Théorèmes limites et processus de Poisson
Notes de cours de M2
Orsay, 2011–2012

Grégory Miermont

Avant-propos

Le cours porte sur deux parties distinctes. La première concerne l'étude de la **convergence en loi de variables aléatoires** à valeurs dans des espaces topologiques plus généraux que $\mathbb{R}$ ou $\mathbb{R}^k$, ce qui nécessite de placer une topologie idoine sur l'ensemble des mesures de probabilités sur l'espace considéré.

La recherche de généralité dans le cadre de la convergence en loi n'est pas gratuite. Par exemple, au lieu de considérer la convergence (en loi) d'une variable aléatoire réelle, on peut naturellement considérer la convergence de fonctions aléatoires, également appelées *processus*. Nous donnons un tel exemple ci-dessous.

Soit $\xi_1, \xi_2, \ldots$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées, définies sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. Nous supposerons de plus que

$$E[\xi_1] = 0\,, \qquad \text{et} \qquad E[(\xi_1)^2] = 1\,.$$

On s'intéresse au comportement de la marche aléatoire $(S_n, n \geq 0)$ associée définie par $S_0 = 0$ et

$$S_n = \xi_1 + \ldots + \xi_n\,, \qquad n \geq 1\,.$$

Par exemple, le théorème de la limite centrale montre que pour tout $a < b$, on a

$$P\left(a < \frac{S_n}{\sqrt{n}} < b\right) \underset{n\to\infty}{\longrightarrow} \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x\,,$$

ou, de façon équivalente

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{loi}} N\,,$$

où N est une variable aléatoire gaussienne centrée réduite (on notera $\mathcal{N}(m, \sigma^2)$ la loi gaussienne de moyenne m et de variance σ^2). Ceci détermine la position asymptotique de S_n lorsque $n \to \infty$: on interprète le résultat précédent en disant que la loi de S_n est proche de celle de $\sqrt{n}N$.

Si l'on veut une information plus précise sur la *trajectoire* de $(S_n, n \geq 0)$, on peut s'intéresser à la position de la marche à des instants $n_1, n_2, \ldots, n_k$ rangés par ordre croissant. Par convention on notera aussi $n_0 = 0$. Les accroissements

$$S_{n_i} - S_{n_{i-1}} = \xi_{n_{i-1}+1} + \ldots + \xi_{n_i}\,, \qquad 1 \leq i \leq k$$

sont alors indépendants, et ont même loi, respectivement, que $S_{n_i - n_{i-1}}$. Supposons que $n_1, \ldots, n_k$ sont des suites (indexées disons par n) et qu'il existe $t_0 = 0 \leq t_1 \leq \ldots \leq t_k$ tels que

$$t_i = \lim_{n\to\infty} \frac{n_i}{n}\,, \qquad 1 \leq i \leq k\,.$$

Alors, une nouvelle utilisation du théorème de la limite centrale (multidimensionnel) montre que

$$\left(\frac{S_{n_i} - S_{n_{i-1}}}{\sqrt{n}}, 1 \leq i \leq k\right) \underset{n\to\infty}{\overset{\text{loi}}{\longrightarrow}} (N_1, \ldots, N_k)\,,$$

où les variables aléatoires $N_1, \ldots, N_k$ sont indépendantes, gaussiennes, de variances respectives $t_1, t_2 - t_1, \ldots, t_k - t_{k-1}$. De façon équivalente, on a que

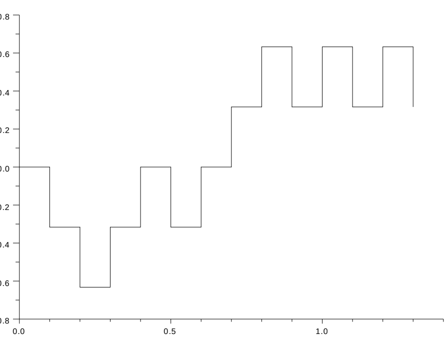
$$\left(\frac{S_{n_i}}{\sqrt{n}}, 1 \leq i \leq k\right) \underset{n\to\infty}{\overset{\text{loi}}{\longrightarrow}} (N_1, N_1 + N_2, \ldots, N_1 + \ldots + N_k)\,, \tag{1}$$

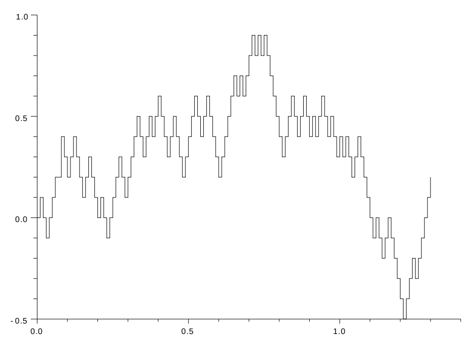
ce qu'on peut interpréter comme la détermination des positions asymptotiques de la marche aléatoire considérée en un nombre fini d'instants $n_1, \ldots, n_k$.

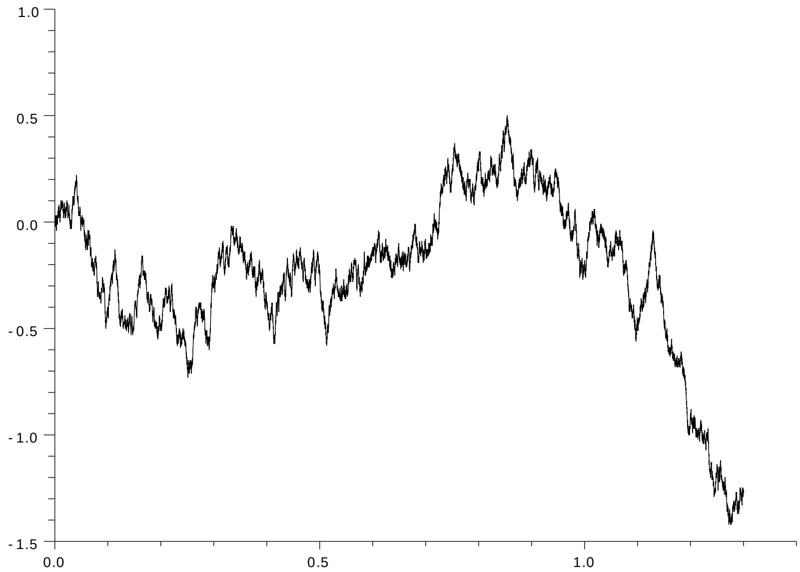
Peut-on aller plus loin et déterminer le comportement asymptotique de la marche aléatoire *dans son intégralité*? Au vu des résultats précédents, il est naturel de considérer une fonction aléatoire d'un paramètre réel positif t, correspondant à la marche aléatoire dont le temps est renormalisé par n, et l'espace par $\sqrt{n}$. Notons donc

$$S_t^{[n]} = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}}\,, \qquad t \geq 0\,,$$

et effectuons une simulation de cette fonction. Par exemple si $P(\xi_1 = 1) = P(\xi_1 = -1) = 1/2$, avec n de plus en plus grand (ici, successivement $10, 100, 10000$), nous obtenons les images suivantes (les sauts de $S^{[n]}$ sont eux aussi matérialisés par des traits verticaux).







On s'aperçoit que, lorsque n augmente, la fonction $S^{[n]}$ semble prendre la forme d'une fonction, de nature aléatoire. Cette fonction est appelée *mouvement brownien* et est un des objets les plus fondamentaux en probabilités. Précisément, un mouvement brownien standard de dimension 1 est une collection $(B_t, t \geq 0)$ de variables aléatoires indexées par $\mathbb{R}_+$, vérifiant les trois propriétés suivantes

- $B_0 = 0$ p.s.,
- la fonction $t \mapsto B_t$ est p.s. continue de $\mathbb{R}_+$ dans $\mathbb{R}$,
- pour tout $0 = t_0 \leq t_1 \leq \ldots \leq t_k$, les variables aléatoires $(B_{t_i} - B_{t_{i-1}}, 1 \leq i \leq k)$ sont indépendantes, respectivement de loi $\mathcal{N}(0, t_i - t_{i-1})$.

