Chapitre 9

Transformation de Fourier et convergence en loi

La transformation de Fourier déjà entrevue à propos des fonctions caractéristiques est un outil important, aussi bien en analyse qu'en théorie des probabilités. Après avoir introduit les propriétés fondamentales de la transformée de Fourier des mesures finies et des fonctions de $\mathrm{L}^1(\lambda_d)$, nous verrons comment la transformée de Fourier permet l'étude de la *convergence des lois de probabilité*. Cette étude débouche naturellement sur l'exemple le plus célèbre de convergence en loi. Il s'agit du *théorème limite central* qui affirme que la suite convenablement normalisée des sommes partielles de variables aléatoires i.i.d. et de carré intégrable suit asymptotiquement une loi gaussienne.

Dans tout le chapitre, $\mathbb{R}^d$ est muni du produit scalaire et de la norme euclidienne standards notés :

$$\langle t, x\rangle = \sum_{j=1}^{d} t_j x_j, \quad \|x\| = \left(x_1^2 + \cdots + x_d^2\right)^{1/2},$$

où $t = (t_1, \ldots, t_d)$ et $x = (x_1, \ldots, x_d)$ désignent deux vecteurs quelconques de $\mathbb{R}^d$.

9.1 Transformée de Fourier d'une mesure

9.1.1 Premières propriétés

Définition 9.1. *Soit μ une mesure finie sur $\left(\mathbb{R}^d, \operatorname{Bor}(\mathbb{R}^d)\right)$. On appelle* transformée de Fourier *de μ la* fonction $\widehat{\mu} : \mathbb{R}^d \to \mathbb{C}$, *définie par :*

$$\forall t \in \mathbb{R}^d, \quad \widehat{\mu}(t) := \int_{\mathbb{R}^d} \exp\left(i\langle t, x\rangle\right) \mathrm{d}\mu(x). \tag{9.1}$$

Lorsque μ est la loi d'un vecteur aléatoire, $\widehat{\mu}$ est la fonction caractéristique *de X.*

$$\text{Si } \mu = P_X, \quad \widehat{\mu}(t) = \varphi_X(t) = \mathbf{E} \exp\left(i\langle t, X\rangle\right). \tag{9.2}$$

On notera que l'hypothèse μ finie (donc $\mu(\mathbb{R}^d) < +\infty$) rend automatiquement μ intégrable sur $\mathbb{R}^d$ la fonction $x \mapsto \exp\bigl(i\langle t, x\rangle\bigr)$, de module constant 1. Il est clair que cette fonction n'est λ_d intégrable sur $\mathbb{R}^d$ pour aucun t et donc (9.1) ne permet pas de définir $\widehat{\lambda_d}$.

Les deux tableaux suivants donnent les fonctions caractéristiques de quelques lois usuelles, discrètes ou à densité. Les calculs de vérification sont laissés en exercice.

Exemple 9.1 (Fonctions caractéristiques de lois discrètes usuelles, $d = 1$).

Loi de X	Paramètres	$\mathbf{P}(X = k)$	$\varphi_X(t)$
δ_a	$a \in \mathbb{R}$	$\mathbf{P}(X = a) = 1$	e^{ita}
$\mathrm{Bern}(p)$	$p \in [0, 1]$	$\mathbf{P}(X = 0) = 1 - p,\ \mathbf{P}(X = 1) = p$	$p\mathrm{e}^{it} + 1 - p$
$\mathrm{Bin}(n, p)$	$p \in [0, 1],\ n \in \mathbb{N}^*$	$C_n^k p^k (1-p)^{n-k}, \quad 0 \le k \le n$	$\bigl(p\mathrm{e}^{it} + (1-p)\bigr)^n$
$\mathrm{Geom}(p)$	$p \in]0, 1[$	$(1-p)^{k-1}p, \quad k \in \mathbb{N}^*$	$\dfrac{p\mathrm{e}^{it}}{1 - (1-p)\mathrm{e}^{it}}$
$\mathrm{Pois}(\alpha)$	$\alpha \in]0, +\infty[$	$\dfrac{\mathrm{e}^{-\alpha}\alpha^k}{k!}, \quad k \in \mathbb{N}$	$\exp\bigl(\alpha(\mathrm{e}^{it} - 1)\bigr)$

Exemple 9.2 (Fonctions caractéristiques de lois usuelles à densité, $d = 1$).

Loi	Paramètres	Densité $f_X(x)$	$\varphi_X(t)$
$\mathrm{Unif}[a, b]$	$a, b \in \mathbb{R}$	$\dfrac{1}{b-a}\mathbf{1}_{[a,b]}(x)$	$\dfrac{\mathrm{e}^{ibt} - \mathrm{e}^{iat}}{(b-a)it}$
$\mathrm{Exp}(a)$	$a \in]0, +\infty[$	$a\mathrm{e}^{-ax}\mathbf{1}_{[0,+\infty[}(x)$	$\dfrac{1}{1 - \frac{it}{a}}$
$\mathrm{Cau}(a, b)$	$a \in \mathbb{R},\ b \in]0, +\infty[$	$\dfrac{1}{\pi b}\dfrac{1}{1 + \left(\frac{x-a}{b}\right)^2}$	$\mathrm{e}^{iat - b\lvert t\rvert}$
Triangulaire		$(1 - \lvert x\rvert)\mathbf{1}_{[-1,1]}(x)$	$\dfrac{2(1 - \cos t)}{t^2}$
$\mathfrak{N}(m, \sigma)$	$m \in \mathbb{R},\ \sigma \in]0, +\infty[$	$\dfrac{1}{\sigma\sqrt{2\pi}}\exp\left(\dfrac{-(x-m)^2}{2\sigma^2}\right)$	$\exp\left(imt - \dfrac{\sigma^2 t^2}{2}\right)$

Le calcul de la fonction caractéristique de la loi gaussienne $\mathfrak{N}(m, \sigma)$ n'est pas immédiat. Il est détaillé ci-dessous (proposition 9.6).

Proposition 9.2. *La transformée de Fourier $\widehat{\mu}$ d'une mesure finie μ est une fonction continue et bornée sur $\mathbb{R}^d$. On a*

$$\forall t \in \mathbb{R}^d, \quad \lvert\widehat{\mu}(t)\rvert \le \mu(\mathbb{R}^d) = \widehat{\mu}(0). \tag{9.3}$$

Vérification. Pour la bornitude, il suffit d'écrire

$$|\widehat{\mu}(t)| = \left| \int_{\mathbb{R}^d} \exp\bigl(i\langle t, x\rangle\bigr)\, \mathrm{d}\mu(x) \right| \leq \int_{\mathbb{R}^d} \bigl|\exp\bigl(i\langle t, x\rangle\bigr)\bigr|\, \mathrm{d}\mu(x) = \int_{\mathbb{R}^d} \mathrm{d}\mu(x) = \mu(\mathbb{R}^d) < +\infty.$$

La continuité est une application immédiate du théorème de continuité sous le signe somme avec pour fonction dominante la constante 1. □

Proposition 9.3. *Soit X une variable aléatoire réelle ayant un moment d'ordre $r \in \mathbb{N}^*$ (i.e. $\mathbf{E}|X|^r < +\infty$) et ϕ sa fonction caractéristique. Alors ϕ est r fois dérivable sur $\mathbb{R}$ et*

$$\forall t \in \mathbb{R}, \quad \phi^{(k)}(t) = i^k \mathbf{E}\left(X^k \mathrm{e}^{itX}\right), \quad 1 \leq k \leq r.$$

En particulier

$$\phi^{(k)}(0) = i^k \mathbf{E} X^k.$$

Vérification. C'est une application du théorème de dérivation sous le signe somme en notant que si $\mathbf{E}|X|^r$ est fini, $\mathbf{E}|X|^k$ l'est aussi pour tout $k \leq r$. Cette remarque jointe à la majoration évidente

$$\left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} \exp\bigl(itX(\omega)\bigr) \right| = \left| i^k X(\omega)^k \mathrm{e}^{itX(\omega)} \right| \leq |X(\omega)|^k,$$

nous permet lors de la k-ième dérivation, d'utiliser $|X|^k$ comme fonction dominante indépendante de t et $\mathbf{P}$-intégrable sur Ω. □

Proposition 9.4. *Si la mesure μ est une mesure produit $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ de mesures finies sur $\mathbb{R}$,*

$$\forall t = (t_1, \ldots, t_d) \in \mathbb{R}^d, \quad \widehat{\mu}(t) = \prod_{j=1}^{d} \widehat{\mu}_j(t_j), \tag{9.4}$$

autrement dit,

$$\widehat{\mu} = \widehat{\mu}_1 \otimes \cdots \otimes \widehat{\mu}_d.$$

Démonstration. Comme $\mu(\mathbb{R}^d) = \mu_1(\mathbb{R}) \times \cdots \times \mu_d(\mathbb{R})$, μ est aussi une mesure finie et $\widehat{\mu}$ est donc bien définie. Comme $\int_{\mathbb{R}^d} |\exp(i\langle t, x\rangle)|\, \mathrm{d}\mu(x) \leq \mu(\mathbb{R}^d)$, la fonction $x \mapsto \exp(i\langle t, x\rangle)$ est $\mu_1 \otimes \cdots \otimes \mu_d$ intégrable et on peut justifier par le corollaire 5.20 le calcul suivant :

$$\begin{aligned}
\widehat{\mu}(t) &= \int_{\mathbb{R}^d} \exp\bigl(i(t_1 x_1 + \cdots + t_d x_d)\bigr)\, \mathrm{d}(\mu_1 \otimes \cdots \otimes \mu_d)(x_1, \ldots, x_d) \\
&= \int_{\mathbb{R}^d} \prod_{j=1}^{d} \exp(it_j x_j)\, \mathrm{d}(\mu_1 \otimes \cdots \otimes \mu_d)(x_1, \ldots, x_d) \\
&= \prod_{j=1}^{d} \int_{\mathbb{R}} \exp(it_j x_j)\, \mathrm{d}\mu_j(x_j) \\
&= \prod_{j=1}^{d} \widehat{\mu}_j(t_j)
\end{aligned}$$

qui établit (9.4). □

Dans le langage des fonctions caractéristiques, le résultat que nous venons de vérifier se traduit comme suit.

Corollaire 9.5. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire à composantes* indépendantes*, notons φ sa fonction caractéristique et pour $j = 1, \dots, d$, φ_j celle de la variable aléatoire réelle X_j. Alors*

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \varphi(t) = \varphi_1(t_1) \dots \varphi_d(t_d),$$

autrement dit

$$\varphi = \varphi_1 \otimes \cdots \otimes \varphi_d.$$

Vérification. Si X est à composantes indépendantes, $P_X = P_{X_1} \otimes \cdots \otimes P_{X_d}$. □

9.1.2 Fonction caractéristique de la loi normale standard

Pour des raisons qui apparaîtront progressivement, il importe de connaître la fonction caractéristique de la loi gaussienne standard $\mathfrak{N}(0, 1)$.

Proposition 9.6. *La loi normale $\mathfrak{N}(0, 1)$ de densité $f : x \mapsto (2\pi)^{-1/2} \exp(-x^2/2)$ a pour fonction caractéristique $\varphi = \sqrt{2\pi} f$:*

$$\forall t \in \mathbb{R}, \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{itx} \mathrm{e}^{-x^2/2} \, \mathrm{d}\lambda(x) = \mathrm{e}^{-t^2/2}. \tag{9.5}$$

Plus généralement, si X suit la loi gaussienne $\mathfrak{N}(m, \sigma)$, de densité

$$f_{m,\sigma} : x \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-m)^2}{2\sigma^2} \right),$$

sa fonction caractéristique $\varphi_{m,\sigma}$ est donnée par

$$\forall t \in \mathbb{R}, \quad \varphi_{m,\sigma}(t) = \exp\Big(imt - \frac{\sigma^2 t^2}{2} \Big). \tag{9.6}$$

Des différentes méthodes possibles pour vérifier (9.5), la plus élégante est sans doute celle qui utilise l'analyse complexe[1]. Celle un peu plus laborieuse que nous exposons ci-dessous, a cependant le mérite de n'utiliser que des techniques du cours d'intégration. L'idée est de développer en série entière e^{itx} et d'intégrer terme à terme. Nous aurons besoin pour cela de connaître les moments de la loi $\mathfrak{N}(0, 1)$.

1. L'intégrale de la fonction *entière* $\exp(-z^2/2)$ le long du contour rectangulaire de sommets R, $R+it$, $-R+it$, $-R$, R est nulle. Elle le reste quand R tend vers l'infini et les intégrales sur les segments verticaux tendent vers 0. En paramétrant les segments horizontaux du contour, on en déduit (9.5). La rédaction complète à partir de cette esquisse est un exercice recommandé aux étudiants suivant le module Variable Complexe.

Lemme 9.7 (Moments de $\mathfrak{N}(0,1)$). *Si la variable aléatoire réelle X suit la loi $\mathfrak{N}(0,1)$, elle a des moments de tout ordre. Ses moments d'ordre impair sont nuls et*

$$\forall k \in \mathbb{N}^*, \quad \mathbf{E}X^{2k} = \prod_{j=1}^{k}(2j-1) = \frac{(2k)!}{2^k k!}. \tag{9.7}$$

Preuve. Pour vérifier l'existence des moments, on écrit grâce au transfert de Ω vers $\mathbb{R}$ par X,

$$\mathbf{E}|X|^n = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^n \mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x|^n \mathrm{e}^{-x^2/2} \,\mathrm{d}x < +\infty, \tag{9.8}$$

la conversion de l'intégrale de Lebesgue en intégrale de Riemann étant justifiée par le fait que la fonction à intégrer est continue sur $\mathbb{R}$ et d'intégrale généralisée absolument convergente. Pour cette convergence, il suffit d'écrire $|x|^n \mathrm{e}^{-x^2/2} = |x|^n \mathrm{e}^{-x^2/4}\mathrm{e}^{-x^2/4}$, de noter que $|x|^n \mathrm{e}^{-x^2/4}$ tend vers 0 quand x tend vers $\pm\infty$ et que $\int_{-\infty}^{+\infty} \mathrm{e}^{-x^2/4} \,\mathrm{d}x$ converge en raison de l'inégalité $\mathrm{e}^{-x^2/4} \leq \mathrm{e}^{-|x|/4}$ pour $|x| \geq 1$, la convergence de $\int_{-\infty}^{+\infty} \mathrm{e}^{-|x|/4} \,\mathrm{d}x$ étant évidente par recours aux primitives. L'existence des moments de tout ordre est établie par (9.8) et le même argument pour la conversion de l'intégrale de Lebesgue en intégrale de Riemann nous permet maintenant d'écrire

$$\forall n \in \mathbb{N}, \quad \mathbf{E}X^n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n \mathrm{e}^{-x^2/2} \,\mathrm{d}x.$$

Quand n est impair, la fonction à intégrer est impaire et cette intégrale de Riemann généralisée absolument convergente[2] est nulle. Reste à calculer les moments d'ordre pair : $c_k := \mathbf{E}X^{2k}$, $k \in \mathbb{N}$. On sait déjà que $c_0 = 1$. L'idée est de chercher une relation de récurrence en intégrant par parties.

$$\frac{1}{\sqrt{2\pi}} \int_a^b x^{2k-2} \mathrm{e}^{-x^2/2} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \left[\frac{x^{2k-1}}{2k-1} \mathrm{e}^{-x^2/2} \right]_a^b - \frac{1}{\sqrt{2\pi}} \int_a^b \frac{x^{2k-1}}{2k-1} (-x) \mathrm{e}^{-x^2/2} \,\mathrm{d}x.$$

En faisant tendre a vers $-\infty$ et b vers $+\infty$, on en déduit

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k-2} \mathrm{e}^{-x^2/2} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}(2k-1)} \int_{-\infty}^{+\infty} x^{2k} \mathrm{e}^{-x^2/2} \,\mathrm{d}x,$$

d'où la relation de récurrence

$$\forall k \in \mathbb{N}^*, \quad c_k = (2k-1)c_{k-1}.$$

On en déduit

$$c_k c_{k-1} \dots c_2 c_1 = (2k-1)(2k-3) \dots 3 c_{k-1} c_{k-2} \dots c_1 c_0,$$

2. La convergence est indispensable ici car sinon on pourrait écrire « $\int_{-\infty}^{+\infty} x \,\mathrm{d}x = 0$ », ce qui est une ânerie...

