En conséquent, RAR^{-1} et A sont dans la même composante connexe par arcs dans G. Par le précédent point, il suit que pour tout $R \in \mathrm{SO}_3(\mathbb{R})$:

$$A = R^{-1}AR$$

Cela impliquerait que la droite Δ associée à A est stable par toutes les rotations de l'espace. Cela n'est possible que si $A = I_3$. D'où $G = \{I_3\}$, et la contradiction recherchée. □

19 Réduction de Jordan pour les nilpotents

Référence : Algèbre linéaire : Réduction des endomorphismes, Roger Mensuy et Rached Mneimné
On se contente d'une partie existence ici. Le point de départ est le suivant :

Proposition 19.1 : Noyaux itérés

Soient $P \in \mathbb{K}[X]$ et $f \in \mathcal{L}(E)$, où E est un $\mathbb{K}$-espace vectoriel de dimension finie. Alors la suite des noyaux itérés $(\ker P^k f)_k$ est d'abord strictement croissante puis constante à partir d'un rang p. De plus, si P est irréductible, $k \in [\![1, p]\!]$ et $x \in \ker P^k f \setminus \ker P^{k-1} f$ alors le polynôme minimal local de f en x est exactement P^k :

$$\mu_{f,x} = P^k$$

On rappelle qu'on a défini le polynôme minimal local comme étant le générateur unitaire de l'idéal $\{P \in \mathbb{K}[X], [P(f)](x) = 0\}$ de $\mathbb{K}[X]$.

Lemme 19.1 : Polynôme minimal et polynôme minimal local

Pour tout endomorphisme $f \in \mathcal{L}(E)$, où E est de dimension finie, il existe $x \in E$ tel que le polynôme minimal local en x et le polynôme minimal sont égaux :

$$\exists x \in E, \mu_{f,x} = \mu_f$$

Démonstration du lemme : On décompose le polynôme minimal μ_f en produit de facteur irréductibles :

$$\mu_f = \prod_{k=1}^{p} P_k^{\alpha_k}$$

Et considérons pour tout k un élément $x_k \in E$ tel que

$$x_k \in \ker P^{\alpha_k} f \setminus \ker P^{\alpha_k - 1} f$$

Cet élément est bien défini par croissance des noyaux itérés. De plus, x_k a pour polynôme minimal $P_k^{\alpha_k}$. Pour conclure ce lemme, il nous suffit de considérer

$$x \overset{\text{déf.}}{=} \sum_{k=1}^{p} x_k \in \bigoplus_{k=1}^{p} \ker P_k^{\alpha_k} f$$

où la somme est bien directe par le lemme des noyaux. Cet élément x vérifie d'une part $\mu_{f,x} f(x) = 0$ et donne alors la décomposition

$$0 = \sum_{k=1}^{p} \left[\mu_{f,x} f\right](x_k)$$

Or, puisque $x_k \in \ker(P_k^{\alpha_k} f)$, tout polynôme en x_k aussi, donc en particulier :

$$\forall k \in [\![1,p]\!], [\mu_{f,x} f](x_k) \in \ker P_k^{\alpha_k} f$$

Ces espaces étant en somme directe, l'unicité de la décomposition donne alors que

$$\forall k \in [\![1,p]\!], [\mu_{f,x} f](x_k) \in \ker P_k^{\alpha_k} f = 0$$

Ainsi, le polynôme minimal de x_k divise bien $\mu_{f,x}$, *id est* P^k divise $\mu_{f,x}$. Il suit que μ_f, produit des P^k, divise bien $\mu_{f,x}$, lui-même divisant μ_f. D'où l'égalité souhaitée. □

Lemme 19.2 : Supplémentaire stable d'un sous-espace cyclique

Soient $f \in \mathcal{L}(E)$ et $x \in E$ tel que $\mu_f = \mu_{f,x}$. Alors le sous-espace

$$E_{f,x} \overset{\text{déf.}}{=} \text{Vect}_{\mathbb{K}}\left(f^j(x), j \in \mathbb{N}\right) = \text{Vect}_{\mathbb{K}}\left(x, f(x), \cdots, f^{p-1}(x)\right)$$

où p est le degré de $\mu_{f,x}$ admet un supplémentaire dans E stable par f.

Démonstration du lemme : On complète la base $(e_1 = x, e_2 = f(x), \ldots, e_p = f^{p-1}(x))$ en une base de E, et on définit F de la sorte :

$$F \overset{\text{déf.}}{=} \left\{y \in E, \forall j \in \mathbb{N}, e_p^*\left(f^j(y)\right) = 0\right\}$$

Il est alors immédiat que F ainsi construit est stable par f. Montrons que F est le supplémentaire tant attendu.