Il n'est pas évident *a priori* qu'une telle famille de variables aléatoires puisse être définie sur un certain espace de probabilités, mais c'est bien le cas — nous en verrons une preuve plus tard.

Avec cette définition, il est clair que la loi de $(B_{t_1}, B_{t_2}, \ldots, B_{t_k})$ est la même que celle de $(N_1, N_1+N_2, \ldots, N_1+\ldots+N_k)$ avec les mêmes notations que ci-dessus. On reformule donc (1) en écrivant que pour tout choix de réels positifs $t_1 \leq t_2 \leq \ldots \leq t_k$, on a

$$\left(S^{[n]}_{t_i}, 1 \leq i \leq k\right) \xrightarrow[n\to\infty]{\text{loi}} (B_{t_i}, 1 \leq i \leq k)\,, \tag{2}$$

Si l'on prend des instants t_i de plus en plus nombreux et resserrés, cette convergence semble bien expliquer que $S^{[n]}$ approche la trajectoire du mouvement brownien B lorsque $n \to \infty$, comme nous l'avons vu sur des simulations. Pourtant, les convergences (2) ne sont pas entièrement satisfaisantes de ce point de vue. Par exemple, il est naturel de conjecturer que l'on a

$$\sup_{0\leq t\leq 1} S^{[n]}_t \xrightarrow[n\to\infty]{\text{loi}} \sup_{0\leq t\leq 1} B_t\,,$$

mais cette convergence *n'est pas* une conséquence des convergences en loi (2). Disons que, même si l'on choisit des instants t_i très resserrés, la convergence (2) n'exclut pas *a priori* que le processus $S^{[n]}$ puisse prendre des valeurs anormalement élevées entre deux de ces instants. C'est le même genre de problème qui advient lorsqu'on a convergence ponctuelle d'une suite de fonctions, et non convergence uniforme. Il nous faut donc développer un concept de convergence en loi pour des variables aléatoires qui sont des fonctions, munies de la topologie uniforme, et c'est l'objet de la première partie du cours.

Une seconde partie du cours sera dédiée à l'étude des **mesures de Poisson**. Dans la description précédente, on peut voir que le mouvement brownien est en quelque sorte le résultat d'une accumulation d'événements contraires extrêmement fréquents : les va-et-vient d'une marche aléatoire. Par contraste, les mesures de Poisson s'intéressent à la notion d'événements rares. Pour fixer les idées, reprenons une marche aléatoire $S_n = \xi_1 + \ldots + \xi_n$, mais cette fois-ci, supposons que $P(\xi_1 = 1) = 1 - P(\xi_1 = 0) = p$. On s'intéresse à nouveau au comportement de la marche aléatoire $(S_n, n \geq 0)$. Plus exactement, donnons-nous des entiers positifs $0 = n_0 < n_1 < \ldots < n_k$. Le nombre sauts de (S_n) qui adviennent dans les intervalles $]n_{i-1}, n_i]$, égaux à $S_{n_i} - S_{n_{i-1}}$ respectivement, sont alors indépendants et suivent respectivement une loi binomiale de paramètres $(n_i - n_{i-1}, p)$. Supposons alors que $p, n_1, \ldots, n_k$ sont en fait des suites, indexées disons par l'entier n. On suppose que p tend vers 0 lorsque $n \to \infty$, de sorte que l'événement de l'occurrence d'un saut de $(S_n, n \geq 0)$ est rare. Précisément, supposons que, lorsque $n \to \infty$, on ait

$$np \longrightarrow \lambda , \qquad \frac{n_i}{n} \longrightarrow t_i , \quad 1 \leq i \leq k ,$$

où $\lambda > 0$ et $t_1 \leq t_2 \leq \ldots \leq t_k$ sont des réels positifs.

Alors les accroissements $(S_{n_i} - S_{n_{i-1}}, 1 \leq i \leq k)$ convergent en loi vers une suite $(Y_1, Y_2, \ldots, Y_k)$ de variables indépendantes, suivant respectivement la loi de Poisson d'intensité $\lambda(t_i - t_{i-1})$ (on a noté $t_0 = 0$). En effet on vérifie aisément que

$$P(S_{n_1} = k) = \binom{n_1}{k} p^k (1-p)^{n-k} \underset{n\to\infty}{\longrightarrow} e^{-\lambda t_1} \frac{(\lambda t_1)^k}{k!} = P(Y_1 = k) ,$$

et le résultat annoncé s'obtient simplement à partir de ce genre de calculs.

Le processus de Poisson d'intensité $\lambda > 0$ est une famille de variables aléatoires $(N_t^{(\lambda)}, t \geq 0)$ indexés par $\mathbb{R}_+$, vérifiant les propriétés suivantes
- $N_0^{(\lambda)} = 0$ p.s.
- p.s. la fonction $t \mapsto N_t^{(\lambda)}$ est croissante continue à droite, à valeurs dans $\mathbb{Z}_+$,
- pour tout $t_0 = 0 \leq t_1 \leq \ldots \leq t_k$, les accroissements $(N_{t_i}^{(\lambda)} - N_{t_{i-1}}^{(\lambda)}, 1 \leq i \leq k)$ sont indépendants, respectivement de loi de Poisson de paramètre $\lambda(t_i - t_{i-1})$.

On peut vérifier que le processus de Poisson peut être défini comme un *processus de comptage*. Soit $R_1, R_2, \ldots$ une suite de variables aléatoires exponentielles indépendantes et de même paramètre λ. On note $T_i = R_1 + \ldots + R_i$, et

$$N_t^{(\lambda)} = \sum_{i=1}^{\infty} \mathbb{1}_{\{R_i \leq t\}} .$$

Autrement dit, le processus $N^{(\lambda)}$ est issu de 0 et effectue un saut de 1 en chaque instant T_i. Alors $N^{(\lambda)}$ vérifie bien les propriétés définissant le processus de Poisson.

On résume les propriétés qui précèdent en disant que, pour tout $t_1 \leq \ldots \leq t_k$ positifs, on a la convergence en loi

$$(S_{\lfloor nt_i \rfloor}, 1 \leq i \leq k) \xrightarrow[n\to\infty]{\text{loi}} (N^{(\lambda)}_{t_i}, 1 \leq i \leq k)\,.$$

On comprend alors que le processus de Poisson, ou plutôt, l'ensemble $\{T_1, T_2, \ldots\}$ de ses sauts, apparaît comme limite des instants de sauts d'une marche aléatoire de Bernoulli, pour laquelle un événement de saut est très rare. Au lieu de raisonner en termes de processus, on peut considérer la mesure $M = \sum_{i\geq 1} \delta_{T_i}$ décrivant l'occurrence de ces événements rares. Les propriétés du processus de Poisson montrent donc que M est une mesure (aléatoire !) vérifiant les deux propriétés fondamentales suivantes

- si $I_1, \ldots, I_k$ sont des intervalles disjoints, alors $M(I_1), \ldots, M(I_k)$ sont des variables aléatoires indépendantes,
- si $0 \leq a < b$, alors la loi de $M(]a, b])$ est la loi de Poisson de paramètre $\lambda(b - a)$.

En effet, on vérifie aisément que $M(]a, b]) = N^{(\lambda)}_b - N^{(\lambda)}_a$. Ces deux propriétés permettent une définition des mesures de Poisson dans des espaces bien plus généraux que $\mathbb{R}_+$, comme on le verra.

Comme application principale des mesures de Poisson, nous donnerons une construction des **processus de Lévy**. Les processus de Lévy constituent une généralisation naturelle du mouvement brownien. On dit que la famille $(X_t, t \geq 0)$ de variables aléatoires indexées par le paramètre $t \in \mathbb{R}_+$ est un processus de Lévy si les trois propriétés suivantes sont vérifiées :

- $X_0 = 0$ p.s.,
- presque-sûrement, la fonction $t \mapsto X_t$ est continue à droite et admet des limites à gauche en tout point,
- pour tout $0 = t_0 \leq t_1 \leq \ldots \leq t_k$, les variables $(X_{t_i} - X_{t_{i-1}}, 1 \leq i \leq k)$ sont indépendantes, respectivement de même loi que $X_{t_i - t_{i-1}}$.