d'où après simplification (aucun des c_j n'est nul), la première égalité de (9.7). Pour la deuxième égalité, il suffit de remarquer que le produit des facteurs *pairs* dans le développement de $(2k)!$ s'écrit

$$2k \times 2(k-1) \times \cdots \times (2 \times 2) \times (2 \times 1) = 2^k k!,$$

ce qui permet d'exprimer le produit des facteurs *impairs* par :

$$\frac{(2k)!}{2^k k!} = (2k-1)(2k-3)\dots 1 = \prod_{j=1}^{k}(2j-1).$$

□

Preuve de la proposition 9.6. Soit X une variable aléatoire de loi $\mathfrak{N}(0,1)$. Sa fonction caractéristique φ peut d'écrire grâce au transfert $(\Omega, \mathbf{P}) \xrightarrow{X} (\mathbb{R}, P_X)$ sous la forme :

$$\varphi(t) = \mathbf{E}\mathrm{e}^{itX} = \int_{\mathbb{R}} \mathrm{e}^{itx} \,\mathrm{d}P_X(x) = \int_{\mathbb{R}} \mathrm{e}^{itx} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x).$$

Dans tout ce qui suit, on travaille avec t quelconque, mais fixé. Posons pour tout $x \in \mathbb{R}$,

$$S_n(x) := \sum_{l=0}^{n} \frac{(itx)^l}{l!} \mathrm{e}^{-x^2/2}.$$

Comme la fonction exponentielle complexe est développable en série entière avec rayon de convergence infini, on a clairement :

$$\forall x \in \mathbb{R}, \quad S_n(x) \xrightarrow[n\to+\infty]{} \mathrm{e}^{itx}\mathrm{e}^{-x^2/2}. \tag{9.9}$$

On voit que cette convergence est *dominée* en écrivant :

$$\forall x \in \mathbb{R}, \quad \bigl|S_n(x)\bigr| \leq \sum_{l=0}^{n} \frac{|tx|^l}{l!} \mathrm{e}^{-x^2/2} \leq \mathrm{e}^{|tx|}\mathrm{e}^{-x^2/2} = \exp\Bigl(-\frac{x^2}{2} + |tx|\Bigr) =: g(x).$$

Cette fonction g ne dépend pas de n. Comme elle est continue sur $\mathbb{R}$, on obtient sa λ-intégrabilité en vérifiant que l'intégrale de Riemann généralisée $\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x$ est absolument convergente. Ceci résulte de la majoration $0 \leq g(x) \leq \exp(-|tx|)$ valable au voisinage de $-\infty$ et de $+\infty$ (plus précisément dès que $|x| \geq 4|t|$). Compte-tenu de (9.9), on peut donc appliquer le théorème de convergence dominée pour obtenir :

$$\lim_{n\to+\infty} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} S_n \,\mathrm{d}\lambda = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{itx}\mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x). \tag{9.10}$$

Par ailleurs, par linéarité de l'intégrale (et grâce à l'existence des moments de tout ordre pour la loi $\mathfrak{N}(0,1)$), on a

$$\forall n \in \mathbb{N}, \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} S_n \,\mathrm{d}\lambda = \sum_{l=0}^{n} \frac{1}{\sqrt{2\pi}} \frac{(it)^l}{l!} \int_{\mathbb{R}} x^l \mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x).$$

En reportant ceci dans (9.10), on aboutit à

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{itx} \mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x) = \sum_{l=0}^{+\infty} \frac{(it)^l}{l!} \int_{\mathbb{R}} x^l \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x).$$

En utilisant maintenant le calcul des moments vu au lemme 9.7, il vient :

$$\varphi(t) = \sum_{k=0}^{+\infty} \frac{(it)^{2k}}{(2k)!} \mathbf{E} X^{2k} = \sum_{k=0}^{+\infty} (-1)^k \frac{t^{2k}}{(2k)!} \frac{(2k)!}{2^k k!} = \sum_{k=0}^{+\infty} \frac{(-t^2/2)^k}{k!} = \mathrm{e}^{-t^2/2}.$$

Comme t était quelconque, cette formule est vraie pour tout $t \in \mathbb{R}$ et (9.5) est démontrée.

On en déduit le calcul de $\varphi_{m,\sigma}$ en remarquant que si X suit la loi $\mathfrak{N}(m, \sigma)$, X a *même loi* que $\sigma Y + m$ où Y suit la loi $\mathfrak{N}(0, 1)$. Par conséquent, X a même fonction caractéristique que $\sigma Y + m$ et

$$\varphi_{m,\sigma}(t) = \mathbf{E} \exp\bigl(it(\sigma Y + m)\bigr) = \mathbf{E}\bigl(\mathrm{e}^{imt} \mathrm{e}^{it\sigma Y}\bigr) = \mathrm{e}^{imt} \mathbf{E}\bigl(\mathrm{e}^{it\sigma Y}\bigr) = \mathrm{e}^{imt} \varphi(\sigma t),$$

d'où l'égalité (9.6). □

9.1.3 Transformée d'un produit de convolution

Une propriété essentielle de la transformation de Fourier est qu'elle permet de ramener le produit de convolution à un produit ordinaire de deux fonctions.

Théorème 9.8. *Si μ et ν sont deux mesures finies sur $\bigl(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\bigr)$,*

$$\forall t \in \mathbb{R}^d, \quad \widehat{\mu * \nu}(t) = \widehat{\mu}(t) \widehat{\nu}(t). \tag{9.11}$$

De même si X et Y sont deux vecteurs aléatoires dans $\mathbb{R}^d$, indépendants, *la fonction caractéristique φ_{X+Y} de leur somme est donnée par*

$$\forall t \in \mathbb{R}^d, \quad \varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t). \tag{9.12}$$

Démonstration. Rappelons que $\mu * \nu$ est la mesure sur $\bigl(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\bigr)$, définie comme la mesure image de $\mu \otimes \nu$ (mesure sur $\bigl(\mathbb{R}^d \times \mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d) \otimes \mathrm{Bor}(\mathbb{R}^d)\bigr)$) par

$$s : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d, \quad (y, z) \mapsto y + z.$$

En particulier, une fonction borélienne $f : \mathbb{R}^d \to \mathbb{C}$ est $\mu * \nu$ intégrable si et seulement si $\tilde{f} : (y, z) \mapsto f(y + z)$ est $\mu \otimes \nu$ intégrable. Dans ce cas on a :

$$\int_{\mathbb{R}^d} f(x) \,\mathrm{d}(\mu * \nu)(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(y + z) \,\mathrm{d}(\mu \otimes \nu)(y, z).$$

Prenons comme fonction $f : x \mapsto \exp(i\langle t, x\rangle)$ pour $t \in \mathbb{R}^d$ fixé. La fonction $\tilde{f}$ est borélienne bornée sur $\mathbb{R}^d \times \mathbb{R}^d$, donc intégrable par rapport à la mesure finie $\mu \otimes \nu$. Par application du théorème de Fubini (plus précisément du corollaire 5.13), on obtient :

$$\begin{aligned}\widehat{\mu * \nu}(t) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle)\, \mathrm{d}(\mu * \nu)(x) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \exp(i\langle t, y + z\rangle)\, \mathrm{d}(\mu \otimes \nu)(y, z) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \exp(i\langle t, y\rangle) \exp(i\langle t, z\rangle)\, \mathrm{d}(\mu \otimes \nu)(y, z) \\ &= \int_{\mathbb{R}^d} \exp(i\langle t, y\rangle)\, \mathrm{d}\mu(y) \int_{\mathbb{R}^d} \exp(i\langle t, z\rangle)\, \mathrm{d}\nu(z) \\ &= \widehat{\mu}(t)\widehat{\nu}(t).\end{aligned}$$

Comme t était quelconque, cette égalité est vraie pour tout t et (9.11) est établie. Pour en déduire (9.12), il suffit de rappeler que si X et Y sont *indépendants*, la loi de leur somme est le produit de convolution : $P_{X+Y} = P_X * P_Y$. □

9.1.4 Caractérisation des mesures finies

Théorème 9.9. *Si deux mesures finies sur* $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$ *ont même transformée de Fourier, elles sont égales.*

Démonstration. Il s'agit de montrer que l'égalité de *fonctions* $\widehat{\mu} = \widehat{\nu}$ implique l'égalité de mesures $\mu = \nu$.

Première étape. On réduit la preuve de l'égalité $\mu = \nu$ à celle de :

$$\forall f : \mathbb{R}^d \to \mathbb{R}, \text{ continue à support compact}, \quad \int_{\mathbb{R}^d} f \, \mathrm{d}\mu = \int_{\mathbb{R}^d} f \, \mathrm{d}\nu. \tag{9.13}$$

Supposons donc (9.13) vraie et soit

$$C := \prod_{j=1}^{d}]a_j, b_j[,$$

un pavé ouvert non vide de $\mathbb{R}^d$ (on a donc $a_j < b_j$ pour tout j). Définissons pour tout $n \geq n_0$, tel que $2/n_0 < \min_{1 \leq j \leq d}(b_j - a_j)$, la fonction continue à support compact f_n par

$$f_n := f_{n,1} \otimes \cdots \otimes f_{n,d},$$

où $f_{n,j}$ est l'unique fonction continue $\mathbb{R} \to [0, 1]$, valant 1 sur $[a_j + 1/n, b_j - 1/n]$, 0 hors de $]a_j, b_j[$ et affine sur chacun des intervalles $[a_j, a_j + 1/n]$ et $[b_j - 1/n, b_j]$, voir la figure 9.1.

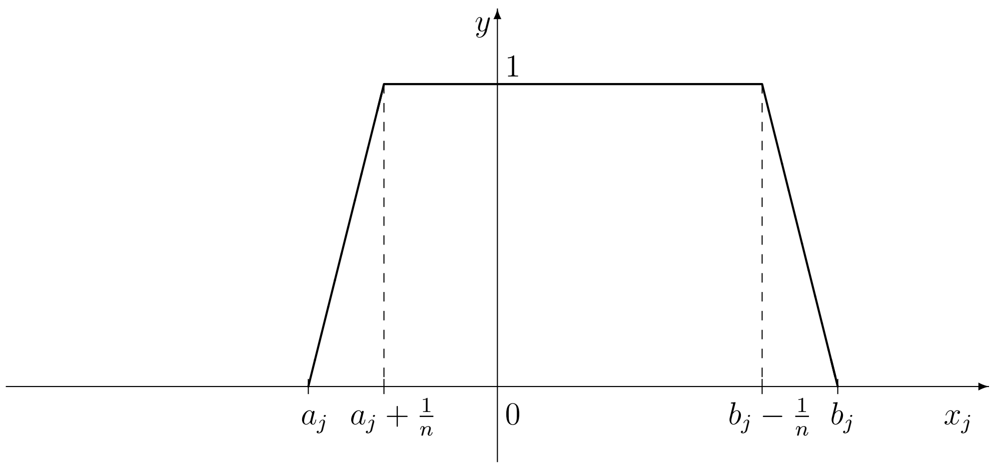

FIG. 9.1 – Fonction $f_{n,j}$

On voit facilement que :

$$\forall x \in \mathbb{R}^d, \quad f_n(x) \xrightarrow[n \to +\infty]{} \mathbf{1}_C(x).$$

Cette convergence ponctuelle[3] sur $\mathbb{R}^d$ est *dominée* par la fonction $\mathbf{1}_C$ qui est bornée, donc μ-intégrable puisque μ est une mesure finie. Par le théorème de convergence dominée, on en déduit :

$$\int_{\mathbb{R}^d} f_n \, \mathrm{d}\mu \xrightarrow[n \to +\infty]{} \int_{\mathbb{R}^d} \mathbf{1}_C \, \mathrm{d}\mu = \mu(C). \tag{9.14}$$

On a évidemment la même convergence d'intégrale avec ν à la place de μ. D'autre part, f_n étant continue à support compact, notre hypothèse provisoire (9.13) nous permet d'écrire :

$$\forall n \geq n_0, \quad \int_{\mathbb{R}^d} f_n \, \mathrm{d}\mu = \int_{\mathbb{R}^d} f_n \, \mathrm{d}\nu.$$

Compte-tenu de (9.14), on peut faire tendre n vers l'infini dans cette égalité pour obtenir $\mu(C) = \nu(C)$. Comme C était quelconque, on a établi la coïncidence des mesures finies μ et ν sur le π-système des pavés ouverts qui engendre la tribu borélienne de $\mathbb{R}^d$, donc μ et ν sont égales. $\square$

Deuxième étape. On réduit la preuve de (9.13) à celle de :

$$\forall f : \mathbb{R}^d \to \mathbb{R} \text{ continue bornée}, \quad \forall n \in \mathbb{N}^*, \quad \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu * \gamma_n) = \int_{\mathbb{R}^d} f \, \mathrm{d}(\nu * \gamma_n), \tag{9.15}$$

3. Remarquons qu'il est important ici d'avoir la convergence pour *tout* $x \in \mathbb{R}^d$ et pas seulement μ-presque partout. En effet la construction de f_n ne dépend aucunement de la mesure μ et le résultat doit être valable pour n'importe quelle mesure finie μ. S'il y avait ne serait-ce qu'un point x_0 tel que $f_n(x_0)$ ne converge pas vers $\mathbf{1}_C(x_0)$, il suffirait de prendre une mesure μ ayant une masse ponctuelle en x_0 pour que la convergence μ-p.p. de f_n vers $\mathbf{1}_C$ soit en défaut.

où $\gamma_n := \mathfrak{N}(0, n^{-1/2})^{\otimes d}$ est une mesure gaussienne. Pour cela il suffit de montrer que

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_n) = \int_{\mathbb{R}^d} f \,\mathrm{d}\mu. \tag{9.16}$$

Remarquons d'abord que γ_n est la mesure image $\gamma_1 \circ h_n^{-1}$, de γ_1 par l'homothétie $h_n : x \mapsto n^{-1/2}x$. On en déduit que toute la masse de γ_n se concentre au voisinage du point 0 quand n tend vers l'infini. Plus précisément :

$$\forall \delta > 0, \quad \lim_{n\to+\infty} \gamma_n\big(\mathbb{R}^d \setminus]-\delta, \delta[^d\big) = 0. \tag{9.17}$$

En effet

$$\gamma_n\big(]-\delta,\delta[^d\big) = \gamma_1\big(h_n^{-1}(]-\delta,\delta[^d)\big) = \gamma_1\big(]-\delta n^{1/2}, \delta n^{1/2}[^d\big) \xrightarrow[n\to+\infty]{} \gamma_1(\mathbb{R}^d) = 1,$$

par continuité séquentielle croissante de la mesure de probabilité γ_1.

En exploitant (9.17) et la continuité de f, nous allons montrer que :

$$\forall x \in \mathbb{R}^d, \quad \int_{\mathbb{R}^d} f(x+z)\,\mathrm{d}\gamma_n(z) \xrightarrow[n\to+\infty]{} f(x). \tag{9.18}$$

En effet, fixons x quelconque dans $\mathbb{R}^d$, il existe pour tout $\varepsilon > 0$, un $\delta > 0$, dépendant de f, de x et de ε tel que

$$\forall z \in]-\delta, \delta[^d, \quad |f(x+z) - f(x)| < \varepsilon. \tag{9.19}$$

Comme γ_n est une probabilité sur $\mathbb{R}^d$, $f(x) = \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\gamma_n(z)$, d'où

$$\begin{aligned}
\Big|\int_{\mathbb{R}^d} f(x+z)\,\mathrm{d}\gamma_n(z) - f(x)\Big| &= \Big|\int_{\mathbb{R}^d} \big(f(x+z) - f(x)\big)\,\mathrm{d}\gamma_n(z)\Big| \\
&\leq \int_{\mathbb{R}^d} \big|f(x+z) - f(x)\big|\,\mathrm{d}\gamma_n(z) \\
&\leq \int_{]-\delta,\delta[^d} \varepsilon\,\mathrm{d}\gamma_n + 2\|f\|_\infty \gamma_n\big(\mathbb{R}^d \setminus]-\delta,\delta[^d\big),
\end{aligned}$$

où l'on a posé $\|f\|_\infty := \sup_{\mathbb{R}^d} |f|$. Grâce à (9.17), il existe un entier n_0 tel que

$$\forall n \geq n_0, \quad 2\|f\|_\infty \gamma_n\big(\mathbb{R}^d \setminus]-\delta,\delta[^d\big) < \varepsilon.$$

Par conséquent,

$$\forall n \geq n_0, \quad \Big|\int_{\mathbb{R}^d} f(x+z)\,\mathrm{d}\gamma_n(z) - f(x)\Big| < \varepsilon\gamma_n\big(]-\delta,\delta[^d\big) + \varepsilon < 2\varepsilon.$$

La convergence (9.18) est ainsi établie [4].