- Soit $y \in E_{f,x} \cap F$. Alors y se décompose de la sorte :

$$y = \sum_{j=0}^{+\infty} a_j f^j(x)$$

où $a_j = 0$ pour $j \geqslant p$. De plus, y vérifie

$$\forall j \in \mathbb{N}, e_p^*(f^j(x)) = 0 = a_j$$

d'où $y = 0$, et don $E_{f,x} \cap F = \{0\}$.

- Déterminons la dimension de F. Puisque $\mu_{f,x} = \mu_f$, on a

$$\mathbb{K}[f] = \text{Vect}_{\mathbb{K}}(\text{id}, f, \cdots, f^{p-1})$$

Ainsi, F s'exprime en terme d'intersection finie d'hyperplans :

$$F = \bigcap_{j=0}^{p-1} \ker\left(e_p^* \circ f^j\right)$$

Montrons que la famille $\left(e_p^* \circ f^j\right)_{0 \leqslant j \leqslant p-1}$ est libre. On se donne $\alpha_0, \cdots, \alpha_{p-1} \in \mathbb{K}$ tels que

$$\sum_{j=0}^{p-1} \alpha_j e_p^* \circ f^j = 0$$

Alors en particulier, l'élément $y \overset{\text{déf.}}{=} \sum_{j=0}^{p-1} \alpha_j f^j(x) \in E_{f,x} \cap F$, donc est nul. Donc tous les α_j sont nuls, donc la famille $\left(e_p^* \circ f^j\right)_{0 \leqslant j \leqslant p-1}$ est bien libre, d'où

$$\dim_{\mathbb{K}} F = n - p$$

Ce qui conclut bien sur l'existence d'un supplémentaire stable par f de $E_{f,x}$. □

Théorème 19.1 : Réduction de JORDAN pour les nilpotents, existence

Soit f un endomorphisme nilpotent sur E un $\mathbb{K}$-espace vectoriel de dimension finie. Alors il existe des entiers $d_1 \geqslant \cdots \geqslant d_r$ et une base $\mathfrak{b}$ de E dans laquelle f s'écrit comme matrice diagonale par blocs de matrices de JORDAN :

$$\text{Mat}_{\mathfrak{b}}(f) = \begin{pmatrix} J_{d_1} & & \\ & \ddots & \\ & & J_{d_r} \end{pmatrix}$$

En particulier, d_1 est le degré du polynôme minimal de f.

Démonstration : On fait la démonstration par récurrence sur la dimension de E.

• Pour E de dimension 2, le résultat est vrai si f est l'endomorphisme nul. Sinon, si $x \in E$ vérifie $f(x) \neq 0$, alors $(x, f(x))$ est une base, par caractère nilpotent de f (la seule valeur propre de f est 0). Dans cette base, f est bien représentée par une matrice de Jordan.

• Supposons le résultat acquis pour tout $\mathbb{K}$-espace vectoriel de dimension inférieure ou égale à n et tout endomorphisme nilpotent sur cet espace vectoriel. Soit E un $\mathbb{K}$-espace vectoriel de dimension $n+1$ et soit $f \in \mathcal{L}(E)$ nilpotent. On va simplement appliquer l'hypothèse de récurrence sur un supplémentaire de $E_{f,x}$, où $\mu_f = \mu_{f,x}$ stable par f, que l'on nomme F. Sur F, il existe une base $\mathfrak{b}_F$ telle que

$$\mathrm{Mat}_{\mathfrak{b}_F}\left(f_{|_F}\right) = \begin{pmatrix} J_{d_2} & & \\ & \ddots & \\ & & J_{d_r} \end{pmatrix}$$

où d_2 est le degré de $\mu_{f_{|_F}}$. On dispose de plus d'une base de l'espace cyclique $E_{f,x}$ donnée par :

$$\mathfrak{b}_x \overset{\text{déf.}}{=} (x, f(x), \cdots, f^{p-1}(x))$$

où p est le degré de μ_f. La base $\mathfrak{b}$ concaténée de $\mathfrak{b}_x$ et de $\mathfrak{b}_F$ permet alors de représenter f matriciellement par une matrice diagonale par blocs composée des blocs $J_p = J_{d_1}$ où $d_1 = p$ est le degré de μ_f, et des blocs J_{d_k}. Enfin, puisque

$$\mu_{f_{|_F}} \mid \mu_f$$

il suit que $d_1 = \deg(\mu_f) \geqslant \deg\left(\mu_{f_{|_F}}\right) = d_2$. Ce qui achève la récurrence. □