Noter que nous avons déjà rencontré deux processus de Lévy dans cette introduction : le mouvement brownien et le processus de Poisson. Nous verrons qu'en un certain sens, ces deux exemples sont les « prototypes » des processus de Lévy généraux.

Je remercie Sanjay Ramassamy pour sa relecture très attentive de ces notes.

Table des matières

1 Convergence étroite des mesures de probabilités

Dans ces notes, on adoptera la notation anglo-saxonne $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Si E est un espace topologique, on notera $\mathcal{B}_E$ la tribu des boréliens, et $\mathcal{M}_f(E)$ l'ensemble des mesures positives finies sur $(E, \mathcal{B}_E)$. On notera également $\mathcal{M}_1(E)$ le sous-ensemble formé par les mesures de probabilités. Si f est une fonction mesurable positive, ou intégrable par rapport à la mesure μ, on utilisera les notations

$$\mu(f) = \langle \mu, f \rangle = \int_E f \, \mathrm{d}\mu \, .$$

On notera $\mathcal{C}(E, F)$ l'espace des fonctions continues de l'espace topologique E vers l'espace topologique F. On notera également $\mathcal{C}_b(E, \mathbb{R}), \mathcal{C}_c(E, \mathbb{R})$ l'ensemble des fonctions de E dans $\mathbb{R}$, qui sont continues et bornées sur E, resp. continues sur E et à support compact.

Dans ce chapitre, les espaces topologiques E considérés seront toujours supposés *métrisables*, c'est-à-dire que l'on peut munir E d'une distance d qui induit la topologie de E. La plupart du temps, E sera dès le début muni d'une distance, néanmoins, il faudra garder à l'esprit que certaines distances peuvent être plus judicieuses que d'autres, dans tous les cas, il sera toujours implicite que toutes les distances considérées sur E induisent la même topologie.

Si d est une distance sur E (induisant la topologie de E), on notera $\mathcal{U}^d(E, \mathbb{R}), \mathcal{U}_b^d(E, \mathbb{R})$ les ensembles des fonctions uniformément continues sur (E, d) et à valeurs dans $\mathbb{R}$ (resp. bornées), et $\mathrm{Lip}_K^d(E, \mathbb{R})$ celui des fonctions à valeurs dans $\mathbb{R}$ qui sont K-lipschitziennes, c'est-à-dire des fonctions $f : E \to \mathbb{R}$ telles que

$$d(f(x), f(y)) \leq K \, d(x, y) \, , \qquad \text{pour tout} \quad x, y \in E \, .$$

On notera également $\mathrm{Lip}^d(E, \mathbb{R})$ la réunion des $\mathrm{Lip}_K^d(E, \mathbb{R})$ pour $K > 0$, et $\mathrm{BL}^d(E, \mathbb{R})$ le sous-ensemble des fonctions lipschitziennes bornées.

On prendra garde au fait que les espaces $\mathcal{U}^d(E, \mathbb{R})$ et $\mathrm{Lip}^d(E, \mathbb{R})$ dépendent du choix de d et non seulement de la topologie de E. Pour s'en convaincre, on pourra considérer $E = \mathbb{R}$ muni successivement des distances

$$d_1(x, y) = |x - y| \, , \qquad d_2(x, y) = |\arctan(x) - \arctan(y)| \, ,$$

qui induisent toutes deux la topologie usuelle de $\mathbb{R}$.

Dans un espace métrique (E, d), on notera $B_d(x, r)$ la boule ouverte de centre x et de rayon r, et $B_d^f(x, r)$ la boule fermée correspondante.

1.1 Rappels sur $\mathbb{R}^k$

Soit k un entier strictement positif. Dans cette section, on se place dans le cas de $E = \mathbb{R}^k$ muni de la topologie usuelle. Sauf mention contraire, on munira toujours $\mathbb{R}^k$ de la norme euclidienne standard notée $|\cdot|$.

Définition 1.1 *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures de probabilités sur $\mathbb{R}^k$. On dit que $(\mu_n)_{n\geq 1}$* converge étroitement *vers $\mu \in \mathcal{M}_1(\mathbb{R}^k)$, et on note $\mu_n \Rightarrow \mu$, si pour toute fonction $f \in \mathcal{C}_b(\mathbb{R}^k, \mathbb{R})$, on a*

$$\mu_n(f) \underset{n\to\infty}{\longrightarrow} \mu(f)\,.$$

Exemples. Si $(x_n)_{n\geq 1}$ est une suite de $\mathbb{R}^k$ qui converge vers x, alors $\delta_{x_n} \Rightarrow \delta_x$ puisque $\delta_{x_n}(f) = f(x_n) \to f(x) = \delta_x(f)$ pour toute fonction f continue.

La suite de mesures sur $\mathbb{R}$

$$\mu_n = \frac{1}{n+1}\sum_{i=0}^{n} \delta_{i/n}\,, \qquad n \geq 0$$

converge étroitement vers la mesure de Lebesgue sur $[0,1]$, puisque pour toute fonction $f \in \mathcal{C}_b(\mathbb{R}, \mathbb{R})$,

$$\mu_n(f) = \frac{1}{n+1}\sum_{i=0}^{n} f\left(\frac{i}{n}\right) \longrightarrow \int_0^1 f(x)\mathrm{d}x\,.$$

En pratique, il n'est pas toujours aisé de vérifier la convergence $\mu_n(f)$ sur toutes les fonctions $f \in \mathcal{C}_b(\mathbb{R}^k, \mathbb{R})$. Un important outil théorique et pratique est le théorème suivant (On notera A° et $\overline{A}$ respectivement l'intérieur et l'adhérence d'un sous-ensemble A d'un espace topologique, et $\partial A = \overline{A} \setminus A^\circ$ sa frontière).

Théorème 1.1 *Soit $(\mu_n)_{n\geq 1}$ une suite de $\mathcal{M}_1(\mathbb{R}^k)$ et $\mu \in \mathcal{M}_1(\mathbb{R}^k)$. On se donne également une distance d sur $\mathbb{R}^k$. Les propositions suivantes sont équivalentes.*

1. *La suite $(\mu_n)_{n\geq 1}$ converge étroitement vers μ*
2. *pour toute fonction $f \in \mathcal{U}_b^d(\mathbb{R}^k, \mathbb{R})$, on a $\lim_n \mu_n(f) = \mu(f)$*
3. *pour toute fonction $f \in \mathrm{BL}^d(\mathbb{R}^k, \mathbb{R})$, on a $\lim_n \mu_n(f) = \mu(f)$*
4. *pour tout ouvert O de $\mathbb{R}^k$, on a $\liminf_n \mu_n(O) \geq \mu(O)$*
5. *pour tout fermé F de $\mathbb{R}^k$, on a $\limsup_n \mu_n(F) \leq \mu(F)$*
6. *pour tout borélien $A \in \mathcal{B}_{\mathbb{R}^k}$ tel que $\mu(\partial A) = 0$, on a $\lim_n \mu_n(A) = \mu(A)$*
7. *pour toute fonction $f : \mathbb{R}^k \to \mathbb{R}$ mesurable, continue μ-presque partout et bornée, on a $\lim_n \mu_n(f) = \mu(f)$.*

Ce théorème étant en fait valable dans un cadre beaucoup plus général, nous n'en donnerons la preuve que plus loin. Par contre, nous allons voir qu'il est utile en pratique pour donner d'autres critères de convergence étroite.