4. Le lecteur attentif aura noté la ressemblance non fortuite avec la preuve du théorème de Fejér.

Comme γ_n est une probabilité, on a la majoration uniforme :

$$\forall x \in \mathbb{R}^d, \quad \Big| \int_{\mathbb{R}^d} f(x+z) \,\mathrm{d}\gamma_n(z) \Big| \leq \|f\|_\infty. \tag{9.20}$$

La mesure μ étant finie, la constante $\|f\|_\infty$ est μ-intégrable. Grâce à (9.18) et (9.20), on peut appliquer le théorème de convergence dominée relativement à la mesure μ pour obtenir :

$$\int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_n) = \int_{\mathbb{R}^d} \Big\{ \int_{\mathbb{R}^d} f(x+z) \,\mathrm{d}\gamma_n(z) \Big\} \,\mathrm{d}\mu(x) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f(x) \,\mathrm{d}\mu(x).$$

Ainsi nous avons établi (9.16) pour toute fonction f continue bornée, donc *a fortiori* pour toute fonction continue à support compact. □

Troisième étape : preuve de (9.15). Il est temps de rappeler que l'hypothèse du théorème en cours de démonstration est l'égalité $\widehat{\mu}(t) = \widehat{\nu}(t)$ pour tout $t \in \mathbb{R}^d$. Pour achever la preuve du théorème, il nous reste à déduire de cette hypothèse l'égalité $\int f \,\mathrm{d}(\mu * \gamma_n) = \int f \,\mathrm{d}(\nu * \gamma_n)$. L'idée est d'exprimer $\int f \,\mathrm{d}(\mu * \gamma_n)$ comme une intégrale où μ n'intervient que par sa transformée de Fourier $\widehat{\mu}$. Il suffira alors de lui substituer $\widehat{\nu}$ pour conclure. Le point crucial est que la transformée de Fourier de la mesure gaussienne γ_n est, à un changement d'échelle près, la densité g_n de γ_n. Plus précisément, $\gamma_n = \mathfrak{N}(0, n^{-1/2})^{\otimes d}$ a pour densité

$$g_n(x) = \Big(\frac{n}{2\pi}\Big)^{d/2} \exp\Big(-\frac{n\|x\|^2}{2}\Big),$$

où $\|x\|$ désigne la norme euclidienne de $\mathbb{R}^d$. En combinant les propositions 9.4 et 9.6 on en déduit immédiatement que

$$\widehat{\gamma_n}(t) = \prod_{j=1}^{d} \exp\Big(-\frac{t_j^2}{2n}\Big) = \exp\Big(-\frac{\|t\|^2}{2n}\Big).$$

En posant $c_n := \big(\frac{n}{2\pi}\big)^{d/2}$, on voit ainsi que

$$\forall z \in \mathbb{R}^d, \quad g_n(z) = c_n \widehat{\gamma_n}(nz). \tag{9.21}$$

On en déduit que

$$\begin{aligned} \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_n) &= \int_{\mathbb{R}^d} \Big\{ \int_{\mathbb{R}^d} f(x+z) g_n(z) \,\mathrm{d}\lambda_d(z) \Big\} \,\mathrm{d}\mu(x) \\ &= c_n \int_{\mathbb{R}^d} \Big\{ \int_{\mathbb{R}^d} f(x+z) \widehat{\gamma_n}(nz) \,\mathrm{d}\lambda_d(z) \Big\} \,\mathrm{d}\mu(x). \end{aligned} \tag{9.22}$$

Le changement de variable $y = x + z$ dans l'intégrale relativement à z dans (9.22) nous donne :

$$\int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_n) = c_n \int_{\mathbb{R}^d} \Big\{ \int_{\mathbb{R}^d} f(y) \widehat{\gamma_n}(ny - nx) \,\mathrm{d}\lambda_d(y) \Big\} \,\mathrm{d}\mu(x)$$

En remarquant que $\widehat{\gamma_n}(ny-nx) = \int_{\mathbb{R}^d} e^{inyu}e^{-inxu}\,d\gamma_n(u)$, on obtient (sans problème pour la justification de l'interversion des intégrations via le théorème de Fubini puisque f est continue à support compact donc $\int_{\mathbb{R}^d}|f|\,d\lambda_d < +\infty$) :

$$\begin{aligned}
\int_{\mathbb{R}^d} f\,d(\mu*\gamma_n) &= c_n\int_{\mathbb{R}^d} f(y)\int_{\mathbb{R}^d} e^{inyu}\int_{\mathbb{R}^d} e^{-inxu}\,d\mu(x)\,d\gamma_n(u)\,d\lambda_d(y) \\
&= c_n\int_{\mathbb{R}^d} f(y)\int_{\mathbb{R}^d} e^{inyu}\widehat{\mu}(-nu)\,d\gamma_n(u)\,d\lambda_d(y) && (9.23)\\
&= c_n\int_{\mathbb{R}^d} f(y)\int_{\mathbb{R}^d} e^{inyu}\widehat{\nu}(-nu)\,d\gamma_n(u)\,d\lambda_d(y) && (9.24)\\
&= \int_{\mathbb{R}^d} f\,d(\nu*\gamma_n). && (9.25)
\end{aligned}$$

Le passage de (9.23) à (9.24) utilise l'hypothèse $\widehat{\mu} = \widehat{\nu}$, celui de (9.24) à (9.25) exploite le fait que le calcul menant à la formule (9.23) pour $\int_{\mathbb{R}^d} f\,d(\mu*\gamma_n)$ est valable pour n'importe quelle mesure finie μ, donc en particulier pour ν. □

Le théorème 9.9 est maintenant complètement démontré. □

Voici une première application du théorème 9.9 à la caractérisation de l'indépendance des variables aléatoires par les fonctions caractéristiques.

Proposition 9.10. *Soit $X = (X_1,\dots,X_d)$ un vecteur aléatoire, notons φ sa fonction caractéristique et φ_j celle de X_j $(j = 1,\dots,d)$. Si*

$$\forall t = (t_1,\dots,t_d) \in \mathbb{R}^d, \quad \varphi(t) = \varphi_1(t_1)\dots\varphi_d(t_d), \tag{9.26}$$

alors les composantes de X sont mutuellement indépendantes.

Preuve. Notons ψ la fonction caractéristique de la loi $\nu := P_{X_1}\otimes\dots\otimes P_{X_d}$. Par la proposition 9.4, $\psi(t) = \varphi_1(t_1)\dots\varphi_d(t_d)$, donc par l'hypothèse (9.26), $\psi = \varphi$, ce qui s'écrit aussi $\widehat{\nu} = \widehat{P_X}$. Le théorème 9.9 nous dit alors que $P_X = \nu$. Nous avons donc vérifié que $P_X = P_{X_1}\otimes\dots\otimes P_{X_d}$, ce qui équivaut à l'indépendance des X_j. □

Comme la proposition 9.10 est la réciproque de la proposition 9.4, l'égalité (9.26) est une condition *nécessaire et suffisante* pour l'indépendance mutuelle de $X_1,\dots,X_d$.

Pour finir cette section, on donne une application des théorèmes 9.8 et 9.9 au calcul de lois de sommes de v.a. indépendantes.

Exemple 9.3 (Trois familles de lois stables par convolution). La famille des lois gaussiennes sur $\mathbb{R}$, celle des lois de Poisson et celle des lois de Cauchy sont chacune stable par convolution. Plus précisément, on a les relations :

$$\begin{aligned}
\mathrm{Pois}(\alpha)*\mathrm{Pois}(\beta) &= \mathrm{Pois}(\alpha+\beta) && (9.27)\\
\mathfrak{N}(m_1,\sigma_1)*\mathfrak{N}(m_2,\sigma_2) &= \mathfrak{N}\big(m_1+m_2,(\sigma_1^2+\sigma_2^2)^{1/2}\big) && (9.28)\\
\mathrm{Cau}(a_1,b_1)*\mathrm{Cau}(a_2,b_2) &= \mathrm{Cau}(a_1+a_2,b_1+b_2). && (9.29)
\end{aligned}$$

Justifions (9.28), la vérification de (9.27) et (9.29) étant analogue sera laissée au lecteur. Pour alléger les écritures, notons $\mu_j := \mathfrak{N}(m_j, \sigma_j)$. En appliquant le théorème 9.8 et la proposition 9.6, on voit que :

$$\begin{aligned}
\forall t \in \mathbb{R}, \quad \widehat{\mu_1 * \mu_2}(t) = \widehat{\mu_1}(t)\widehat{\mu_2}(t) &= \exp\left(im_1 t - \frac{\sigma_1^2 t^2}{2}\right)\exp\left(im_2 t - \frac{\sigma_2^2 t^2}{2}\right) \\
&= \exp\left(i(m_1+m_2)t - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}\right) \\
&= \widehat{\mu}(t),
\end{aligned}$$

où $\mu := \mathfrak{N}\big(m_1+m_2, (\sigma_1^2+\sigma_2^2)^{1/2}\big)$. Ainsi $\mu_1 * \mu_2$ et μ ont même transformée de Fourier, donc d'après le théorème 9.9, $\mu_1 * \mu_2 = \mu$, ce qui n'est que l'écriture allégée de (9.28). Un énoncé équivalent à (9.28) est : *si X_1 et X_2 sont deux variables aléatoires gaussiennes* indépendantes *de paramètres respectifs (m_1, σ_1) et (m_2, σ_2), alors $S := X_1 + X_2$ est aussi gaussienne de paramètres (m, σ) avec $m = m_1 + m_2$ et $\sigma^2 = \sigma_1^2 + \sigma_2^2$.* Remarquons que dès que l'on sait que S est gaussienne, les relations entre les paramètres peuvent se retrouver en se rappelant que $m = \mathbf{E}S$ et $\sigma^2 = \operatorname{Var} S$. Il suffit alors d'appliquer la linéarité de l'espérance et la relation $\operatorname{Var} S = \operatorname{Var} X_1 + \operatorname{Var} X_2$ pour X_1 et X_2 *indépendantes.*

9.2 Transformée de Fourier d'une fonction

Définition 9.11. *Soit $f \in \mathrm{L}^1(\lambda_d)$, on définit sa transformée de Fourier $\widehat{f}$ par*

$$\forall t \in \mathbb{R}^d, \quad \widehat{f}(t) := \int_{\mathbb{R}^d} f(x)\exp(i\langle t, x\rangle)\,\mathrm{d}\lambda_d(x). \tag{9.30}$$

Remarque 9.12. Si μ est une mesure finie de densité f par rapport à λ_d, alors $\widehat{\mu} = \widehat{f}$.

Remarque 9.13. Il y a dans la littérature diverses autres définitions de la transformée de Fourier d'une fonction. Leur forme générale est $\widehat{f}(t) := a\int_{\mathbb{R}^d} f(x)\exp(ib\langle t, x\rangle)\,\mathrm{d}\lambda_d(x)$, le choix des constantes réelles a et b étant motivé par l'utilisation principale envisagée pour la transformée de Fourier. Par exemple si on veut la considérer comme isométrie euclidienne de $\mathrm{L}^2(\mathbb{R})$, il est commode de prendre $a = (2\pi)^{-1/2}$ et $b = 1$ ou -1. Nous avons adopté le choix qui permet d'identifier transformée de Fourier d'une densité de probabilité et fonction caractéristique de la loi correspondante.

9.2.1 Propriétés de la transformation de Fourier dans L^1

Nous exposons maintenant les propriétés de la transformée de Fourier en nous restreignant pour simplifier un peu, au cas $d = 1$.

Proposition 9.14. *Soient $f \in \mathrm{L}^1(\mathbb{R})$, $a \in \mathbb{R}$ et $c > 0$ quelconques.*

i) Si $g(x) = f(x)\mathrm{e}^{iax}$, alors $\widehat{g}(t) = \widehat{f}(t+a)$.

ii) Si $g(x) = f(x-a)$, alors $\widehat{g}(t) = \widehat{f}(t)\mathrm{e}^{ita}$.

iii) Si $g(x) = f(x/c)$, alors $\widehat{g}(t) = c\widehat{f}(ct)$.

Vérification. Si $g(x) = f(x)e^{iax}$,

$$\widehat{g}(t) = \int_{\mathbb{R}} f(x)e^{iax}e^{itx}\, d\lambda_1(x) = \int_{\mathbb{R}} f(x)e^{i(a+t)x}\, d\lambda_1(t) = \widehat{f}(t+a).$$

Si $g(x) = f(x-a)$, le changement de variable $y = x - a$ et l'invariance de λ_1 par translation nous donnent :

$$\widehat{g}(t) = \int_{\mathbb{R}} f(x-a)e^{itx}\, d\lambda_1(x) = \int_{\mathbb{R}} f(y)e^{it(y+a)}\, d\lambda_1(y) = e^{ita}\int_{\mathbb{R}} f(y)e^{ity}\, d\lambda_1(y) = \widehat{f}(t)e^{ita}.$$

Enfin si $g(x) = f(x/c)$, le changement de variable $y = x/c$ (la mesure image de λ_1 est alors $c\lambda_1$) donne :

$$\widehat{g}(t) = \int_{\mathbb{R}} f\Big(\frac{x}{c}\Big)e^{itx}\, d\lambda_1(x) = \int_{\mathbb{R}} f(y)e^{itcy}c\, d\lambda_1(y) = c\widehat{f}(ct).$$

□

Théorème 9.15. *Pour toutes f et $g \in L^1(\mathbb{R})$,*

$$\widehat{f*g} = \widehat{f}\cdot\widehat{g}. \tag{9.31}$$

Preuve. Rappelons (définition 5.56) que si f et g sont dans $L^1(\mathbb{R})$, leur produit de convolution est la classe de fonctions dont un représentant est défini λ-p.p. par :

$$(f*g)(y) := \int_{\mathbb{R}} f(x)g(y-x)\, d\lambda_1(x) = \int_{\mathbb{R}} f(y-x)g(x)\, d\lambda_1(x).$$

On sait (proposition 5.57) que si $f, g \in L^1(\mathbb{R})$, alors $f*g$ est aussi un élément de L^1. Donc sa transformée de Fourier est bien définie et s'écrit :

$$\widehat{f*g}(t) = \int_{\mathbb{R}} e^{ity}(f*g)(y)\, d\lambda_1(y) = \int_{\mathbb{R}} e^{ity}\left\{\int_{\mathbb{R}} f(x)g(y-x)\, d\lambda_1(x)\right\} d\lambda_1(y).$$

Comme $|f|*|g|$ est aussi dans L^1, on peut appliquer le théorème de Fubini pour permuter l'ordre des intégrations :

$$\widehat{f*g}(t) = \int_{\mathbb{R}}\left\{\int_{\mathbb{R}} e^{ity}f(x)g(y-x)\, d\lambda_1(y)\right\} d\lambda_1(x).$$

Pour x fixé, on effectue dans l'intégrale intérieure le changement de variable $u = y - x$ qui nous donne (λ_1 étant invariante par translation) :

$$\begin{aligned}
\widehat{f*g}(t) &= \int_{\mathbb{R}}\left\{\int_{\mathbb{R}} e^{it(x+u)}f(x)g(u)\, d\lambda_1(u)\right\} d\lambda_1(x)\\
&= \int_{\mathbb{R}} e^{itx}f(x)\left\{\int_{\mathbb{R}} e^{itu}g(u)\, d\lambda_1(u)\right\} d\lambda_1(x)\\
&= \left\{\int_{\mathbb{R}} e^{itu}g(u)\, d\lambda_1(u)\right\}\int_{\mathbb{R}} e^{itx}f(x)\, d\lambda_1(x)\\
&= \widehat{f}(t)\widehat{g}(t).
\end{aligned}$$

□

Proposition 9.16 (Dérivée d'une transformée de Fourier). *Si f et la fonction $g : x \mapsto xf(x)$ sont toutes deux dans $\mathrm{L}^1(\mathbb{R})$, alors $\widehat{f}$ est dérivable et*

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{f}(t) = i\,\widehat{g}(t) \tag{9.32}$$

Vérification. C'est une application directe du théorème de dérivation sous le signe somme. □

Théorème 9.17 (Transformée de Fourier d'une dérivée). *Si f et f' sont dans L^1, alors*

$$\widehat{\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)}(t) = \widehat{f'}(t) = -it\widehat{f}(t). \tag{9.33}$$

Preuve. Examinons d'abord le cas particulier où f est C^1 à support compact $[a,b]$. Alors f et f' sont nulles aux points a et b et en dehors de $[a,b]$. On obtient facilement (9.33) en passant aux intégrales de Riemann et en intégrant par parties :

$$\begin{aligned}
\widehat{f'}(t) = \int_{[a,b]} \mathrm{e}^{itx} f'(x)\,\mathrm{d}\lambda_1(x) &= \int_a^b \mathrm{e}^{itx} f'(x)\,\mathrm{d}x = \left[\mathrm{e}^{itx} f(x)\right]_a^b - it\int_a^b \mathrm{e}^{itx} f(x)\,\mathrm{d}x \\
&= -it\int_{-\infty}^{+\infty} \mathrm{e}^{itx} f(x)\,\mathrm{d}x \\
&= -it\int_{\mathbb{R}} \mathrm{e}^{itx} f(x)\,\mathrm{d}\lambda_1(x).
\end{aligned}$$

Nous admettrons que l'on peut passer au cas général par un argument de densité. □

Remarque 9.18. Le théorème 9.17 énonce la deuxième propriété essentielle de la transformée de Fourier (la première étant l'effet sur la convolution). La transformée de Fourier transforme une dérivation en une multiplication par la fonction monôme $t \mapsto -it$. Plus généralement si D est un opérateur différentiel de la forme :

$$D := \sum_{k=1}^{n} c_k \frac{\mathrm{d}^k}{\mathrm{d}x^k},$$

on a pour toute f telle que f, $f', \dots, f^{(n)}$ soient dans L^1,

$$\widehat{Df}(t) = Q_n(t)\widehat{f}(t),$$

où Q est un polynôme. Cette propriété est utile pour la résolution d'équations différentielles de la forme $Df = g$ qui deviennent par transformée de Fourier $Q_n\widehat{f} = \widehat{g}$, d'où $\widehat{f} = Q_n^{-1}\widehat{g}$. On peut en déduire une solution f si l'on sait calculer la transformée de Fourier inverse de $Q_n^{-1}\widehat{g}$.