1.1.1 Restriction de l'ensemble des fonctions-test

Proposition 1.1 *Soit $(\mu_n)_{n\geq 1}$ une suite de $\mathcal{M}_1(\mathbb{R}^k)$ et $\mu \in \mathcal{M}_1(\mathbb{R}^k)$. Soit enfin H un ensemble de fonctions mesurables bornées, dont l'adhérence pour la norme uniforme contient $\mathcal{C}_c(\mathbb{R}^k, \mathbb{R})$. Si $\mu_n(f)$ converge vers $\mu(f)$ pour tout $f \in H$, alors $\mu_n \Rightarrow \mu$.*

Preuve. Dans un premier temps on suppose que $H = \mathcal{C}_c(\mathbb{R}^k, \mathbb{R})$. Pour tout $r > 0$, considérons une fonction χ_r continue, positive, bornée par 1, égale à 1 sur la boule ouverte $B(0,r)$ de $\mathbb{R}^k$, et nulle en dehors de $B(0,r+1)$. En particulier, on a que $\mathbb{1}_{B(0,r)} \leq \chi_r \leq \mathbb{1}_{B(0,r+1)}$. Par conséquent, comme $\chi_r \in \mathcal{C}_c(\mathbb{R}^k, \mathbb{R})$, on a

$$\limsup_{n\to\infty} \mu_n(\mathbb{R}^k \setminus B(0, r+1)) \leq 1 - \liminf_{n\to\infty} \mu_n(\chi_r) = 1 - \mu(\chi_r) \leq \mu(\mathbb{R}^k \setminus B(0,r)) .$$

Soit $f \in \mathcal{C}_b(\mathbb{R}^k, \mathbb{R})$. Fixons $\varepsilon > 0$ et choisissons $r > 1$ tel que

$$\mu(\mathbb{R}^k \setminus B(0, r-1)) \leq \varepsilon/4\|f\|_\infty .$$

Par hypothèse, on peut trouver une fonction $g \in \mathcal{C}_c(\mathbb{R}^k, \mathbb{R})$ telle que $\|f\chi_r - g\|_\infty \leq \varepsilon/4$. Alors

$$\begin{aligned} |\mu_n(f) - \mu(f)| &\leq |\mu_n(f) - \mu_n(f\chi_r)| + |\mu_n(f\chi_r) - \mu_n(g)| + |\mu_n(g) - \mu(g)| \\ &\quad + |\mu(g) - \mu(f\chi_r)| + |\mu(f\chi_r) - \mu(f)| \\ &\leq \|f\|_\infty \mu_n(\mathbb{R}^k \setminus B(0,r)) + \varepsilon/4 + |\mu_n(g) - \mu(g)| \\ &\quad + \varepsilon/4 + \|f\|_\infty \mu(\mathbb{R}^k \setminus B(0,r)) . \end{aligned}$$

On a donc

$$\begin{aligned} \limsup_n |\mu_n(f) - \mu(f)| &\leq \frac{\varepsilon}{2} + \|f\|_\infty \Big(\mu(\mathbb{R}^k \setminus B(0, r-1)) + \mu(\mathbb{R}^k \setminus B(0,r))\Big) \\ &\quad + \limsup_n |\mu_n(g) - \mu(g)| \leq \varepsilon , \end{aligned}$$

par hypothèse. Comme $\varepsilon > 0$ était arbitraire, on conclut.

Si maintenant H est un ensemble général satisfaisant les hypothèses de l'énoucé. Soit $f \in \mathcal{C}_c(\mathbb{R}^k, \mathbb{R})$ et $\varepsilon > 0$. On peut trouver $g \in H$ tel que $\|f - g\|_\infty \leq \varepsilon$, et comme $|\mu_n(f) - \mu_n(g)| \leq \|f - g\|_\infty$, on a alors

$$\limsup_{n\to\infty} |\mu_n(f) - \mu(f)| \leq \varepsilon + \limsup_{n\to\infty} |\mu_n(g) - \mu(g)| = \varepsilon ,$$

par hypothèse. Donc $\mu_n(f)$ converge vers $\mu(f)$ pour toute fonction $f \in \mathcal{C}_c(\mathbb{R}^k, \mathbb{R})$, et on conclut par la première partie de la preuve que $\mu_n \Rightarrow \mu$. □

1.1.2 Fonctions de répartition

Dans le cas où $k = 1$, il existe un critère utile de convergence étroite. Si $\mu \in \mathcal{M}_f(\mathbb{R})$, on note $F_\mu(x) = \mu(]-\infty, x])$ sa fonction de répartition. C'est une fonction croissante, et on pourra noter que x est un point où F_μ n'est pas continue si et seulement si $\mu(\{x\}) > 0$, plus précisément

$$\mu(\{x\}) = F_\mu(x) - F_\mu(x-) ,$$

où l'on a noté $F_\mu(x-)$ la limite à gauche de F_μ en x.

Proposition 1.2 *Soit* $(\mu_n)_{n\geq 1}$ *une suite de mesures de probabilités sur* $\mathbb{R}$*, et* $\mu \in \mathcal{M}_1(\mathbb{R})$*. Alors* $\mu_n \Rightarrow \mu$ *si et seulement si*

$$F_{\mu_n}(x) \underset{n\to\infty}{\longrightarrow} F_\mu(x)$$

pour tout point x *qui est un point de continuité de* F_μ*.*

Preuve. Supposons dans un premier temps que $\mu_n \Rightarrow \mu$, et soit $x \in \mathbb{R}$ un point qui n'est pas un atome de μ. Alors la fonction $y \mapsto \mathbb{1}_{\{y\leq x\}}$ est μ-presque partout continue (son seul point de discontinuité est x). Donc la propriété 7. du théorème 1.1 implique que $\mu_n(]-\infty, x]) \to \mu(]-\infty, x])$, comme voulu.

Réciproquement, si $F_{\mu_n}(x) \to F_\mu(x)$ pour tout x qui n'est pas un atome de μ, on en déduit immédiatement que $\mu_n(g) \to \mu(g)$ pour toute fonction g de la forme $\sum_{i=1}^r \alpha_i \mathbb{1}_{]x_i,y_i]}$, où les extrémités des intervalles $]x_i, y_i]$ ne sont pas des atomes de μ. Comme toute fonction continue à support compact est limite uniforme de telles fonctions, on peut appliquer la proposition 1.1 et conclure que $\mu_n \Rightarrow \mu$. □

Cette reformulation de la convergence étroite des mesures sur $\mathbb{R}$ permet de déduire un intéressant résultat de type « compacité ».

Proposition 1.3 *Soit* $(\mu_n)_{n\geq 1}$ *une suite de mesures de probabilités sur* $\mathbb{R}$*. On suppose que pour tout* $\varepsilon > 0$*, il existe* $A > 0$ *tel que*

$$\sup_{n\geq 1} \mu_n(\mathbb{R} \setminus [-A, A]) < \varepsilon\,.$$

Alors, il existe une extraction $\psi(n)$ *telle que* $\mu_{\psi(n)}$ *converge étroitement lorsque* $n \to \infty$*.*

Preuve. Considérons une énumération $\{q_1, q_2, \dots\}$ des nombres rationnels. Pour tout $m \geq 1$, la suite $(F_{\mu_n}(q_m))_{n\geq 1}$ est bornée. Donc un argument d'extraction diagonale assure l'existence d'une extraction ψ telle que $F_{\mu_{\psi(n)}}(q)$ converge vers une limite notée $F_0(q)$ pour tout $q \in \mathbb{Q}$.

Définissons alors une fonction F par

$$F(x) = \lim_{q\downarrow x} F_0(q)\,, \qquad x \in \mathbb{R}$$

où la limite est considérée implicitement le long des rationnels $q > x$. Il est facile de vérifier que F est croissante et continue à droite, ce que nous laissons en exercice. Par ailleurs, par hypothèse, pour tout $\varepsilon > 0$, on peut trouver $A > 1$ tel que

$$\begin{aligned}
\mu_n(\mathbb{R} \setminus [-A, A]) &= \mu_n(]-\infty, -A[) + \mu_n(]A, \infty[) \\
&= F_{\mu_n}((-A)-) + (1 - F_{\mu_n}(A)) \\
&\leq \varepsilon\,.
\end{aligned}$$

Soit alors $q > A$ un nombre rationnel. Comme F_{μ_n} est croissante, on a que $F_{\mu_n}(-q) \leq \varepsilon$ et que $F_{\mu_n}(q) \geq 1-\varepsilon$. En passant à la limite le long de $\psi(n)$, on en déduit que $F_0(-q) \leq \varepsilon$

et $F_0(q) \geq 1-\varepsilon$, et enfin, si $B > q$, que $F(-B) \leq \varepsilon$ et $F(B) \geq 1-\varepsilon$. De ce fait, la fonction F a pour limite 0 en $-\infty$ et 1 en $+\infty$.

Par un théorème standard de théorie de la mesure, il existe donc une unique mesure de probabilités $\mu \in \mathcal{M}_1(\mathbb{R})$ telle que $F = F_\mu$: c'est la mesure de Stieltjes associée à F, et parfois notée dF.

Il reste à vérifier que $\mu_{\psi(n)} \Rightarrow \mu$. Soit $x \in \mathbb{R}$ un point de continuité de F. Alors il vient aisément que

$$F(x) = \lim_{q\uparrow x} F_0(q) = \lim_{q\downarrow x} F_0(q)\,,$$

les deux limites étant considérées le long des rationnels, respectivement $< x$ et $> x$. Finalement, par monotonie, si $q < x < q'$ et $q, q' \in \mathbb{Q}$, on a

$$F_0(q) = \lim_n F_{\mu_{\psi(n)}}(q) \leq \liminf_n F_{\mu_{\psi(n)}}(x) \leq \limsup_n F_{\mu_{\psi(n)}}(x) \leq \lim_n F_{\mu_{\psi(n)}}(q') = F_0(q')\,,$$

et en faisant tendre q et q' vers x on obtient finalement que $\lim_n F_{\mu_{\psi(n)}}(x) = F(x)$. La proposition 1.2 montre donc que $\mu_{\psi(n)} \Rightarrow \mu$. □

1.1.3 Fonctions caractéristiques

Si $\mu \in \mathcal{M}_1(\mathbb{R}^k)$, on note $\phi_\mu(u) = \int_{\mathbb{R}^k} e^{\mathrm{i}u\cdot x}\mu(\mathrm{d}x)$ la fonction caractéristique de μ, où $u \in \mathbb{R}^k$ et $u \cdot x$ désigne le produit scalaire usuel.

Un résultat fondamental liant fonction caractéristique et convergence étroite est le suivant.

Proposition 1.4 *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures de probabilités sur $\mathbb{R}^k$, et $\mu \in \mathcal{M}_1(\mathbb{R}^k)$. Alors $\mu_n \Rightarrow \mu$ si et seulement si*

$$\phi_{\mu_n}(u) \underset{n\to\infty}{\longrightarrow} \phi_\mu(u) \qquad \textit{pour tout} \quad u \in \mathbb{R}^k\,.$$

Le sens direct est bien sûr évident. Pour le sens réciproque, on peut en fait donner un résultat plus précis, dû à Lévy.

Théorème 1.2 (Théorème de Lévy) *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures de probabilités sur $\mathbb{R}^k$. On suppose que*
- *pour tout $u \in \mathbb{R}^k$, $\phi_{\mu_n}(u)$ converge vers une limite notée $\phi(u)$, et que*
- *la fonction $u \mapsto \phi(u)$ est continue en 0.*

Alors, il existe une mesure $\mu \in \mathcal{M}_1(\mathbb{R}^k)$ telle que $\phi = \phi_\mu$, et l'on a $\mu_n \Rightarrow \mu$.

Nous allons montrer ce théorème uniquement dans le cas où $k = 1$. Nous avons d'abord recours au résultat utile suivant.

Lemme 1.1 *Il existe une constante $C > 0$ telle que, pour tout $\mu \in \mathcal{M}_1(\mathbb{R})$ et pour tout $K > 0$, on a*

$$\mu(\mathbb{R} \setminus [-K, K]) \leq C\,K \int_{-1/K}^{1/K} (1 - \Re\phi_\mu(u))\mathrm{d}u\,.$$

Preuve. Remarquons que $\Re\phi_\mu(u) = \int_{\mathbb{R}} \cos(ux)\mu(\mathrm{d}x)$. De ce fait, le théorème de Fubini donne

$$\begin{aligned} K\int_{-1/K}^{1/K}(1-\Re\phi_\mu(u))\mathrm{d}u &= K\int_{\mathbb{R}}\mu(\mathrm{d}x)\int_{-1/K}^{1/K}(1-\cos(ux))\mathrm{d}u \\ &= 2\int_{\mathbb{R}}\mu(\mathrm{d}x)\left(1-\frac{\sin(x/K)}{x/K}\right) . \end{aligned}$$

La fonction $y \mapsto \sin(y)/y$, prolongée en 0 par continuité, est majorée par 1 et vérifie que

$$\frac{\sin(y)}{y} \leq \sin(1) , \qquad |y| \geq 1 ,$$

comme on peut s'en convaincre par une étude de fonctions. Autrement dit, pour tout $y \in \mathbb{R}$,

$$(1-\sin(1))\mathbb{1}_{\mathbb{R}\setminus[-1,1]}(y) \leq \left(1-\frac{\sin(y)}{y}\right) ,$$

d'où l'on tire

$$K\int_{-1/K}^{1/K}(1-\Re\phi_\mu(u))\mathrm{d}u \geq 2(1-\sin(1))\int_{\mathbb{R}}\mu(\mathrm{d}x)\mathbb{1}_{\mathbb{R}\setminus[-1,1]}(x/K) ,$$

ce qu'il fallait démontrer, en prenant $C = (2(1-\sin(1)))^{-1}$ (cette constante est bien sûr loin d'être optimale). □

Preuve du théorème 1.2. Nous allons d'abord montrer qu'il existe une extraction $\psi(n)$ le long de laquelle on a la convergence étroite de μ_n.

Pour cela, on a recours à la proposition 1.3. Le lemme 1.1 appliqué à μ_n implique que pour tout $K > 0$,

$$\limsup_n \mu_n(\mathbb{R}\setminus[-K,K]) \leq C\,K \limsup_{n\to\infty}\int_{-1/K}^{1/K}(1-\Re\phi_{\mu_n}(u))\mathrm{d}u .$$

Par hypothèse, on a convergence ponctuelle de ϕ_{μ_n} vers une fonction ϕ. Comme ϕ_{μ_n} est une fonction caractéristique, elle est de module borné par 1, et on peut appliquer le théorème de convergence dominée pour obtenir que

$$\lim_{n\to\infty}\int_{-1/K}^{1/K}(1-\Re\phi_{\mu_n}(u))\mathrm{d}u = \int_{-1/K}^{1/K}(1-\Re\phi(u))\mathrm{d}u$$

Par ailleurs, comme ϕ est supposée continue en 0, et que $\phi(0) = 1$ (puisque $\phi_{\mu_n}(0) = 1$ pour tout n), on a que

$$\lim_{K\to\infty} K\int_{-1/K}^{1/K}(1-\Re\phi(u))\mathrm{d}u = 2(1-\Re\phi(0)) = 0 ,$$

donc pour tout $\varepsilon > 0$, on peut choisir $K_1 > 0$ tel que $CK_1 \int_{-1/K_1}^{1/K_1} (1 - \Re\phi(u))\mathrm{d}u \leq \varepsilon/2$. On en déduit alors que pour tout n assez grand, disons, $n > n_0$, on a

$$\mu_n(\mathbb{R} \setminus [-K_1, K_1]) \leq \varepsilon\,.$$

Par ailleurs, choisissons K_2 tel que

$$\max_{1 \leq n \leq n_0} \mu_n(\mathbb{R} \setminus [-K_2, K_2]) \leq \varepsilon\,.$$

Alors, en posant $A = \max(K_1, K_2)$, on a bien

$$\sup_{n \geq 1} \mu_n(\mathbb{R} \setminus [-A, A]) \leq \varepsilon\,.$$

La proposition 1.3 permet bien de conclure à l'existence d'une extraction ψ et de $\mu \in \mathcal{M}_1(\mathbb{R})$ telles que $\mu_{\psi(n)} \Rightarrow \mu$.

On applique alors le sens facile de la proposition 1.4, et on en déduit que pour tout $u \in \mathbb{R}$, $\phi_{\psi(n)}(u)$ converge vers $\phi_\mu(u)$. Donc ϕ est bien la fonction caractéristique d'une mesure de probabilités.