Théorème 9.19 (Riemann-Lebesgue). *Si $f \in \mathrm{L}^1(\mathbb{R})$, $\widehat{f} \in \mathcal{C}_0(\mathbb{R})$, autrement dit $\widehat{f}$ est continue sur $\mathbb{R}$ et tend vers 0 en $-\infty$ et en $+\infty$. On a*

$$\|\widehat{f}\|_\infty \leq \|f\|_1.$$

Preuve. La continuité de $\widehat{f}$ résulte d'une application immédiate du théorème de continuité sous le signe somme.

Pour la limite en $\pm\infty$, on peut écrire en remarquant que $\mathrm{e}^{i\pi} = -1$,

$$\forall t \in \mathbb{R}^*, \quad \widehat{f}(t) = -\int_{\mathbb{R}} f(x)\exp\bigl(it(x+\pi/t)\bigr)\,\mathrm{d}\lambda_1(x) = -\int_{\mathbb{R}} f(y-\pi/t)\mathrm{e}^{ity}\,\mathrm{d}\lambda_1(y).$$

On en déduit

$$2\widehat{f}(t) = \int_{\mathbb{R}} \bigl(f(y) - f(y-\pi/t)\bigr)\mathrm{e}^{ity}\,\mathrm{d}\lambda_1(y),$$

d'où

$$\forall t \in \mathbb{R}^*, \quad 2|\widehat{f}(t)| \leq \|f - f(\,.-\pi/t)\|_1.$$

On conclut en faisant tendre t vers $+\infty$ (ou $-\infty$) et avec le lemme suivant. □

Lemme 9.20. *Pour $1 \leq p < +\infty$, les translations* opèrent continûment *sur* $\mathrm{L}^p(\mathbb{R})$*, ce qui signifie que pour tout $f \in \mathrm{L}^p(\mathbb{R})$,*

$$\lim_{h\to 0} \|f - f(\,.+h)\|_p = 0. \tag{9.34}$$

Preuve. Vérifions d'abord (9.34) dans le cas particulier où f est continue à support compact inclus dans $[a,b]$. Dans ce cas on a pour tout $h \in [-1,1]$,

$$\|f - f(\,.+h)\|_p^p = \int_{[a-1,b+1]} |f(x) - f(x+h)|^p\,\mathrm{d}\lambda_1(x)$$

et on obtient facilement (9.34) par convergence dominée, en prenant pour fonction dominante la constante $2^p \sup_{\mathbb{R}} |f|^p$.

Dans le cas général, soit f un élément quelconque de $\mathrm{L}^p(\mathbb{R})$ et $\varepsilon > 0$. Par le théorème 6.26, l'espace $\mathcal{C}_c(\mathbb{R})$ des fonctions continues à support compact est dense dans $\mathrm{L}^p(\mathbb{R})$, on peut donc trouver une fonction $g \in \mathcal{C}_c(\mathbb{R})$ telle que $\|f - g_\varepsilon\|_p < \varepsilon$. En raison de l'invariance de la mesure de Lebesgue par translation, il est facile de voir que $\|f(\,.+h) - g_\varepsilon(\,.+h)\|_p = \|f - g_\varepsilon\|_p$. Par inégalité triangulaire dans $\mathrm{L}^p(\mathbb{R})$, on a :

$$\begin{aligned}\|f - f(\,.+h)\|_p &\leq \|f - g_\varepsilon\|_p + \|g_\varepsilon - g_\varepsilon(\,.+h)\|_p + \|f(\,.+h) - g_\varepsilon(\,.+h)\|_p \\ &\leq \|g_\varepsilon - g_\varepsilon(\,.+h)\|_p + 2\varepsilon.\end{aligned}$$

Par le cas particulier étudié ci-dessus, il existe $\delta > 0$ tel que pour tout h tel que $|h| < \delta$, $\|g_\varepsilon - g_\varepsilon(\,.+h)\|_p < \varepsilon$, d'où $\|f - f(\,.+h)\|_p < 3\varepsilon$. Ce raisonnement étant valable pour tout $\varepsilon > 0$, on a établi (9.34). □

Corollaire 9.21. *Si $f \in \mathrm{L}^1(\mathbb{R})$, $\widehat{f}$ est* uniformément *continue sur* $\mathbb{R}$.

Justification. C'est un exercice classique d'analyse : si f est continue sur $\mathbb{R}$ et tend vers 0 en $-\infty$ et $+\infty$, elle est uniformément continue sur $\mathbb{R}$. □

9.2.2 Formule d'inversion

Si l'on veut exploiter pleinement les bonnes propriétés de la transformée de Fourier pour résoudre des équations de convolution ou des équations différentielles, il importe de savoir l'inverser.

Théorème 9.22. *Si f et $\widehat{f}$ sont toutes deux dans $\mathrm{L}^1(\mathbb{R})$, on a la formule d'inversion,*

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(t) \mathrm{e}^{-ixt} \, \mathrm{d}\lambda_1(t) \quad \textit{pour } \lambda_1\textit{-presque tout } x \in \mathbb{R}. \tag{9.35}$$

Nous admettrons ce théorème.

9.3 Convergence en loi

9.3.1 Discussion introductive

Nous abordons maintenant l'étude de la convergence des lois de probabilité. Si μ_n et μ sont des probabilités sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$, la première notion de convergence de μ_n vers μ qui vienne à l'esprit est la convergence *ponctuelle* de la suite de *fonctions d'ensembles* μ_n définies sur $\mathrm{Bor}(\mathbb{R}^d)$ vers μ. Cette convergence est définie par :

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d), \quad \mu_n(B) \xrightarrow[n\to+\infty]{} \mu(B). \tag{9.36}$$

Il se trouve que cette notion est trop restrictive quand on veut l'appliquer à l'étude de la convergence des lois des variables aléatoires. L'exemple simple suivant (avec $d = 1$) devrait nous en convaincre.

Exemple 9.4. Considérons la suite de variables aléatoires réelles $(X_n)_{n\geq 1}$, où pour chaque $n \geq 1$, X_n suit la loi uniforme sur l'intervalle $[-1/n, 1/n]$. Il est facile de vérifier que X_n converge vers la v.a. constante 0 pour chacun des modes de convergence déjà vus (p.s., en probabilité et dans tous les espaces $\mathrm{L}^p(\Omega)$, y compris $\mathrm{L}^\infty(\Omega)$). On s'attend donc légitimement à ce que la loi μ_n de X_n converge vers la loi de la v.a. constante 0, c'est à dire vers $\mu = \delta_0$, masse de Dirac en 0. Or cette convergence ne peut avoir lieu au sens de (9.36). En effet pour $B = [0, 1]$, on a clairement $\mu_n(B) = 1/2$ pour tout $n \geq 1$, la suite $\mu_n(B)$ converge donc vers $1/2$, tandis que $\mu(B) = \delta_0([0, 1]) = 1$. On a le même problème avec $B' =]0, 1]$, $\mu_n(B') = 1/2$ et $\mu(B') = 0$. Soit maintenant un intervalle I d'extrémités a et b avec $a \neq 0$ et $b \neq 0$. Vérifions que $\mu_n(I)$ converge vers $\delta_0(I)$. Puisque μ_n est une loi uniforme en dimension 1, $\mu_n(I)$ se calcule comme un rapport de longueurs :

$$\mu_n(I) = \frac{\lambda\big([-1/n, 1/n] \cap I\big)}{\lambda\big([-1/n, 1/n]\big)} = \frac{n}{2}\lambda\big([-1/n, 1/n] \cap I\big).$$

Comme ni a ni b n'est nul, nous avons seulement deux cas à envisager, $0 \in]a, b[$ et $0 \in \mathbb{R} \setminus [a, b]$. Dans le premier cas, pour tout n supérieur à un certain n_0, on a l'inclusion $[-1/n, 1/n] \subset]a, b[\subset I$, d'où $\mu_n(I) = \frac{n}{2}\lambda\big([-1/n, 1/n]\big) = 1$. Dans ce cas, $\mu_n(I)$ converge vers 1, donc vers $\delta_0(I)$, puisque si $0 \in I$, $\delta_0(I) = 1$. Dans le deuxième cas, on a pour tout n supérieur à un certain n_1, $[-1/n, 1/n] \cap [a, b] = \emptyset$, d'où *a fortiori* $[-1/n, 1/n] \cap I = \emptyset$ et $\mu_n(I) = 0$. Alors $\mu_n(I)$ converge vers $0 = \delta_0(I)$ $(0 \notin I)$.

Il semble donc que la pathologie présentée par l'exemple 9.4 soit liée à la présence sur la frontière du borélien B de masse ponctuelle pour la mesure limite μ. L'exemple suivant permet d'affiner cette impression et de voir que le problème n'est pas tant le fait que μ ait une masse ponctuelle sur la frontière, mais plus généralement que la frontière ait une μ-mesure non nulle.

Exemple 9.5. Considérons la suite de vecteurs aléatoires (X_n, Y_n) de $\mathbb{R}^2$ de loi μ_n uniforme sur $[0, 1] \times [0, 1/n]$. Intuitivement, la loi de μ_n devrait converger vers la loi μ du vecteur aléatoire (X, Y), où X suit la loi uniforme sur $[0, 1]$ et Y est la v.a. constante zéro. Autrement dit $\mu = \mathrm{Unif}[0, 1] \otimes \delta_0$. À nouveau on constate que cette convergence ne peut avoir lieu au sens de (9.36), puisque pour $B = [0, 1/2] \times [-1, 0]$, $\mu_n(B) = 0$ pour tout n, tandis que $\mu(B) = \lambda_1([0, 1/2]) = 1/2$. Ici il n'y a pas de masse ponctuelle pour μ, mais la mesure de la frontière de B est non nulle.

Au vu des exemples 9.4 et 9.5 et avec un peu de bonne volonté, on se résout à modifier comme suit la définition de la convergence de μ_n vers μ proposée par (9.36) :

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d) \text{ tel que } \mu(\partial B) = 0, \quad \mu_n(B) \xrightarrow[n \to +\infty]{} \mu(B). \tag{9.37}$$

Nous avons noté $\partial B = \overline{B} \setminus \overset{\circ}{B}$ la frontière de B. Pour une autre illustration de la pertinence de la condition $\mu(\partial B) = 0$, voir la remarque 9.39 sur le théorème de De Moivre-Laplace.

En fait, pour des raisons de maniabilité et pour permettre un lien ultérieur avec la notion de convergence faible étudiée en analyse fonctionnelle, nous choisirons comme définition de la convergence de μ_n vers μ la condition suivante dont l'équivalence avec (9.37) sera établie ci-dessous (théorème 9.32) :

$$\forall f : \mathbb{R}^d \to \mathbb{R}, \text{ continue bornée,} \quad \int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n \xrightarrow[n \to +\infty]{} \int_{\mathbb{R}^d} f \, \mathrm{d}\mu. \tag{9.38}$$

À première vue, (9.38) semble très différente de (9.36) et (9.37). En fait la différence n'est pas si grande si l'on se rappelle que $\mu_n(B)$ peut aussi s'écrire sous la forme $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n$ avec $f = \mathbf{1}_B$. Le fait de « tester » la convergence de μ_n vers μ sur les intégrales des fonctions continues bornées plutôt que sur celles des indicatrices des boréliens permet de « lisser » un peu les choses et d'éviter les pathologies présentées ci-dessus. On pourra vérifier à titre d'exercice que la condition (9.38) est réalisée dans les exemples 9.4 et 9.5.

9.3.2 Définition et propriétés

Définition 9.23 (Convergence étroite). *Soient μ_n $(n \geq 1)$ et μ des mesures de probabilité sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$. On dit que μ_n converge* étroitement *vers μ quand n tend vers l'infini si pour toute fonction continue bornée $f : \mathbb{R}^d \to \mathbb{R}$, $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n$ converge vers $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu$ quand n tend vers l'infini.*

Définition 9.24 (Convergence en loi). *Soient X_n ($n \geq 1$) et X des vecteurs aléatoires de $\mathbb{R}^d$. On dit que X_n converge* en loi *vers X si la loi de X_n converge étroitement vers celle de X quand n tend vers l'infini. Autrement dit,*

$$X_n \xrightarrow[n\to+\infty]{\text{loi}} X \quad si \quad \forall f : \mathbb{R}^d \to \mathbb{R} \text{ continue bornée,} \quad \mathbf{E}f(X_n) \xrightarrow[n\to+\infty]{} \mathbf{E}f(X). \tag{9.39}$$

Remarque 9.25. Il n'est pas nécessaire dans la définition 9.24 de supposer les X_n définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$, puisque seule intervient leur loi. On peut très bien envisager qu'ils soient définis sur des espaces probabilisés différents : $X_n : (\Omega_n, \mathcal{F}_n, \mathbf{P}_n) \to \big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$. L'écriture de la convergence étroite de P_{X_n} vers P_X sous la forme (9.39) utilise implicitement le théorème de transfert puisque

$$\mathbf{E}f(X_n) = \int_{\Omega_n} f \circ X_n \, \mathrm{d}\mathbf{P}_n = \int_{\mathbb{R}^d} f \, \mathrm{d}P_{X_n}.$$

Remarque 9.26. Bien que traditionnelle et admise par l'ensemble des probabilistes et statisticiens, la définition 9.24 contient un *grossier abus de langage* qui est la source de bien des erreurs pour les débutants. La convergence en loi de la suite (X_n) vers le vecteur aléatoire X n'est pas une vraie convergence de suite de vecteurs aléatoires, c'est seulement la *convergence de la loi* de X_n vers la *loi* de X au sens de la définition 9.23 (qui elle, est une vraie convergence de suite de *mesures*). On peut donc dire que X_n converge aussi en loi vers *n'importe quel* vecteur aléatoire Y de *même loi* que X. Par conséquent, *il n'y a pas unicité de la limite* pour la convergence en loi. Il convient donc d'être très prudent avec la manipulation de cette convergence. Par exemple si X_n converge en loi vers X et Y_n converge en loi vers Y, on ne peut pas dire que $X_n + Y_n$ converge en loi vers $X + Y$. Pour le voir (avec $d = 1$), prenons X_n de loi symétrique, disons $\mathbf{P}(X_n = -1) = \mathbf{P}(X_n = +1) = 1/2$. Posons pour tout n, $Y_n := -X_n$. Il est clair que Y_n et X_n ont même loi et comme cette loi ne dépend pas de n, il y a bien convergence en loi de la suite (X_n) vers une variable aléatoire X de même loi que X_n. Comme Y_n a même loi que X_n pour tout n, il est correct de dire que Y_n converge elle aussi en loi vers X. On n'a pourtant pas convergence de $X_n + Y_n$ vers $X + X$, puisque pour tout n, $X_n + Y_n = 0$ donc $X_n + Y_n$ a pour loi δ_0 ne dépendant pas de n. Ainsi $X_n + Y_n$ converge en loi vers 0. La loi de $X + X$ n'est évidemment pas δ_0, c'est $\frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_2$. Bien sûr, on peut *aussi* dire que Y_n converge en loi vers $Y := -X$ et avec ce choix particulier de Y (mais encore une fois une infinité d'autres choix sont possibles), $X_n + Y_n$ converge en loi vers $X + Y$.

Remarque 9.27. Maintenant que l'on a bien compris que dans la convergence en loi de X_n vers un vecteur aléatoire X de loi μ, la limite X n'est pas unique, on s'autorisera un abus de langage supplémentaire nous permettant d'écrire quand on ne souhaite pas choisir X parmi tous les vecteurs aléatoires ayant la loi μ : $X_n \xrightarrow{\text{loi}} \mu$. Nous userons de cette facilité dans l'énoncé du théorème limite central en écrivant

$$S_n^* \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0,1).$$

Voici un premier exemple de convergence étroite, obtenu par utilisation directe de la définition 9.23.