Il reste à montrer que μ_n converge étroitement vers μ, et pas seulement le long de l'extraction ψ. Supposons par l'absurde que ce n'est pas le cas. Alors on pourrait trouver une nouvelle extraction ψ', et une fonction $f \in \mathcal{C}_b(\mathbb{R}, \mathbb{R})$, tels que $|\mu_{\psi'(n)}(f) - \mu(f)| \geq \varepsilon_0$ pour un certain $\varepsilon_0 > 0$ et tout $n \geq 1$. Mais les hypothèses de la proposition 1.3 s'appliquent toujours à la suite $(\mu_{\psi'(n)})_{n\geq 1}$ au lieu de $(\mu_n)_{n\geq 1}$. Donc on peut ré-extraire de cette sous-suite, une sous-sous-suite $(\mu_{\psi''(n)})_{n\geq 1}$ qui converge étroitement vers une limite μ'. Comme ϕ_{μ_n} converge ponctuellement vers $\phi = \phi_\mu$, et $\phi_{\mu_{\psi''(n)}}$ converge ponctuellement vers $\phi_{\mu'}$, toujours par le sens facile de la proposition 1.4, on en déduit que $\phi_{\mu'} = \phi_\mu$. Par la propriété d'injectivité de la transformée de Fourier, on en déduit que $\mu = \mu'$. Donc $\mu_{\psi''(n)}$ converge vers μ, et par définition $\mu_{\psi''(n)}(f)$ converge vers $\mu(f)$, ce qui est contradictoire avec le fait que $|\mu_{\psi'(n)}(f) - \mu(f)| \geq \varepsilon_0$ pour tout $n \geq 1$. □

Ce mode de raisonnement en deux étapes (preuve de l'existence de limites selon une sous-suite, puis identification de la limite) est extrêmement utile lorsqu'on manipule la convergence étroite des mesures, et il le deviendra d'autant plus que l'espace sous-jacent considéré sera compliqué.

Question. Le théorème 1.2 reste-t-il vrai si, toutes choses égales par ailleurs, au lieu de supposer que $\phi_{\mu_n}(u)$ converge vers $\phi(u)$ pour tout $u \in \mathbb{R}^k$, on suppose seulement que cette convergence a lieu sur un voisinage de 0 ?

1.2 Mesures de probabilités sur un espace métrique

Dans la suite de ce chapitre, on se concentre sur l'étude de l'ensemble $\mathcal{M}_1(E)$ des mesures de probabilités boréliennes sur un espace topologique E. Nous supposerons *toujours* à partir de maintenant que l'espace E vérifie les propriétés suivantes :

– L'espace E est séparable : il existe une suite dense dans E.

– Il existe une distance d sur E (induisant la topologie de E) telle que (E, d) est un espace métrique complet.

Définition 1.2 *Un espace topologique satisfaisant la définition précédente est appelé* espace polonais.

En pratique, la distance d est souvent « fournie » avec l'espace E, et on appelle souvent *espace polonais* un espace métrique, séparable et complet. Néanmoins, il faut se souvenir que la notion de complétude dépend de la métrique choisie, comme la notion de fonctions uniformément continues. Il peut donc arriver que l'on doive choisir une « meilleure » distance que celle qui est livrée naturellement avec E.

Par exemple, l'espace $\mathbb{R}$ muni de la topologie usuelle est polonais. En effet, $\mathbb{R}$ est séparable, et la topologie de $\mathbb{R}$ est induite par la distance $d_1(x, y) = |x - y|$, qui est complète. Par contre, $\mathbb{R}$ n'est pas complet pour la distance $d_2(x, y) = |\arctan(x) - \arctan(y)|$, même si cette dernière induit bien la même topologie. Donc il ne faut pas s'y tromper : l'espace métrique $(\mathbb{R}, d_2)$ (ou plutôt l'espace topologique induit) est bien un espace polonais...

La grande majorité des espaces dans lesquels les probabilistes travaillent, c'est-à-dire considèrent des variables aléatoires, sont des espaces polonais, car ces espaces ont de « bonnes » propriétés vis-à-vis de la notion de convergence étroite que l'on va considérer maintenant. Cependant, une grande partie des notions que nous allons étudier s'étend à des espaces topologiques plus compliqués. On consultera par exemple l'ouvrage de Billingsley [1] pour un traitement très général.

Rappelons quelques propriétés élémentaires des mesures de probabilités sur un espace métrique.

Proposition 1.5 *Soit $\mu \in \mathcal{M}_1(E)$ où (E, d) est un espace métrique. Alors pour tout $A \in \mathcal{B}_E$, on a*

$$\begin{aligned} \mu(A) &= \inf\{\mu(O) : O \textit{ ouvert contenant } A\} \\ &= \sup\{\mu(F) : F \textit{ fermé contenu dans } A\} . \end{aligned}$$

De plus, la mesure μ est déterminée par les intégrales $\mu(f)$, où f décrit l'ensemble des fonctions lipschitziennes sur (E, d).

Bien sûr, on aurait pu, dans la seconde partie de la proposition, se restreindre aux fonctions qui sont, par exemple, 1-lipschitziennes (il suffit de diviser f par sa constante de Lipschitz et utiliser la linéarité de l'intégrale). Notons aussi que la notion de fonction lipschitziennes varie énormément selon le choix de d : la précédente proposition est valide pour tout tel choix. Par contraste, la première partie de la proposition ne fait pas intervenir la distance, et la seule hypothèse est donc que E est métrisable.

Preuve. Soit $\mathcal{A}$ l'ensemble des parties boréliennes A de E telles que A vérifie

$$\begin{aligned} \mu(A) &= \inf\{\mu(O) : O \text{ ouvert contenant } A\} \\ &= \sup\{\mu(F) : F \text{ fermé contenu dans } A\} . \end{aligned}$$

Nous vérifions que $\mathcal{A}$ contient les ouverts. Soit donc O un ouvert de E. La première égalité est évidente. Par ailleurs, pour tout $k \geq 1$, l'ensemble $F_k = \{x \in E : d(x, E \setminus O) \geq 1/k\}$ est clairement un fermé, si l'on note $d(x, B) = \inf_{y \in B} d(x, y)$. Les fermés $F_k, k \geq 1$ forment une suite croissante d'union O, car tout point de O est à distance strictement positive de son complémentaire. Par continuité des mesures par limite croissante, on a bien $\mu(O) = \lim \uparrow \mu(F_k)$.

Par ailleurs, $\mathcal{A}$ est une tribu. En effet, elle contient $\varnothing$. Elle est de plus stable par complémentaire, car si $A \in \mathcal{A}$ alors pour tout $\varepsilon > 0$ on peut trouver $F \subset A \subset O$ avec F fermé et O ouvert, tels que $\mu(O \setminus F) \leq \varepsilon$. Mais alors on a aussi $E \setminus O \subset E \setminus A \subset E \setminus F$ et on a bien $\mu((E \setminus F) \setminus (E \setminus O)) \leq \varepsilon$, et donc $E \setminus A$ est aussi dans $\mathcal{A}$. Enfin, soit $A_1, A_2, \ldots$ une suite de A. Fixons $\varepsilon > 0$ et donnons-nous des fermés $F_1, F_2, \ldots$ et des ouverts $O_1, O_2, \ldots$ tels que

$$F_i \subset A_i \subset O_i \qquad \text{et} \qquad \mu(O_i \setminus F_i) \leq \frac{\varepsilon}{2^i}\,, \qquad i \geq 1\,.$$

L'ensemble $O = \bigcup_{i \geq 1} O_i$ est alors un ouvert contenant $A = \bigcup_{i \geq 1} A_i$. En revanche, la réunion F' des F_i n'est pas nécessairement un ouvert. De ce fait, on se donne N fixé tel que

$$\mu\left(F' \setminus \bigcup_{i=1}^{N} F_i\right) \leq \varepsilon\,,$$

et on note $F = \bigcup_{i=1}^{N} F_i$, qui est fermé (noter qu'un tel N existe bien par la continuité des mesures de probabilités par limite décroissante). On a alors

$$\begin{aligned} \mu(O \setminus F) &\leq \mu(O \setminus F') + \mu(F' \setminus F) \\ &\leq \sum_{i \geq 1} \mu(O_i \setminus F') + \varepsilon \\ &\leq \sum_{i \geq 1} \mu(O_i \setminus F_i) + \varepsilon \\ &\leq 2\varepsilon\,. \end{aligned}$$

On en déduit que $A \in \mathcal{A}$. Finalement, $\mathcal{A}$ est une tribu contenant les ouverts, c'est donc la tribu borélienne tout entière.