Exemple 9.6 (Convergence de lois uniformes discrètes). Pour $n \geq 2$, soit μ_n la loi uniforme discrète sur l'ensemble fini $E_n := \{1/n, 2/n, \dots, n/n\}$. Autrement dit :

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{k/n}.$$

Quand n tend vers l'infini, μ_n converge étroitement vers μ loi uniforme sur $[0,1]$, i.e. μ est la mesure de densité $\mathbf{1}_{[0,1]}$ par rapport à λ_1. En effet, soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. On a

$$\int_{\mathbb{R}} f \,\mathrm{d}\mu_n = \frac{1}{n}\sum_{k=1}^{n}\int_{\mathbb{R}} f \,\mathrm{d}\delta_{k/n} = \frac{1}{n}\sum_{k=1}^{n} f\Big(\frac{k}{n}\Big).$$

On reconnaît là une *somme de Riemann* associée à f et au partage de $[0,1]$ constitué par E_n. Il est bien connu que cette somme converge (puisque f est continue) vers l'intégrale de Riemann de f entre 0 et 1 :

$$\int_{\mathbb{R}} f \,\mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int_0^1 f(x)\,\mathrm{d}x = \int_{[0,1]} f \,\mathrm{d}\lambda_1 = \int_{\mathbb{R}} f\mathbf{1}_{[0,1]} \,\mathrm{d}\lambda_1 = \int_{\mathbb{R}} f \,\mathrm{d}\mu.$$

Comme la fonction continue bornée f était quelconque, ceci établit la convergence étroite de μ_n vers μ. Sur le même sujet, voir aussi [ICP] section 8.1.

Le théorème suivant est un outil très puissant pour obtenir des convergences en loi.

Théorème 9.28 (de l'image continue). *Soient X_n ($n \geq 1$) et X des vecteurs aléatoires de $\mathbb{R}^d$ et $g : \mathbb{R}^d \to \mathbb{R}^k$ une application continue. Si X_n converge en loi vers X, alors la suite de vecteurs aléatoires $g(X_n)$ dans $\mathbb{R}^k$ converge en loi vers le vecteur aléatoire $g(X)$.*

Preuve. Par hypothèse, $\mathbf{E}f(X_n)$ tend vers $\mathbf{E}f(X)$ pour toute $f : \mathbb{R}^d \to \mathbb{R}$ continue bornée. Posons $Y_n := g(X_n)$ et $Y := g(X)$. Il s'agit de montrer que $\mathbf{E}h(Y_n)$ tend vers $\mathbf{E}h(Y)$ pour toute $h : \mathbb{R}^k \to \mathbb{R}$, continue bornée. Or $\mathbf{E}h(Y_n) = \mathbf{E}\big((h\circ g)(X_n)\big)$ et $\mathbf{E}h(Y) = \mathbf{E}\big((h\circ g)(X)\big)$, il suffit alors pour conclure de remarquer que $f := h\circ g$ est une fonction continue bornée de $\mathbb{R}^d \to \mathbb{R}$ et d'appliquer l'hypothèse. Il importe de remarquer que nous n'avons pas supposé g bornée. □

Voici une application simple mais bien utile du théorème de l'image continue.

Corollaire 9.29. *Si le vecteur aléatoire (X_n, Y_n) de $\mathbb{R}^2$ converge en loi vers le vecteur aléatoire (X, Y), alors*

a) $X_n \xrightarrow[n\to+\infty]{\text{loi}} X$ *et* $Y_n \xrightarrow[n\to+\infty]{\text{loi}} Y$.

b) $\forall (a,b) \in \mathbb{R}^2$, $aX_n + bY_n \xrightarrow[n\to+\infty]{\text{loi}} aX + bY$, *c)* $X_nY_n \xrightarrow[n\to+\infty]{\text{loi}} XY$.

Vérification. Il suffit de remarquer que d'après le théorème 9.28, si $g : \mathbb{R}^2 \to \mathbb{R}$ est continue, $g(X_n, Y_n)$ converge en loi vers $g(X, Y)$. Voici le choix explicite de g dans chaque cas :

a) $g(x, y) = x$, puis $g(x, y) = y$.

b) $g(x, y) = ax + by$.

c) $g(x, y) = xy$.

□

Remarque 9.30. Il convient de noter que la réciproque du a) est fausse, car sinon on pourrait grâce au b) en déduire que les convergences en loi respectives de X_n vers X et de Y_n vers Y impliquent celle de $X_n + Y_n$ vers $X + Y$. On sait que cette implication est fausse cf. remarque 9.25.

La remarque suivante sur le même sujet est à sauter en première lecture.

Remarque 9.31. Par ailleurs, la convergence en loi de X_n vers X et de Y_n vers Y n'interdit pas que pour chaque n, X_n et Y_n soient définis sur des espaces probabilisés $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ et $(\Omega'_n, \mathcal{F}'_n, \mathbf{P}'_n)$ différents[5]. La définition de X_n+Y_n comme application mesurable n'est alors même pas claire. Une possibilité serait de la définir comme application de $(\Omega_n \times \Omega'_n, \mathcal{F}_n \otimes \mathcal{F}'_n)$ dans $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ en posant $(X_n + Y_n)(\omega, \omega') := X_n(\omega) + Y_n(\omega')$. Ceci laisse entière la question du choix de la probabilité sur $(\Omega_n \times \Omega'_n, \mathcal{F}_n \otimes \mathcal{F}'_n)$. Par exemple, munir cet espace mesurable de $\mathbf{P}_n \otimes \mathbf{P}'_n$ reviendrait à supposer que X_n et Y_n sont indépendantes, ce qui n'est pas dans l'hypothèse...Ceci explique pourquoi dans l'étude de la convergence en loi de sommes $S_n = X_{n,1} + \cdots + X_{n,k_n}$ de variables aléatoires, on suppose en général que pour chaque n, les termes $X_{n,j}$ de S_n sont des variables aléatoires définies sur le même espace espace probabilisé $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$.

Le théorème suivant rassemble les diverses caractérisations de la convergence étroite, à l'exception de celle relative à la convergence des transformées de Fourier qui fait l'objet de la sous-section 9.3.3. Dans la littérature anglaise, il est connu sous le sobriquet de *portmanteau theorem*, qu'il serait erroné de traduire par « théorème du portemanteau », puisque le faux-ami *portmanteau* désigne en fait une grande valise à deux compartiments[6]. Le premier compartiment *i)*–*iii)* contient les caractérisations fonctionnelles (i.e. en termes de convergences d'intégrales de fonctions) le deuxième *iv)*–*vi)* contient les caractérisations ensemblistes (convergences de mesures de boréliens).

5. Cette difficulté ne se présente pas quand on parle du *vecteur aléatoire* (X_n, Y_n) qui est par définition, une application mesurable d'un espace $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ dans $(\mathbb{R}^2, \mathrm{Bor}(\mathbb{R}^2))$. Dans ce cas les composantes X_n et Y_n sont des applications mesurables du même espace probabilisé $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ dans $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$.

6. D'après le Collins Cobuild : **1** *A portmanteau is a large travelling case which opens out into two equal compartments ; an old fashioned use.* **2** *Portmanteau is used to describe* **2.1** *a word that combines parts of the forms and meanings of two other words. For example, 'brunch' is formed from 'breakfast' and 'lunch'...* **2.2** *someone or something that combines many different features or uses.*

Théorème 9.32. *Soient μ_n ($n \geq 1$) et μ des mesures de probabilité sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$. Les fonctions f ci-dessous sont toutes définies sur $\mathbb{R}^d$ et à valeurs dans $\mathbb{R}$. Les assertions suivantes sont équivalentes.*

i) Pour toute f continue bornée, $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu$.

ii) Pour toute f uniformément *continue sur $\mathbb{R}^d$ et bornée, $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu$.*

iii) Pour toute f continue à support compact, $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu$.

iv) Pour tout fermé C de $\mathbb{R}^d$, $\limsup\limits_{n\to+\infty} \mu_n(C) \leq \mu(C)$.

v) Pour tout ouvert G de $\mathbb{R}^d$, $\liminf\limits_{n\to+\infty} \mu_n(G) \geq \mu(G)$.

vi) Pour tout borélien B tel que $\mu(\partial B) = 0$, $\lim\limits_{n\to+\infty} \mu_n(B) = \mu(B)$, où l'on a noté $\partial B := \overline{B} \setminus \overset{\circ}{B}$ la frontière de B.

Si $d = 1$, en notant F_n et F les fonctions de répartition respectives de μ_n et μ, les conditions i)–vi) sont équivalentes à :

vii) Pour tout point de continuité x de F, $\lim\limits_{n\to+\infty} F_n(x) = F(x)$.

Démonstration. La preuve de l'équivalence des conditions *i*) à *vi*) s'articule selon le schéma :

$$i) \Rightarrow ii) \Rightarrow iii) \Rightarrow i); \quad i) \Rightarrow iv); \quad iv) \Leftrightarrow v); \quad iv) \text{ et } v) \Rightarrow vi); \quad vi) \Rightarrow i).$$

Les implications *i*) $\Rightarrow$ *ii*) $\Rightarrow$ *iii*) sont évidentes par emboîtement des ensembles de fonctions f concernées.

Preuve de iii) $\Rightarrow$ *i*). La propriété cruciale permettant de remonter de *iii*) à *i*) est le fait que pour une probabilité μ sur $\mathbb{R}^d$, toute la masse est portée à ε près par un *compact*. On dit que μ est *tendue*. Plus précisément, nous allons utiliser la version fonctionnelle suivante de cette propriété :

$$\forall \varepsilon > 0, \exists h_\varepsilon : \mathbb{R}^d \to [0,1], \text{ continue à support compact}, \quad 0 \leq 1 - \int_{\mathbb{R}^d} h_\varepsilon \,\mathrm{d}\mu < \varepsilon. \quad (9.40)$$

En effet, la suite croissante (pour l'inclusion) des compacts $B(k) := [-k,k]^d$, $k \in \mathbb{N}^*$ a pour réunion $\mathbb{R}^d$, donc par continuité séquentielle croissante de μ, $\mu\big(B(k)\big) \uparrow \mu(\mathbb{R}^d) = 1$. On peut donc trouver un entier k_ε tel que $\mu\big(B(k_\varepsilon)\big) > 1 - \varepsilon$. On définit alors

$$h_\varepsilon = h_{\varepsilon,1} \otimes \cdots \otimes h_{\varepsilon,d},$$

où $h_{\varepsilon,j} : \mathbb{R} \to [0,1]$ est la fonction continue valant 0 hors de $[-k_\varepsilon - 1, k_\varepsilon + 1]$, valant 1 sur $[-k_\varepsilon, k_\varepsilon]$ et est affine sur chacun des intervalles $[-k_\varepsilon - 1, -k_\varepsilon]$ et $[k_\varepsilon, k_\varepsilon + 1]$. Cette fonction h_ε est bien continue à support compact (le support est ici l'hypercube fermé

$[-k_\varepsilon - 1, k_\varepsilon + 1]^d$). De plus h_ε vérifie l'encadrement $1 \geq h_\varepsilon \geq \mathbf{1}_{B(k_\varepsilon)}$, qui donne après intégration sur $\mathbb{R}^d$ relativement à μ :

$$1 \geq \int_{\mathbb{R}^d} h_\varepsilon \, \mathrm{d}\mu \geq \mu\big(B(k_\varepsilon)\big) > 1 - \varepsilon.$$

Ainsi (9.40) est vérifiée. Notons que la fonction h_ε dépend de ε et de μ.

Soit f continue bornée $\mathbb{R}^d \to \mathbb{R}$ et fixons $\varepsilon > 0$. Par inégalité triangulaire, on a la majoration :

$$\left| \int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \right| \leq T_1 + T_2 + T_3, \tag{9.41}$$

où

$$T_1 := \left| \int_{\mathbb{R}^d} f(1 - h_\varepsilon) \, \mathrm{d}\mu_n \right| \quad T_2 := \left| \int_{\mathbb{R}^d} f h_\varepsilon \, \mathrm{d}\mu_n - \int_{\mathbb{R}^d} f h_\varepsilon \, \mathrm{d}\mu \right| \quad T_3 := \left| \int_{\mathbb{R}^d} f(1 - h_\varepsilon) \, \mathrm{d}\mu \right|.$$

D'après (9.40) on a immédiatement (noter la positivité de $1 - h_\varepsilon$) :

$$T_3 \leq \int_{\mathbb{R}^d} |f|(1 - h_\varepsilon) \, \mathrm{d}\mu \leq \|f\|_\infty \int_{\mathbb{R}^d} (1 - h_\varepsilon) \, \mathrm{d}\mu = \|f\|_\infty \left\{ 1 - \int_{\mathbb{R}^d} h_\varepsilon \, \mathrm{d}\mu \right\} < \|f\|_\infty \, \varepsilon. \tag{9.42}$$

De même avec μ_n au lieu de μ,

$$\forall n \in \mathbb{N}^*, \quad T_1 \leq \|f\|_\infty \left\{ 1 - \int_{\mathbb{R}^d} h_\varepsilon \, \mathrm{d}\mu_n \right\}. \tag{9.43}$$

La fonction h_ε étant continue à support compact, l'hypothèse iii) nous donne la convergence de $\int_{\mathbb{R}^d} h_\varepsilon \, \mathrm{d}\mu_n$ vers $\int_{\mathbb{R}^d} h_\varepsilon \, \mathrm{d}\mu$, qui avec (9.42) et (9.43) nous assure de l'existence d'un entier $n_1 = n_1(\varepsilon)$ tel que

$$\forall n \geq n_1, \quad T_1 \leq 2\|f\|_\infty \, \varepsilon. \tag{9.44}$$

Enfin, la fonction $f h_\varepsilon$ est continue (comme produit de deux fonctions continues) et à support compact (puisque nulle en dehors du support compact de h). L'hypothèse *iii)* appliquée à cette fonction nous dit que T_2 tend vers 0 quand n tend vers l'infini, il existe donc un $n_2 = n_2(\varepsilon)$ tel que :

$$\forall n \geq n_2, \quad T_2 \leq \varepsilon. \tag{9.45}$$

En rassemblant (9.41), (9.42), (9.44) et (9.45), on voit que

$$\forall n \geq n_0 := \max(n_1, n_2), \quad \left| \int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \right| < \big(3\|f\|_\infty + 1\big)\varepsilon.$$

Comme ε est arbitraire, on a ainsi prouvé la convergence de $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n$ vers $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu$. Ce résultat étant valable pour toute f continue bornée, l'implication *iii)* $\Rightarrow$ *i)* est établie. □

Preuve de $i) \Rightarrow iv)$. Fixons un fermé quelconque C de $\mathbb{R}^d$ et définissons la fonction continue $\operatorname{dist}(\,.\,, C)$ par :

$$\forall x \in \mathbb{R}^d, \quad \operatorname{dist}(x, C) := \inf\bigl\{\|x - y\|;\ y \in C\bigr\}.$$

Rappelons que puisque C est *fermé*,

$$\operatorname{dist}(x, C) = 0 \Leftrightarrow x \in C. \tag{9.46}$$

Définissons pour tout $k \in \mathbb{N}^*$, la fonction $h_k : \mathbb{R}_+ \to [0, 1]$ par

$$h_k(t) := \begin{cases} 1 - kt & \text{si } 0 \leq t \leq 1/k, \\ 0 & \text{si } t > 1/k. \end{cases}$$

Cette fonction est continue bornée et de plus

$$\lim_{k\to+\infty} h_k(t) = \begin{cases} 1 & \text{si } t = 0, \\ 0 & \text{si } t > 0. \end{cases} \tag{9.47}$$

En posant

$$\forall x \in \mathbb{R}^d, \quad f_k(x) := h_k\bigl(\operatorname{dist}(x, C)\bigr),$$

on construit une fonction continue bornée f_k telle que $\mathbf{1}_C \leq f_k \leq 1$ et compte-tenu de (9.46) et (9.47), f_k converge ponctuellement sur $\mathbb{R}^d$ vers $\mathbf{1}_C$. Cette convergence étant dominée par la constante 1 qui est μ intégrable (μ est une mesure finie),

$$\lim_{k\to+\infty} \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu = \int_{\mathbb{R}^d} \mathbf{1}_C \,\mathrm{d}\mu = \mu(C). \tag{9.48}$$

D'autre part pour chaque $k \in \mathbb{N}^*$, l'inégalité $\mathbf{1}_C \leq f_k$ nous donne :

$$\forall n \in \mathbb{N}^*, \quad \mu_n(C) \leq \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu_n. \tag{9.49}$$

Si l'on fait tendre n vers l'infini (k restant fixé), l'intégrale au second membre de (9.49) converge vers $\int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu$ d'après l'hypothèse $i)$. Par contre rien ne nous dit que la suite $\mu_n(C)$ converge. Pour remédier à cet inconvénient, on prend les limites supérieures[7]

$$\limsup_{n\to+\infty} \mu_n(C) \leq \limsup_{n\to+\infty} \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu_n. \tag{9.50}$$

Grâce à l'hypothèse $i)$ le second membre de (9.50) s'écrit aussi $\lim_{n\to+\infty} \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu_n = \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu$, d'où

$$\forall k \in \mathbb{N}^*, \quad \limsup_{n\to+\infty} \mu_n(C) \leq \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu.$$

7. On pourrait aussi prendre les limites inférieures et dérouler sans inconvénient la fin de la preuve. On aboutirait simplement à l'inégalité $\liminf_n \mu_n(C) \leq \mu(C)$ qui est plus faible que $iv)$.

Dans cette inégalité le premier membre est une constante, de sorte qu'en faisant tendre k vers l'infini, on obtient grâce à (9.48),

$$\limsup_{n\to+\infty} \mu_n(C) \leq \mu(C).$$

Comme le fermé C était quelconque, l'implication $i) \Rightarrow iv)$ est établie. □

Preuve de iv) $\Leftrightarrow$ *v).* La preuve de cette équivalence est immédiate en remarquant que si G est un ouvert, son complémentaire $C := \mathbb{R}^d \setminus G$ est un fermé (et vice versa), que $\mu_n(G) = 1 - \mu_n(C)$ et en utilisant les égalités[8]

$$\limsup_{n\to+\infty}(1-u_n) = 1 - \liminf_{n\to+\infty} u_n, \quad \liminf_{n\to+\infty}(1-v_n) = 1 - \limsup_{n\to+\infty} v_n, \tag{9.51}$$

avec $u_n = \mu_n(G)$ et $v_n = \mu_n(C)$. □

Preuve de $\big(iv)$ *et* $v)\big) \Rightarrow vi)$. Soit B un borélien de $\mathbb{R}^d$ tel que $\mu(\partial B) = 0$. Comme $\overline{B}$ est l'union disjointe de son intérieur $\overset{\circ}{B}$ et de sa frontière ∂B, on a $\mu(\overline{B}) = \mu(\overset{\circ}{B}) + \mu(\partial B) = \mu(\overset{\circ}{B})$. En raison de l'inclusion $\overset{\circ}{B} \subset B \subset \overline{B}$, on en déduit :

$$\mu(\overset{\circ}{B}) = \mu(B) = \mu(\overline{B}). \tag{9.52}$$

L'inclusion $B \subset \overline{B}$ et l'hypothèse $iv)$ nous donnent

$$\limsup_{n\to+\infty} \mu_n(B) \leq \limsup_{n\to+\infty} \mu_n(\overline{B}) \leq \mu(\overline{B}), \tag{9.53}$$

tandis que $v)$ et l'inclusion $\overset{\circ}{B} \subset B$ nous donnent

$$\mu(\overset{\circ}{B}) \leq \liminf_{n\to+\infty} \mu_n(\overset{\circ}{B}) \leq \liminf_{n\to+\infty} \mu_n(B). \tag{9.54}$$

En combinant (9.52), (9.53) et (9.54), on en déduit

$$\mu(B) \leq \liminf_{n\to+\infty} \mu_n(B) \leq \limsup_{n\to+\infty} \mu_n(B) \leq \mu(B),$$

d'où

$$\liminf_{n\to+\infty} \mu_n(B) = \limsup_{n\to+\infty} \mu_n(B) = \mu(B),$$

ce qui signifie que $\mu_n(B)$ converge vers $\mu(B)$. □

8. Si g est continue et décroissante $\overline{\mathbb{R}} \to \overline{\mathbb{R}}$ et $E \subset \overline{\mathbb{R}}$, on a $\sup g(E) = g(\inf E)$ et $\inf g(E) = g(\sup E)$ (exercice). En appliquant ceci avec $g(x) = 1 - x$ et en revenant aux définitions de $\liminf$ et $\limsup$, on obtient (9.51).

Preuve de vi) $\Rightarrow$ i). Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction continue bornée. Nous allons construire pour chaque $\varepsilon > 0$, une fonction étagée $g = \sum_{k=1}^{l} c_k \mathbf{1}_{B_k}$ telle que *pour toute probabilité ν sur* $\bigl(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\bigr)$, $I(\nu) := \bigl|\int_{\mathbb{R}^d}(g-f)\,\mathrm{d}\nu\bigr| < 2\varepsilon$. Cette majoration uniforme relativement à la mesure permet de ramener le contrôle de la différence $|\int_{\mathbb{R}^d} f\,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f\,\mathrm{d}\mu|$ à celui de $|\int_{\mathbb{R}^d} g\,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} g\,\mathrm{d}\mu|$. Cette différence d'intégrales s'écrit comme une combinaison linéaire finie des $\mu_n(B_k) - \mu(B_k)$ et on la majore par ε pour n assez grand grâce à l'hypothèse *vi*), pour peu que l'on ait réussi à choisir les B_k tels que $\mu(\partial B_k) = 0$.

Pour réaliser ce programme, commençons par fixer $\varepsilon > 0$. Comme f est bornée, $f(\mathbb{R}^d)$ est une partie bornée de $\mathbb{R}$, on peut donc définir $a = a(\varepsilon)$ et $b = b(\varepsilon)$ par :

$$a := \max\bigl\{k\varepsilon;\ k \in \mathbb{Z},\ (k+1)\varepsilon \leq \inf f\bigr\}, \quad b := \min\bigl\{k\varepsilon;\ k \in \mathbb{Z},\ (k-1)\varepsilon \geq \sup f\bigr\}.$$

Ainsi $[a, b]$ déborde de chaque côté de l'intervalle $[\inf f, \sup f]$ d'au moins ε. L'application continue f étant borélienne, la mesure image $\mu \circ f^{-1}$ est bien définie. C'est une mesure finie (et même une probabilité) sur $\bigl(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\bigr)$. Notons F sa fonction de répartition. On sait que F est discontinue au point $y \in \mathbb{R}$ si et seulement si $(\mu \circ f^{-1})(\{y\}) \neq 0$ et que l'ensemble D de ces points de discontinuité de F est au plus dénombrable. Ainsi pour tout intervalle non vide J, $D \cap J$ est au plus dénombrable alors que J est infini non dénombrable. Par conséquent J contient au moins un point de D^c, autrement dit un point c tel que $(\mu \circ f^{-1})(\{c\}) = 0$. En appliquant cet argument à chacun des intervalles $[k\varepsilon, (k+1)\varepsilon[$ pour $a \leq k\varepsilon < b$, on obtient une suite finie $(c_k)_{0 \leq k \leq l}$ vérifiant :

$$\forall k = 0, \dots, l, \quad (\mu \circ f^{-1})(\{c_k\}) = 0. \tag{9.55}$$

La construction de cette suite nous montre de plus que $c_0 < \inf f$, $c_l > \sup f$ et que puisque $c_k \in [k\varepsilon, (k+1)\varepsilon[$,

$$\forall k = 1, \dots, l, \quad 0 < c_k - c_{k-1} < 2\varepsilon. \tag{9.56}$$

Posons maintenant

$$g := \sum_{k=1}^{l} c_k \mathbf{1}_{B_k}, \quad \text{où} \quad B_k := f^{-1}\bigl(]c_{k-1}, c_k]\bigr).$$

Par construction les B_k sont deux à deux disjoints et comme $c_0 < \inf f \leq \sup f < c_l$, $]c_0, c_l]$ contient $f(\mathbb{R}^d)$ et par conséquent l'union des B_k est $\mathbb{R}^d$. On en déduit que pour toute mesure de probabilité ν sur $\mathbb{R}^d$,

$$I(\nu) := \left|\int_{\mathbb{R}^d} (g-f)\,\mathrm{d}\nu\right| = \sum_{k=1}^{l} \int_{B_k} (c_k - f)\,\mathrm{d}\nu \leq 2\varepsilon \sum_{k=1}^{l} \nu(B_k) = 2\varepsilon\nu(\mathbb{R}^d) = 2\varepsilon, \tag{9.57}$$

en notant que si $x \in B_k$, $0 \leq c_k - f(x) \leq c_k - c_{k-1} < 2\varepsilon$, d'après (9.56).

L'inégalité triangulaire et (9.57) appliquée à $\nu = \mu_n$ puis à $\nu = \mu$, donnent

$$\begin{aligned}\left|\int_{\mathbb{R}^d} f\,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f\,\mathrm{d}\mu\right| &\leq I(\mu_n) + \left|\int_{\mathbb{R}^d} g\,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} g\,\mathrm{d}\mu\right| + I(\mu) \\ &\leq 4\varepsilon + \sum_{k=1}^{l} |c_k||\mu_n(B_k) - \mu(B_k)|. \end{aligned} \tag{9.58}$$

Vérifions que $\mu(\partial B_k) = 0$ pour chaque k. Par continuité de f,

– $G_k := f^{-1}\big(]c_{k-1}, c_k[\big)$ est un *ouvert*, inclus dans B_k, donc $G_k \subset \overset{\circ}{B_k}$;
– $C_k = f^{-1}\big([c_{k-1}, c_k]\big)$ est un *fermé*, contenant B_k, donc $\overline{B_k} \subset C_k$.

Alors

$$\partial B_k = \overline{B_k} \setminus \overset{\circ}{B_k} \subset C_k \setminus G_k = \{x \in \mathbb{R}^d;\ f(x) = c_{k-1} \text{ ou } f(x) = c_k\},$$

d'où grâce à (9.55),

$$\mu(\partial B_k) = (\mu \circ f^{-1})(\{c_{k-1}\}) + (\mu \circ f^{-1})(\{c_k\}) = 0.$$

On voit maintenant en appliquant l'hypothèse *vi)* que le sigma dans (9.58) est la somme d'un nombre fini (l dépendant de ε) de termes tendant chacun vers 0 quand n tend vers $+\infty$. On peut ainsi trouver un $n_0 = n_0(\varepsilon)$ tel que ce sigma soit majoré par ε à partir du rang n_0. Ainsi

$$\exists n_0 = n_0(\varepsilon),\ \forall n \geq n_0, \qquad \left| \int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \,\mathrm{d}\mu \right| < 5\varepsilon.$$

Ceci étant vrai pour tout $\varepsilon > 0$, $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n$ converge vers $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu$. Comme la fonction continue bornée f était quelconque, *i)* est vérifiée. □

Pour achever la preuve du théorème, il ne nous reste plus qu'à établir la caractérisation de la convergence étroite par les fonctions de répartition dans le cas $d = 1$. Exploitant l'équivalence déjà établie des conditions *i)* à *vi)*, il nous suffit pour cela de vérifier que :

$$vi) \Rightarrow vii) \Rightarrow iii).$$

Preuve de vi) ⇒ vii). En notant F_n et F les fonctions de répartition respectives de μ_n et μ, on a :

$$\forall x \in \mathbb{R}, \quad F_n(x) = \mu_n(]-\infty, x]), \quad F(x) = \mu(]-\infty, x]).$$

Supposons maintenant que x est point de continuité de F. On sait qu'alors $\mu(\{x\}) = 0$. Or $\{x\}$ est exactement la frontière du borélien $B :=]-\infty, x]$, donc par *vi)*, $\mu_n(B)$ converge vers $\mu(B)$, autrement dit $F_n(x)$ converge vers $F(x)$. □

Preuve de vii) ⇒ iii). Soit $f : \mathbb{R} \to \mathbb{R}$, continue à support compact. Alors f est nulle en dehors d'un intervalle compact $[a, b]$ de $\mathbb{R}$ et sur cet intervalle, elle est *uniformément* continue :

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall x, y \in [a, b], \quad |x - y| \leq \delta \Rightarrow |f(x) - f(y)| < \varepsilon. \tag{9.59}$$

Fixons $\varepsilon > 0$ et un δ correspondant donné par (9.59). Découpons $[a, b]$ en intervalles consécutifs de longueur $\delta/2$ en partant de a (le dernier aura une longueur au plus $\delta/2$). L'ensemble des points de discontinuité de F étant au plus dénombrable, on peut trouver dans chacun de ces intervalles un point de continuité de F. On obtient ainsi une suite finie $x_1, \dots, x_l$ de points de continuité de F. Par construction, la distance entre x_{k-1} et x_k est au plus δ, d'où grâce à (9.59),

$$\forall k = 1, \dots, l,\ \forall x \in [x_{k-1}, x_k], \quad |f(x) - f(x_k)| < \varepsilon. \tag{9.60}$$

Considérons la fonction étagée

$$g := \sum_{k=1}^{l} f(x_k) \mathbf{1}_{]x_{k-1}, x_k]},$$

qui comme f, est nulle hors de $[a, b]$. Pour toute mesure de probabilité ν sur $\mathbb{R}$, on a

$$\begin{aligned} I(\nu) &:= \left| \int_{\mathbb{R}} (g - f) \,\mathrm{d}\nu \right| & (9.61) \\ &\leq \int_{[a,x_0]} |f| \,\mathrm{d}\nu + \sum_{k=1}^{l} \int_{]x_{k-1},x_k]} \big|f(x) - f(x_k)\big| \,\mathrm{d}\nu(x) + \int_{]x_l,b]} |f| \,\mathrm{d}\nu. & (9.62) \end{aligned}$$

Comme $f(a) = 0$ et $|x_0 - a| \leq \delta$, $|f(x)| = |f(x) - f(a)|$ est majoré par ε sur $[a, x_0]$, d'où

$$\int_{[a,x_0]} |f| \,\mathrm{d}\nu \leq \varepsilon \nu([a, x_0]).$$

Par le même argument avec b à la place de a,

$$\int_{]x_l,b]} |f| \,\mathrm{d}\nu \leq \varepsilon \nu(]x_l, b]).$$

De plus par (9.60),

$$\forall k = 1, \dots, l, \quad \int_{]x_{k-1},x_k]} \big|f(x) - f(x_k)\big| \,\mathrm{d}\nu(x) \leq \varepsilon \nu(]x_{k-1}, x_k]).$$

En reportant ces majorations dans (9.62), il vient :

$$I(\nu) \leq \varepsilon \nu([a, x_0]) + \varepsilon \sum_{k=1}^{l} \nu(]x_{k-1}, x_k]) + \varepsilon \nu(]x_l, b]) = \varepsilon \nu([a, b]) \leq \varepsilon. \tag{9.63}$$

Par inégalité triangulaire et (9.63) appliquée avec $\nu = \mu_n$, puis $\nu = \mu$, on obtient

$$\begin{aligned} \left| \int_{\mathbb{R}} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}} f \,\mathrm{d}\mu \right| &\leq I(\mu_n) + \left| \int_{\mathbb{R}} g \,\mathrm{d}\mu_n - \int_{\mathbb{R}} g \,\mathrm{d}\mu \right| + I(\mu) \\ &\leq 2\varepsilon + \sum_{k=1}^{l} |f(x_k)| \big|\mu_n(]x_{k-1}, x_k]) - \mu(]x_{k-1}, x_k])\big|. & (9.64) \end{aligned}$$

Remarquons maintenant que

$$\begin{aligned}\mu_n(]x_{k-1},x_k]) - \mu(]x_{k-1},x_k]) &= F_n(x_k) - F_n(x_{k-1}) - F(x_k) + F(x_{k-1}) \\ &= \big(F_n(x_k) - F(x_k)\big) + \big(F(x_{k-1}) - F_n(x_{k-1})\big)\end{aligned}$$

tend vers 0 quand n tend vers $+\infty$, en raison de l'hypothèse *vii*) puisque les x_k sont des points de continuité de F. En notant que pour l'instant, ε est fixé et les x_k et l dépendent de ε et de F, mais pas de n, on en déduit l'existence d'un $n_0 = n_0(\varepsilon, F, f)$ tel que pour tout $n \geq n_0$, le sigma dans (9.64) est strictement inférieur à ε. Ainsi

$$\forall n \geq n_0 = n_0(\varepsilon, F, f), \qquad \left| \int_{\mathbb{R}} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}} f \, \mathrm{d}\mu \right| < 3\varepsilon.$$

Comme ε était arbitraire, ceci établit la convergence de $\int_{\mathbb{R}} f \, \mathrm{d}\mu_n$ vers $\int_{\mathbb{R}} f \, \mathrm{d}\mu$. La fonction continue à support compact f étant quelconque, on obtient finalement *iii*). □

Le théorème 9.32 est maintenant complètement démontré. □

Voici un exemple où la caractérisation de la convergence étroite par la convergence des f.d.r. au sens de *vii*) est un outil bien adapté.