Nous venons de montrer que la mesure μ est déterminée par sa valeur sur les ouverts, ou sur les fermés. Soit donc F un fermé. La fonction

$$f_{F,K}(x) = \max(1 - K\, d(x, F), 0) \tag{3}$$

est alors K-lipschitzienne, bornée par 1, et converge en décroissant vers la fonction indicatrice de F lorsque $K \to \infty$. En effet, on a utilisé la propriété que $d(x, F) = 0$ si et seulement si $x \in F$, qui vient du fait que F est fermé. On déduit, par théorème de convergence dominée, que $\mu(F) = \lim_{K \to \infty} \mu(f_{F,K})$. On en déduit bien que les intégrales μ contre les fonctions lipschitziennes déterminent les valeurs de μ sur les fermés, et donc déterminent la mesure μ elle-même. □

1.3 Convergence étroite sur $\mathcal{M}_1(E)$

La définition de la convergence étroite sur un espace topologique (polonais) est la même que dans $\mathbb{R}^k$. Ainsi

Définition 1.3 *Soit $(\mu_n)_{n\geq 1}$ une suite de $\mathcal{M}_1(E)$, et $\mu \in \mathcal{M}_1(E)$. On dit que μ_n converge étroitement vers μ, et on note $\mu_n \Rightarrow \mu$, si pour toute fonction $f \in \mathcal{C}_b(E, \mathbb{R})$ on a $\mu_n(f) \to \mu(f)$.*

Nous montrons maintenant un résultat fondamental sur la convergence étroite, qui est l'exact analogue du théorème 1.1 dans notre cas général, et dont la preuve avait été laissée en suspens. Noter que dans l'énoncé suivant, on ne demande pas nécessairement que (E, d) soit un espace métrique complet, ni même séparable.

Théorème 1.3 *Soit $(\mu_n)_{n\geq 1}$ une suite de $\mathcal{M}_1(E)$ et $\mu \in \mathcal{M}_1(E)$. On se donne également une distance d sur E. Les propositions suivantes sont équivalentes.*

1. *La suite $(\mu_n)_{n\geq 1}$ converge étroitement vers μ*
2. *pour toute fonction $f \in \mathcal{U}_b^d(E, \mathbb{R})$, on a $\lim_n \mu_n(f) = \mu(f)$*
3. *pour toute fonction $f \in \mathrm{BL}^d(E, \mathbb{R})$, on a $\lim_n \mu_n(f) = \mu(f)$*
4. *pour tout ouvert O de E, on a $\liminf_n \mu_n(O) \geq \mu(O)$*
5. *pour tout fermé F de E, on a $\limsup_n \mu_n(F) \leq \mu(F)$*
6. *pour tout borélien $A \in \mathcal{B}_E$ tel que $\mu(\partial A) = 0$, on a $\lim_n \mu_n(A) = \mu(A)$*
7. *pour toute fonction $f : E \to \mathbb{R}$ mesurable, continue μ-presque partout et bornée, on a $\lim_n \mu_n(f) = \mu(f)$.*

Preuve. Les implications 1. $\Longrightarrow$ 2. $\Longrightarrow$ 3. et 7. $\Longrightarrow$ 1. sont évidentes, de même que l'équivalence entre 4. et 5. par un simple passage au complémentaire.

Montrons que 3. $\Longrightarrow$ 5. Soit donc F un fermé de E. On considère la fonction $f_{F,K}$ définie en (3), qui est lipschitzienne et vérifie $\mathbb{1}_F \leq f_{F,K} \leq 1$. Ainsi, pour tout $n \geq 1$ on a $\mu_n(F) \leq \mu_n(f_{F,K})$, et comme on a supposé 3., on en déduit que $\limsup_n \mu_n(F) \leq \mu(f_{F,K})$. Comme $f_{F,K}$ converge vers $\mathbb{1}_F$ ponctuellement et est bornée par 1, on en déduit par convergence dominée que $\limsup_n \mu_n(F) \leq \mu(F)$

Montrons que 4. et 5. impliquent 6. Soit donc $A \in \mathcal{B}_E$, et supposons que $\mu(\partial A) = 0$, de sorte que $\mu(A) = \mu(A^\circ) = \mu(\overline{A})$. On applique 4. et 5. aux ensembles A° et $\overline{A}$, et on trouve

$$\begin{aligned}\mu(A) = \mu(A^\circ) &\leq \liminf_n \mu_n(A^\circ) \leq \liminf_n \mu_n(A) \\ &\leq \limsup_n \mu_n(A) \leq \limsup_n \mu_n(\overline{A}) \leq \mu(\overline{A}) = \mu(A)\,,\end{aligned}$$

et c'est ce que l'on voulait montrer.

Montrons enfin que 6. $\Longrightarrow$ 7. Soit donc f une fonction continue μ-presque partout et bornée. Sans perte de généralité, on peut supposer que f est positive (on peut en effet écrire $f = f_+ - f_-$ et raisonner sur chaque terme). Soit D l'ensemble des points de discontinuité de f. Notons d'abord que pour tout $\nu \in \mathcal{M}_1(E)$, on a

$$\nu(f) = \int_0^\infty \nu(\{f \geq y\})\mathrm{d}y\,. \tag{4}$$

C'est en effet une conséquence immédiate du théorème de Fubini, en écrivant

$$\nu(f) = \int_E \nu(\mathrm{d}x) \int_0^\infty \mathbb{1}_{[0,f(x)]}(y)\mathrm{d}y = \int_0^\infty \mathrm{d}y \int_E \nu(\mathrm{d}x)\mathbb{1}_{\{f(x)\geq y\}}\,.$$

Par ailleurs, pour tout $y \geq 0$, notons $A_y = \{x : f(x) \geq y\}$. Soit $x \in \overline{A}_y$, de sorte que x est limite d'une suite x_n telle que $f(x_n) \geq y$. Si $x \notin D$, c'est-à-dire si x est point de continuité de f, alors on a aussi $f(x) \geq y$. Donc $\overline{A}_y \subseteq A_y \cup D$. Par ailleurs, si $f(x) > y$ et $x \notin D$, alors on a également $f(x') > y$ pour x' dans un voisinage de x. Donc $\{f > y\} \setminus D \subseteq A_y^\circ$. Finalement, on en déduit que $\partial A_y \subseteq \{f = y\} \cup D$.

Par ailleurs, l'ensemble $\{y \geq 0 : \mu(\{f = y\}) > 0\}$ est au plus dénombrable. Il est en effet la réunion des ensembles $\{y \geq 0 : \mu(\{f = y\}) \geq 1/r\}, r \geq 1$, qui sont respectivement de cardinal au plus r, puisque les ensembles $\{f = y\}$ sont deux-à-deux disjoints. Par 6., on en déduit que pour Lebesgue-presque tout $y \geq 0$, on a $\mu_n(\{f \geq y\}) \to \mu(\{f \geq y\})$ quand $n \to \infty$. Donc par convergence dominée, en utilisant (4) et le fait que f est bornée, on a

$$\begin{aligned}\mu_n(f) = \int_0^\infty \mu_n(\{f \geq y\})\mathrm{d}y \quad &= \quad \int_0^{\|f\|_\infty} \mu_n(\{f \geq y\})\mathrm{d}y \\ &\underset{n\to\infty}{\longrightarrow} \int_0^{\|f\|_\infty} \mu(\{f \geq y\})\mathrm{d}y = \mu(f)\,,\end{aligned}$$

comme voulu. □

1.4 Topologie étroite et métrisabilité de $\mathcal{M}_1(E)$

Rappelons que le dual topologique V' d'un espace vectoriel normé V est l'espace vectoriel des formes linéaires continues sur V. En voyant $\mathcal{M}_1(E)$ comme sous-ensemble de $\mathcal{C}_b(E,\mathbb{R})'$, on voit que la convergence étroite correspond à la notion de convergence pour la topologie faible-* en restriction à $\mathcal{M}_1(E)$, également appelée topologie étroite. C'est la topologie la moins fine rendant continues les applications $\mu \mapsto \mu(f)$ pour $f \in \mathcal{C}_b(E,\mathbb{R})$. Une base d'ouverts est donnée par les ensembles

$$\mathcal{V}(\nu; f_1, \ldots, f_r; \varepsilon_1, \ldots, \varepsilon_r) = \{\mu \in \mathcal{M}_1(E) : |\mu(f_i) - \nu(f_i)| < \varepsilon_i, 1 \leq i \leq r\}\,,$$

indexés par une mesure $\nu \in \mathcal{M}_1(E)$, des fonctions $f_1, \ldots, f_r \in \mathcal{C}_b(E,\mathbb{R})$, et des réels $\varepsilon_1, \ldots, \varepsilon_r > 0$.