Exemple 9.7 (Une loi limite d'extrêmes). Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires indépendantes et de même loi avec fonction de répartition commune F. Définissons la suite de variables aléatoires (M_n) par :

$$M_n := \max_{1\leq k\leq n} X_k \qquad n \in \mathbb{N}^*. \tag{9.65}$$

Connaissant F, il est facile d'obtenir la fonction de répartition G_n de M_n :

$$G_n(x) = \mathbf{P}(M_n \leq x) = \mathbf{P}\big(\forall k \in \{1, \dots, n\},\, X_k \leq x\big) = \mathbf{P}\Big(\bigcap_{k=1}^{n} \{X_k \leq x\}\Big).$$

En utilisant l'indépendance des X_k, puis le fait qu'elles ont même loi, on en déduit :

$$G_n(x) = \prod_{k=1}^{n} \mathbf{P}(X_k \leq x) = \big(F(x)\big)^n. \tag{9.66}$$

Supposons désormais que les X_k ont pour loi commune la loi exponentielle de paramètre α. On a alors :

$$F(x) = 1 - \mathrm{e}^{-\alpha x} \text{ si } x \geq 0, \quad F(x) = 0 \text{ si } x < 0.$$

$$G_n(x) = \big(1 - \mathrm{e}^{-\alpha x}\big)^n \text{ si } x \geq 0, \quad G_n(x) = 0 \text{ si } x < 0.$$

Donc pour x réel fixé, on a $\lim_{n\to+\infty} G_n(x) = 0$. La signification intuitive de ce résultat est que M_n aura tendance à être grand pour les grandes valeurs de n. Afin de préciser cette idée, on essaie de normaliser M_n pour contrôler sa croissance aléatoire vers $+\infty$. Pour cela, on examine le comportement asymptotique de $\mathbf{P}(M_n - \alpha^{-1}\ln n \leq x)$:

$$\mathbf{P}\left(M_n - \frac{\ln n}{\alpha} \leq x\right) = G_n\left(x + \frac{\ln n}{\alpha}\right) = \big(1 - \mathrm{e}^{-\alpha x - \ln n}\big)^n = \left(1 - \frac{\mathrm{e}^{-\alpha x}}{n}\right)^n. \tag{9.67}$$

On en déduit que :

$$\lim_{n\to+\infty} \mathbf{P}\left(M_n - \frac{\ln n}{\alpha} \leq x\right) = \exp\bigl(-\mathrm{e}^{-\alpha x}\bigr) \tag{9.68}$$

Le calcul (9.67) est valable pour $\ln n \geq -\alpha x$, donc pour tout $n \in \mathbb{N}^*$ et tout $x \geq 0$. Pour $x < 0$ fixé, on aura $\ln n \geq -\alpha x$ pour $n \geq n_0(x)$ donc (9.68) est valable pour tout x réel. On peut donc dire qu'*asymptotiquement*, M_n est de l'ordre de grandeur de $\alpha^{-1} \ln n$ et que la dispersion aléatoire de M_n autour de cette valeur est donnée par la loi de fonction de répartition :

$$H(x) = \exp\bigl(-\mathrm{e}^{-\alpha x}\bigr) \qquad x \in \mathbb{R}. \tag{9.69}$$

Il s'agit d'une loi de Gumbel. D'après le théorème 9.32 *vii*) et (9.68), on peut reformuler la conclusion en disant que la suite de variables aléatoires $M_n - \alpha^{-1} \ln n$ converge en loi vers une v.a. suivant la loi de Gumbel définie par sa f.d.r. (9.69).

Théorème 9.33. *La convergence en probabilité implique la convergence en loi : si X_n ($n \geq 1$) et X sont des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ telles que X_n converge en probabilité vers X, alors X_n converge aussi en loi vers X.*

Preuve. La convergence en probabilité de X_n vers X n'a de sens que si les X_n et X sont définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. Grâce au théorème 9.32, il suffit de montrer que $\int_{\mathbb{R}} f \,\mathrm{d}P_{X_n}$ converge vers $\int_{\mathbb{R}} f \,\mathrm{d}P_X$ pour toute fonction f uniformément continue bornée $\mathbb{R} \to \mathbb{R}$. La continuité uniforme se traduit par

$$\forall \varepsilon > 0,\ \exists \delta > 0, \quad |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon.$$

On en déduit en prenant $x = X_n(\omega)$ et $y = X(\omega)$ que pour tout $\omega \in \{|X_n - X| < \delta\}$, $\bigl|f\bigl(X_n(\omega)\bigr) - f\bigl(X_n(\omega)\bigr)\bigr| < \varepsilon$. Par transfert, on a

$$\left|\int_{\mathbb{R}} f \,\mathrm{d}P_{X_n} - \int_{\mathbb{R}} f \,\mathrm{d}P_X\right| = \left|\int_{\Omega} \bigl(f(X_n) - f(X)\bigr) \,\mathrm{d}\mathbf{P}\right| \leq \int_{\Omega} \bigl|f(X_n) - f(X)\bigr| \,\mathrm{d}\mathbf{P}. \tag{9.70}$$

Découpant cette dernière intégrale suivant la partition de Ω en $\{|X_n - X| < \delta\}$ et son complémentaire, on obtient

$$\begin{aligned} \int_{\Omega} \bigl|f(X_n) - f(X)\bigr| \,\mathrm{d}\mathbf{P} &\leq \int_{\{|X_n - X| < \delta\}} \varepsilon \,\mathrm{d}\mathbf{P} + \int_{\{|X_n - X| \geq \delta\}} 2\|f\|_\infty \,\mathrm{d}\mathbf{P} \\ &\leq \varepsilon + 2\|f\|_\infty \mathbf{P}\bigl(|X_n - X| \geq \delta\bigr). \end{aligned} \tag{9.71}$$

Cette majoration est valable pour tout n. En exploitant l'hypothèse de convergence en probabilité de X_n vers X, on peut trouver n_0 (dépendant de δ, donc de ε et de f) tel que pour tout $n \geq n_0$, le second terme dans (9.71) soit majoré par ε. En reportant dans (9.70), on en déduit que $\bigl|\int_{\mathbb{R}} f \,\mathrm{d}P_{X_n} - \int_{\mathbb{R}} f \,\mathrm{d}P_X\bigr| < 2\varepsilon$ pour tout $n \geq n_0(\varepsilon, f)$. Ceci établit la convergence de $\int_{\mathbb{R}} f \,\mathrm{d}P_{X_n}$ vers $\int_{\mathbb{R}} f \,\mathrm{d}P_X$. Ce résultat étant valable pour toute f uniformément continue bornée, P_{X_n} converge étroitement vers P_X par le théorème 9.32 *ii*), autrement dit, X_n converge en loi vers X. □

9.3.3 Convergence en loi et fonctions caractéristiques

Théorème 9.34. *Soient X_n ($n \in \mathbb{N}^*$) et X des vecteurs aléatoires de $\mathbb{R}^d$ de fonctions caractéristiques respectives φ_{X_n} et φ_X. Alors X_n converge en loi vers X si et seulement si*

$$\forall t \in \mathbb{R}^d, \quad \varphi_{X_n}(t) \xrightarrow[n\to+\infty]{} \varphi_X(t). \tag{9.72}$$

Démonstration. On vérifie facilement la nécessité de (9.72) pour la convergence en loi de X_n vers X. En effet pour t quelconque fixé, les fonctions $x \mapsto \cos(\langle t, x\rangle)$ et $x \mapsto \sin(\langle t, x\rangle)$ sont continues et bornées sur $\mathbb{R}^d$. La convergence en loi de X_n vers X implique alors celle de $\mathbf{E}\cos(\langle t, X_n\rangle)$ vers $\mathbf{E}\cos(\langle t, X\rangle)$ et de $\mathbf{E}\sin(\langle t, X_n\rangle)$ vers $\mathbf{E}\sin(\langle t, X\rangle)$. Par linéarité de l'espérance, on en déduit la convergence $\mathbf{E}\exp(i\langle t, X_n\rangle) = \varphi_{X_n}(t)$ vers $\mathbf{E}\exp(i\langle t, X\rangle) = \varphi_X(t)$. Comme t était quelconque, (9.72) est vérifiée.

La réciproque demande un peu plus de travail. Notons pour alléger μ_n la loi de X_n et μ celle de X et supposons (9.72) vérifiée, *i.e.* que $\widehat{\mu}_n$ converge ponctuellement vers $\widehat{\mu}$ sur $\mathbb{R}^d$. Nous allons montrer la convergence étroite de μ_n vers μ en vérifiant que pour toute fonction f continue à support compact sur $\mathbb{R}^d$, $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n$ converge vers $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu$ (cf. théorème 9.32 *iii*)).

Suivant la même méthode que dans la preuve du théorème 9.9, introduisons dans le problème la loi gaussienne $\gamma_k := \mathfrak{N}(0, k^{-1/2})^{\otimes d}$. D'après la formule (9.23) établie dans la preuve du théorème 9.9 on a (en y remplaçant γ_n par γ_k) pour toute mesure de probabilité ν sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$:

$$\forall f \in \mathcal{C}_c(\mathbb{R}^d), \quad \int_{\mathbb{R}^d} f \,\mathrm{d}(\nu * \gamma_k) = c_k \int_{\mathbb{R}^d} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{ikyu} \widehat{\nu}(-ku) \,\mathrm{d}\gamma_k(u) \,\mathrm{d}\lambda_d(y). \tag{9.73}$$

Par inégalité triangulaire,

$$\left| \int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \,\mathrm{d}\mu \right| \leq I_k(\mu_n) + \left| \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu_n * \gamma_k) - \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_k) \right| + I_k(\mu), \tag{9.74}$$

où l'on a posé pour toute mesure de probabilité ν sur $\mathbb{R}^d$:

$$I_k(\nu) := \left| \int_{\mathbb{R}^d} f \,\mathrm{d}\nu - \int_{\mathbb{R}^d} f \,\mathrm{d}(\nu * \gamma_k) \right|.$$

Nous aurons besoin d'une majoration de $I_k(\nu)$, uniforme par rapport à la mesure ν. Voici comment l'obtenir. La fonction f étant continue à support compact, elle est uniformément continue sur $\mathbb{R}^d$:

$$\forall \varepsilon > 0, \ \exists \delta > 0, \ \forall x \in \mathbb{R}^d, \ \forall y \in]-\delta, \delta[^d, \quad |f(x+y) - f(x)| < \varepsilon. \tag{9.75}$$

La mesure γ_k étant une probabilité sur $\mathbb{R}^d$, la constante $\int_{\mathbb{R}^d} f \,\mathrm{d}\nu$ est γ_k intégrable sur $\mathbb{R}^d$ et on a

$$\int_{\mathbb{R}^d} f \,\mathrm{d}\nu = \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} f(x) \,\mathrm{d}\nu(x) \right\} \mathrm{d}\gamma_k(y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x) \,\mathrm{d}(\nu \otimes \gamma_k)(x, y).$$

En écrivant $\int_{\mathbb{R}^d} f \,\mathrm{d}(\nu * \gamma_k)$ à l'aide de $\nu \otimes \gamma_k$, on en déduit :

$$\begin{aligned} I_k(\nu) &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(x+y) - f(x)| \,\mathrm{d}(\nu \otimes \gamma_k)(x,y) \\ &= \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} |f(x+y) - f(x)| \,\mathrm{d}\gamma_k(y) \right\} \mathrm{d}\nu(x). \end{aligned} \tag{9.76}$$

Dans l'intégrale relative à γ_k, découpons $\mathbb{R}^d$ en $]-\delta, \delta[^d$ et son complémentaire, majorons $|f(x+y) - f(x)|$ par ε sur $]-\delta, \delta[^d$ (grâce à (9.75)) et par $2\|f\|_\infty$ sur son complémentaire. Le majorant ainsi obtenu pour l'intégrale est une constante indépendante de x. En reportant cette constante dans (9.76) et en intégrant sur $\mathbb{R}^d$ par rapport à la mesure de probabilité ν, on obtient finalement :

$$I_k(\nu) \leq \varepsilon + 2\|f\|_\infty \gamma_k\big(\mathbb{R}^d \setminus]-\delta, \delta[^d\big), \tag{9.77}$$

où δ est choisi en fonction de ε, conformément à (9.75). C'est bien la majoration uniforme en ν annoncée.

Fixons maintenant $\varepsilon > 0$ et le δ correspondant. Nous avons déjà vu, cf. (9.17), que $\gamma_k\big(\mathbb{R}^d \setminus]-\delta, \delta[^d\big)$ tend vers 0 quand k tend vers l'infini. On peut donc trouver un $k_0 = k_0(\varepsilon, f)$ tel que $I_{k_0}(\nu) < 2\varepsilon$ et ceci *quelle que soit la mesure de probabilité* ν sur $\mathbb{R}^d$. Ceci vaut en particulier pour $\nu = \mu_n$ (pour tout n) et $\nu = \mu$. Par conséquent en choisissant $k = k_0$ dans (9.74), on a :

$$\forall n \in \mathbb{N}^*, \quad \left| \int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \,\mathrm{d}\mu \right| \leq \left| \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu_n * \gamma_{k_0}) - \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_{k_0}) \right| + 4\varepsilon. \tag{9.78}$$

Admettons provisoirement que pour toute $f \in \mathcal{C}_c(\mathbb{R}^d)$,

$$\lim_{n \to +\infty} \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu_n * \gamma_{k_0}) = \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_{k_0}). \tag{9.79}$$

Alors on peut trouver un $n_0 = n_0(\varepsilon, f)$ tel que :

$$\forall n \geq n_0(\varepsilon), \quad \left| \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu_n * \gamma_{k_0}) - \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_{k_0}) \right| < \varepsilon.$$

En reportant dans (9.78), on obtient

$$\forall n \geq n_0(\varepsilon), \quad \left| \int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \,\mathrm{d}\mu \right| < 5\varepsilon.$$

Comme ε était quelconque, ceci établit la convergence de $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n$ vers $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu$. Ce résultat étant valable pour toute $f \in \mathcal{C}_c(\mathbb{R}^d)$, la condition *iii*) du théorème 9.32 est vérifiée, donc μ_n converge étroitement vers μ.

Pour compléter la preuve, il ne reste plus qu'à justifier la convergence (9.79). En posant

$$h_n(y) := c_{k_0} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{ik_0 yu} \widehat{\mu}_n(-k_0 u) \,\mathrm{d}\gamma_{k_0}(u), \quad h(y) := c_{k_0} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{ik_0 yu} \widehat{\mu}(-k_0 u) \,\mathrm{d}\gamma_{k_0}(u),$$

la formule (9.73) nous permet de réécrire la convergence (9.79) sous la forme

$$\int_{\mathbb{R}^d} f\,\mathrm{d}(\mu_n * \gamma_{k_0}) = \int_{\mathbb{R}^d} h_n(y)\,\mathrm{d}\lambda_d(y) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} h(y)\,\mathrm{d}\lambda_d(y) = \int_{\mathbb{R}^d} f\,\mathrm{d}(\mu * \gamma_{k_0}). \quad (9.80)$$

Par hypothèse, $\widehat{\mu}_n$ converge ponctuellement sur $\mathbb{R}^d$ vers $\widehat{\mu}$. On en déduit que

$$\forall y \in \mathbb{R}^d,\ \forall u \in \mathbb{R}^d, \quad \mathrm{e}^{ik_0yu}\widehat{\mu}_n(-k_0u) \xrightarrow[n\to+\infty]{} \mathrm{e}^{ik_0yu}\widehat{\mu}(-k_0u).$$

Pour chaque $y \in \mathbb{R}^d$, cette convergence est dominée relativement à n par la fonction constante 1 (considérée ici comme fonction de u) qui est γ_{k_0} intégrable sur $\mathbb{R}^d$ puisque γ_{k_0} est une probabilité. Par une première application du théorème de convergence dominée, relativement à la mesure γ_{k_0}, on en déduit :

$$\forall y \in \mathbb{R}^d, \quad h_n(y) \xrightarrow[n\to+\infty]{} h(y).$$

À son tour, cette convergence est dominée, en effet :

$$\forall n \in \mathbb{N}^*,\ \forall y \in \mathbb{R}^d, \quad |h_n(y)| \leq c_{k_0}|f(y)| \int_{\mathbb{R}^d} \left|\mathrm{e}^{ik_0yu}\widehat{\mu}_n(-k_0u)\right| \mathrm{d}\gamma_{k_0}(u) \leq c_{k_0}|f(y)|$$

et la fonction $c_{k_0}|f|$ indépendante de n est λ_d intégrable sur $\mathbb{R}^d$ parce que f est continue à support *compact*. Une deuxième application du théorème de convergence dominée, cette fois relativement à λ_d, nous donne finalement la convergence de $\int_{\mathbb{R}^d} h_n\,\mathrm{d}\lambda_d$ vers $\int_{\mathbb{R}^d} h\,\mathrm{d}\lambda_d$, ce qui établit (9.80) et achève la preuve du théorème. □

9.4 Théorème limite central

Théorème 9.35. *Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$, indépendantes, de même loi et de carré intégrable (et non* p.s. *constantes). Notons $m := \mathbf{E}X_1$, $\sigma^2 := \operatorname{Var} X_1$ avec $\sigma > 0$. Alors*

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0,1).$$

Démonstration. On se ramène à des variables Y_k centrées et de variance 1 en écrivant

$$S_n^* = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k \quad \text{où} \quad Y_k := \frac{X_k - m}{\sigma}.$$

Les Y_k sont indépendantes et de même loi. Notons ϕ la fonction caractéristique de Y_1 (et donc aussi de chaque Y_k) et φ_n celle de S_n^*. La fonction caractéristique de chacune des $n^{-1/2}Y_k$ est $t \mapsto \phi(n^{-1/2}t)$. Par indépendance et équidistribution de ces variables, on en déduit :

$$\forall t \in \mathbb{R}, \quad \varphi_n(t) = \left\{\phi\left(n^{-1/2}t\right)\right\}^n. \quad (9.81)$$

Comme Y_1 est de carré intégrable, ϕ est deux fois dérivable (cf. proposition 9.3), d'où

$$\forall u \in \mathbb{R}, \quad \phi(u) = \phi(0) + \phi'(0)u + \phi''(0)\frac{u^2}{2} + u^2\varepsilon(u)$$

avec $\varepsilon(u)$ tendant vers 0 quand u tend vers 0. La propositon 9.3 nous dit de plus que $\phi'(0) = i\mathbf{E}Y_1 = 0$ et $\phi''(0) = -\mathbf{E}Y_1^2 = -1$. Prenons $u = n^{-1/2}t$, il vient

$$\forall n \geq 1, \quad \phi\big(n^{-1/2}t\big) = 1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\big(n^{-1/2}t\big).$$

Pour t fixé, on a par le lemme 9.36 b) ci-dessous :

$$\lim_{n\to+\infty} \left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\big(n^{-1/2}t\big)\right)^n = \exp\Big(\frac{-t^2}{2}\Big). \tag{9.82}$$

Ce résultat étant valable pour tout t réel, on vient de montrer (compte-tenu de (9.81)) que φ_n converge ponctuellement sur $\mathbb{R}$ vers la fonction caractéristique de la loi $\mathfrak{N}(0,1)$. On conclut avec le théorème 9.34 que S_n^* converge en loi vers $\mathfrak{N}(0,1)$. □

Le lemme suivant permet d'éviter le recours au logarithme complexe pour justifier (9.82).

Lemme 9.36.

a) Pour tout $n \geq 1$, si $z_1, \dots, z_n$ et $z'_1, \dots, z'_n$ sont des nombres complexes de module au plus un, on a

$$|z_1 \dots z_n - z'_1 \dots z'_n| \leq \sum_{k=1}^{n} |z_k - z'_k|. \tag{9.83}$$

b) Pour tout $t \in \mathbb{R}$, si $\varepsilon(u)$ est une fonction à valeurs complexes tendant vers 0 quand u tend vers 0,

$$\left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\big(n^{-1/2}t\big)\right)^n \xrightarrow[n\to+\infty]{} \exp(-t^2/2). \tag{9.84}$$

Preuve. Le a) se vérifie par récurrence. Le cas $n=1$ est trivial. Supposons (9.83) vraie au rang n. Posons $p_n := z_1 \dots z_n$ et $p'_n := z'_1 \dots z'_n$. On peut alors écrire (justifications *a posteriori*)

$$\begin{aligned}
|p_n z_{n+1} - p'_n z'_{n+1}| &= |(p_n - p'_n)z_{n+1} + p'_n(z_{n+1} - z'_{n+1})| \\
&\leq |p_n - p'_n||z_{n+1}| + |p'_n||z_{n+1} - z'_{n+1}| && (9.85) \\
&\leq |p_n - p'_n| + |z_{n+1} - z'_{n+1}| && (9.86) \\
&\leq \sum_{k=1}^{n} |z_k - z'_k| + |z_{n+1} - z'_{n+1}| = \sum_{k=1}^{n+1} |z_k - z'_k|. && (9.87)
\end{aligned}$$

On a pu passer de (9.85) à (9.86) parce que $|z_{n+1}| \leq 1$ et $|p'_n| \leq 1$ (produit de n nombres complexes de module au plus 1). L'utilisation de l'hypothèse de récurrence pour majorer $|p_n - p'_n|$ a permis le passage de (9.86) à (9.87). Le a) est donc établi.

Pour prouver b), en prenant dans a) $z_1 = \cdots = z_n = v$ et $z'_1 = \cdots = z'_n = w$, on voit que pour tous nombres complexes v et w de module au plus 1,

$$\left|v^n - w^n\right| \leq n|v - w|. \tag{9.88}$$

Il est d'ailleurs facile d'obtenir directement cette inégalité en utilisant la factorisation de $v^n - w^n$ par $(v - w)$. Prenons maintenant

$$v := 1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\big(n^{-1/2}t\big), \quad w := \exp\Big(\frac{-t^2}{2n}\Big).$$

Pour $n \geq n_0 = n_0(t)$, on a $|v| \leq 1$ et pour tout n, $|w| < 1$. En remarquant que

$$\exp\Big(\frac{-t^2}{2n}\Big) = 1 - \frac{t^2}{2n} + \frac{\delta_n(t)}{n}, \qquad \lim_{n\to+\infty} \delta_n(t) = 0,$$

on déduit alors de (9.88) que

$$\begin{aligned} \forall n \geq n_0(t), \quad \left|\Big(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\big(n^{-1/2}t\big)\Big)^n - \exp\Big(\frac{-t^2}{2}\Big)\right| &\leq n\left|\frac{t^2}{n}\varepsilon\big(n^{-1/2}t\big) - \frac{\delta_n(t)}{n}\right| \\ &\leq t^2\left|\varepsilon\big(n^{-1/2}t\big)\right| + |\delta_n(t)| \end{aligned}$$

et ce majorant tend vers 0 quand n tend vers l'infini. □

Théorème 9.37 (De Moivre-Laplace). *Si S_n est une variable aléatoire de loi binomiale de paramètres n et $p \in]0,1[$, on a avec $q := 1 - p$,*

$$S_n^* := \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}}\Big(\frac{S_n}{n} - p\Big) \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0,1).$$

Comme la fonction de répartition Φ de $\mathfrak{N}(0,1)$ est continue sur $\mathbb{R}$, ceci équivaut à

$$\forall x \in \mathbb{R}, \quad \mathbf{P}(S_n^* \leq x) \xrightarrow[n\to+\infty]{} \Phi(x) = \int_{-\infty}^{x} \exp\Big(-\frac{t^2}{2}\Big)\frac{\mathrm{d}t}{\sqrt{2\pi}}.$$

Preuve. C'est un simple corollaire du théorème 9.35 en remarquant que S_n a *même loi* que $X_1 + \cdots + X_n$, où les X_k sont des variables aléatoires de Bernoulli indépendantes et de même paramètre p et en rappelant que l'espérance et la variance de la loi $\mathrm{Bin}(n,p)$ sont respectivement np et npq. □

La démonstration historique du théorème de De Moivre-Laplace repose sur un bon contrôle des coefficients binomiaux *via* la formule de Stirling. L'intérêt de cette approche « élémentaire » est de donner une idée de la vitesse de convergence qui est en $O(n^{-1/2})$, cf. [ICP] chap. 7.

Une application directe de ce théorème est la construction d'*intervalles de confiance* pour l'estimation d'une probabilité inconnue p à partir de l'observation d'un *échantillon* de n variables de Bernoulli indépendantes de paramètre p. Considérons pour $t > 0$ l'événement

$$A_{n,t} := \Big\{\omega \in \Omega; -t \leq \sqrt{\frac{n}{pq}}\Big(\frac{S_n(\omega)}{n} - p\Big) \leq t\Big\}.$$

Le théorème de De Moivre-Laplace nous dit que pour n assez grand, on peut utiliser l'approximation :

$$\mathbf{P}\big(A_{n,t}\big) \simeq \Phi(t) - \Phi(-t) = 2\Phi(t) - 1.$$

Ceci peut se réécrire

$$\mathbf{P}\Big(\frac{S_n}{n} - t\sqrt{\frac{pq}{n}} \leq p \leq \frac{S_n}{n} + t\sqrt{\frac{pq}{n}}\Big) = 2\Phi(t) - 1 + \varepsilon_n.$$

On ignore la valeur de p, donc *a fortiori* celle de $\sqrt{pq}$. Heureusement, il est possible de la majorer car $p(1-p)$ est maximal pour $p = 1/2$. D'où

$$\sqrt{pq} \leq \sqrt{\frac{1}{4}} = \frac{1}{2}, \tag{9.89}$$

de sorte qu'en notant

$$B_{n,t} := \Big\{\omega \in \Omega; \frac{S_n(\omega)}{n} - \frac{t}{2\sqrt{n}} \leq p \leq \frac{S_n(\omega)}{n} + \frac{t}{2\sqrt{n}}\Big\},$$

l'inclusion $A_{n,t} \subset B_{n,t}$ nous donne :

$$\mathbf{P}\big(B_{n,t}\big) \geq 2\Phi(t) - 1 + \varepsilon_n. \tag{9.90}$$

En pratique, n est fixé et on a observé des valeurs numériques explicites $x_1, \dots, x_n$ que l'on interprète comme les valeurs de $X_1(\omega), \dots, X_n(\omega)$ pour *un même* ω tiré au sort (suivant $\mathbf{P}$). On est donc en présence d'une valeur numérique explicite, $S_n(\omega)/n = (x_1 + \cdots + x_n)/n$, disons pour fixer les idées $S_n(\omega)/n = 0{,}53$. Proposer pour le paramètre inconnu p l'intervalle de confiance

$$I_{n,t} = \Big[0{,}53 - \frac{t}{2\sqrt{n}} ; 0{,}53 + \frac{t}{2\sqrt{n}}\Big],$$

c'est faire le *pari* que le ω observé est bien dans $B_{n,t}$. La probabilité de gagner ce pari est minorée par $2\Phi(t) - 1 + \varepsilon_n$. On dit que $I_{n,t}$ est un intervalle de confiance pour p avec un *niveau*[9] d'au moins $2\Phi(t) - 1 + \varepsilon_n$. En pratique, on laisse tomber le ε_n et on

9. Il y a ici un piège sémantique : supposons qu'on ait trouvé $I_{n,t} = [0{,}51; 0{,}55]$ avec un niveau de confiance de 95%. Il est tout à fait incorrect d'écrire « $\mathbf{P}(p \in [0{,}51; 0{,}55]) \geq 0{,}95$ ». En effet, p n'a rien d'aléatoire, c'est une constante. L'aléatoire concerne notre ignorance sur sa valeur. Même si on considère p comme une variable aléatoire constante, la probabilité de son appartenance à $[0{,}51; 0{,}55]$ vaut 0 ou 1, et comme on ne peut pas exclure le premier cas, on ne peut pas minorer cette probabilité par 0,95.

détermine t de façon approchée grâce à la tabulation de Φ. Par exemple pour un niveau de confiance de 95%, on est ramené à la résolution de l'équation $\Phi(t) = 1{,}95/2 = 0{,}975$ d'où $t \simeq 1{,}96$, ce qui nous donne l'intervalle

$$I_n = \left[\frac{S_n(\omega)}{n} - \frac{1{,}96}{2\sqrt{n}}; \frac{S_n(\omega)}{n} + \frac{1{,}96}{2\sqrt{n}}\right], \qquad \text{au niveau de confiance 95\%.}$$

En fait les statisticiens préfèrent une variante de cette méthode pour obtenir des intervalles de confiance plus étroits, notamment quand p n'est pas trop proche de $1/2$. L'idée est de remplacer la variance inconnue pq de X_1 par un *estimateur* au lieu de la majorer de façon certaine par (9.89). Ainsi en estimant pq par $M_n(1 - M_n)$ où $M_n := S_n/n$, on obtient au niveau de confiance 95% l'intervalle

$$J_n = \left[M_n(\omega) - 1{,}96\sqrt{\frac{M_n(\omega)(1 - M_n(\omega))}{n}}; M_n(\omega) + 1{,}96\sqrt{\frac{M_n(\omega)(1 - M_n(\omega))}{n}}\right].$$

La figure 9.2 représente (verticalement) les intervalles I_n et J_n calculés sur le même échantillon simulé de taille 200 de la loi Bern(0.75). Sur 100 simulations de ce type, on observe 4 intervalles I ne contenant pas la vraie valeur de p (les numéros 50, 60, 69 et 95) et 6 intervalles J ne contenant pas p (numéros 2, 29, 50, 60, 69 et 95).

Corollaire 9.38 (du th. 9.35, approximation gaussienne d'une loi de Poisson). *Si Y_n suit une loi de Poisson de paramètre $\alpha_n = na$ (a constante),*

$$Y_n^* := \frac{Y_n - \alpha_n}{\sqrt{\alpha_n}} \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0,1).$$

Preuve. Il suffit de remarquer que Y_n a *même loi* que $X_1 + \cdots + X_n$ où les X_k sont indépendantes et de même loi Pois(a). Par le théorème 9.35, le S_n^* bâti sur la suite des X_k converge en loi vers $\mathfrak{N}(0,1)$. Comme $\mathbf{E}Y_n = \alpha_n$ et $\operatorname{Var} Y_n = \alpha_n$, on voit que Y_n^* a même loi que S_n^*. Il en résulte que Y_n^* converge en loi vers la même limite que S_n^*. □

Ce corollaire explique que l'approximation poissonienne d'une binomiale Bin(n,p) finisse par devenir gaussienne quand $\alpha_n = np$ augmente. En pratique, on préfère l'approximation gaussienne dès que $np \geq 20$.

Remarque 9.39. Revenons au théorème de De Moivre-Laplace. Notons μ_n la loi de S_n^*, $\mu := \mathfrak{N}(0,1)$,

$$E_n := \left\{\frac{k - nm}{\sigma\sqrt{n}}; \quad 0 \leq k \leq n\right\}$$

l'ensemble fini des valeurs possibles de S_n^* et B le borélien dénombrable $B := \cup_{n\geq 1} E_n$. Comme $\mu_n(E_n) = 1$ et $E_n \subset B$, on a $\mu_n(B) = 1$ pour tout n. Ainsi $\mu_n(B)$ converge vers 1. Or $\mu(B) = 0$ puisque B est dénombrable et μ est à densité par rapport à λ_1. Donc $\mu_n(B)$ ne converge pas vers $\mu(B)$. Il n'y a pas contradiction avec la convergence étroite de μ_n vers μ car B est d'intérieur vide (il ne contient aucun intervalle) et dense dans $\mathbb{R}$ (exercice), donc de fermeture $\mathbb{R}$. La frontière de B est donc…$\mathbb{R}$ lui même!

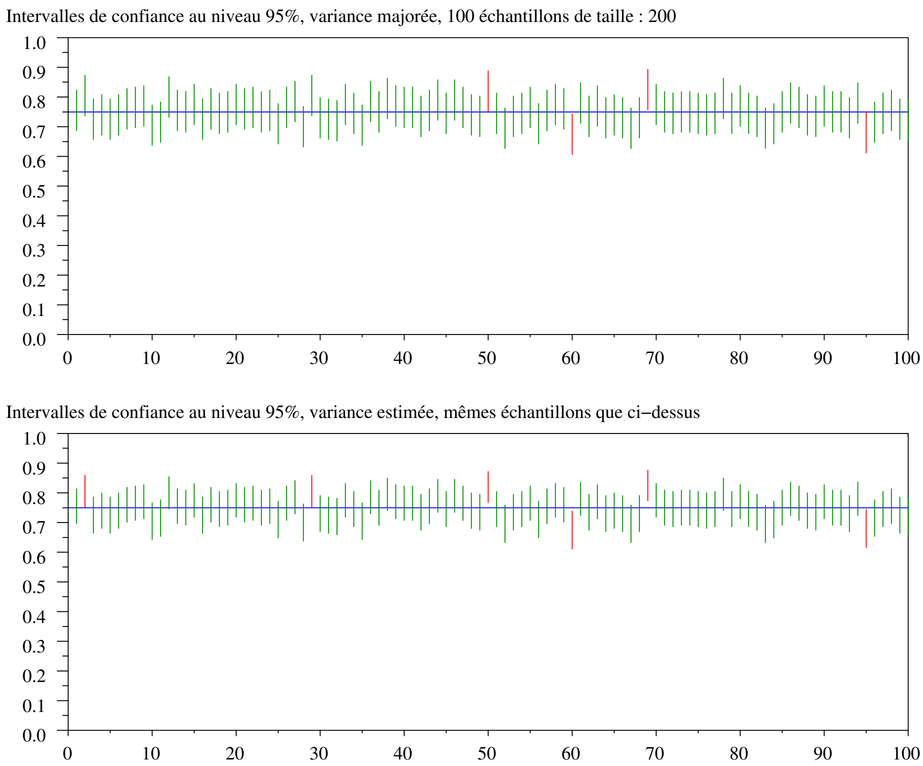

FIG. 9.2 – Intervalles I_{200} et J_{200} pour 100 échantillons simulés avec $p = 0.75$.

Donc $\mu(\partial B) = 1 \neq 0$. On voit ainsi que la restriction $\mu(\partial B) = 0$ dans (9.37) n'est pas seulement destinée à éliminer les cas pathologiques. Elle intervient aussi dans le plus classique et le plus utilisé de tous les exemples de convergence en loi.

On a le même phénomène avec l'exemple 9.6, où $B = \mathbb{Q} \cap]0, 1]$ et plus généralement, chaque fois que l'on a convergence étroite d'une suite de lois discrètes vers une loi continue.