On peut se demander si la topologie étroite est métrisable : ce genre de question est importante si l'on veut, par exemple, avoir une caractérisation séquentielle des fonctions continues pour la topologie étroite. Il se trouve que c'est le cas sous l'hypothèse que E est polonais (en fait, métrique séparable suffirait).

Proposition 1.6 *La topologie étroite sur $\mathcal{M}_1(E)$ est métrisable.*

Pour le montrer, notons quelques faits importants sur les espaces séparables. Rappelons qu'un espace métrique est dit précompact si pour tout $\varepsilon > 0$, on peut le recouvrir par un nombre fini de boules de rayon ε.

Lemme 1.2 *Un espace métrique séparable peut être muni d'une distance qui induit la même topologie, et pour laquelle il est précompact.*

Preuve. Soit (E, d) un espace séparable métrique, et $(\alpha_n, n \geq 1)$ une suite dense. Alors l'application

$$x \mapsto h(x) = \left(\frac{d(x, \alpha_n)}{1 + d(x, \alpha_n)} \right)_{n \geq 1}$$

est une application continue de E vers $[0,1]^{\mathbb{N}}$ (muni de la topologie produit), qui réalise un homéomorphisme de E sur son image. La continuité est évidente, vérifions d'abord que h est injective. Pour cela, il suffit de voir que si $x \neq y$ alors comme on peut trouver un $n \geq 1$ tel que $d(x, \alpha_n) < d(x, y)/2$, on a alors forcément $d(x, \alpha_n) \neq d(y, \alpha_n)$. Enfin, la fonction réciproque de h est continue, car si $(x_m)_{m \geq 1}$ est une suite telle que $d(x_m, \alpha_n)$ converge vers $d(x, \alpha_n)$ pour tout n, on peut choisir $\varepsilon > 0$ et un n tel que $d(x, \alpha_n) < \varepsilon/2$, et on obtient que $d(x_m, \alpha_n) < \varepsilon/2$ pour tout m assez grand, d'où l'on a $d(x_m, x) \leq \varepsilon$ pour m assez grand, et donc x_m converge vers x.

Si $\mathbf{a}, \mathbf{b}$ sont dans $[0,1]^{\mathbb{N}}$, on note alors

$$D(\mathbf{a}, \mathbf{b}) = \sum_{n \geq 1} \frac{|a_n - b_n|}{2^n} ,$$

et on vérifie sans peine que ceci définit une distance induisant la topologie produit sur $[0,1]^{\mathbb{N}}$. Donc par ce qui précède, si on pose

$$\hat{d}(x, y) = D(h(x), h(y)) , \qquad x, y \in E ,$$

alors $\hat{d}$ est une distance sur E, qui rend ce dernier précompact (car h le réalise comme sous-espace du compact $[0,1]^{\mathbb{N}}$). □

Lemme 1.3 *Si (E, d) est un espace métrique précompact, alors l'ensemble $\mathcal{U}^d(E, \mathbb{R}) = \mathcal{U}^d_b(E, \mathbb{R})$ est séparable pour la norme uniforme.*

Preuve. Soit $(\hat{E}, d)$ le complété de (E, d). Alors il est précompact et complet, d'où l'on déduit qu'il est compact (exercice). Soit (α_n) une suite dense de E. Alors l'algèbre des polynômes à coefficients rationnels en les fonctions

$$x \in \hat{E} \mapsto d(x, \alpha_n) , \qquad n \geq 1$$

est une algèbre dénombrable qui sépare les points. Par le théorème de Stone-Weierstraß, on en déduit que cette algèbre est dense dans $\mathcal{C}(\hat{E}, \mathbb{R}) = \mathcal{U}^d(\hat{E}, \mathbb{R})$ pour la norme uniforme. On obtient le résultat en restreignant à E, après avoir remarqué que toute fonction uniformément continue sur (E, d) s'étend de façon unique en une fonction de $\mathcal{C}(\hat{E}, \mathbb{R})$. □

Preuve de la Proposition 1.6. Par le lemme 1.2, nous pouvons munir E d'une distance $\hat{d}$ qui le rend précompact. Le lemme 1.3 assure que $\mathcal{U}^{\hat{d}}(E,\mathbb{R}) = \mathcal{U}_b^{\hat{d}}(E,\mathbb{R})$ est séparable. Soit donc $(g_n, n \geq 1)$ une suite dense de telles fonctions. On pose

$$\operatorname{dist}(\mu,\nu) = \sum_{n\geq 1} \frac{1}{2^n} |\mu(g_n) - \nu(g_n)| \wedge 1 \,.$$

Alors dist est une distance sur $\mathcal{M}_1(E)$, comme le lecteur s'en convaincra facilement. La topologie induite par cette distance est clairement la topologie la moins fine rendant continues les applications $\mu \mapsto \mu(g_n)$ pour $n \geq 1$, et de ce fait, cette topologie est moins fine que la topologie étroite. En particulier, si $\mu_n \Rightarrow \mu$ lorsque $n \to \infty$, alors $\operatorname{dist}(\mu_n, \mu) \to 0$.

Réciproquement, si $\operatorname{dist}(\mu_n,\mu) \to 0$ lorsque $n \to \infty$, cela signifie que pour tout $m \geq 1$, on a $\mu_n(g_m) \to \mu(g_m)$. Soit alors f une fonction de $\mathcal{U}^d(E,\mathbb{R})$ et $\varepsilon > 0$, on peut trouver $m \geq 1$ tel que $\|f - g_m\|_\infty \leq \varepsilon/2$, et donc

$$\limsup_n |\mu_n(f) - \mu(f)| \leq \varepsilon + \limsup_n |\mu_n(g_m) - \mu(g_m)| = \varepsilon \,,$$

ce qui signifie que $\mu_n \Rightarrow \mu$ par le 2. du théorème 1.3.

On en déduit que si $\mathcal{V}$ est un voisinage de $\mu \in \mathcal{M}_1(E)$ pour la topologie étroite, c'est aussi un voisinage de μ pour $(\mathcal{M}_1(E), \operatorname{dist})$. En effet, dans le cas contraire, on pourrait trouver une suite $(\mu_n)_{n\geq 1}$ telle que $\mu_n \notin \mathcal{V}$ pour tout n mais $\operatorname{dist}(\mu_n,\mu) \to 0$ lorsque $n \to \infty$. Mais cela implique que $\mu_n \Rightarrow \mu$ et donc que μ_n est dans $\mathcal{V}$ pour n assez grand, ce qui est une contradiction. □

En fait, on a le résultat suivant, plus fort.

Théorème 1.4 *L'espace $\mathcal{M}_1(E)$ (muni de la topologie étroite) est polonais.*

Nous verrons une preuve de ce résultat plus loin, lorsque nous nous intéresserons à d'autres types de métriques sur $\mathcal{M}_1(E)$. Nous justifions cependant dès maintenant une partie de ce résultat.

Proposition 1.7 *L'espace $\mathcal{M}_1(E)$ est séparable.*

Preuve. Nous utilisons d'abord les lemmes 1.2 et 1.3 : soit d est une distance telle que (E,d) est précompact, et si $g_1, g_2, \ldots$ est une suite de $\mathcal{U}^d(E,\mathbb{R})$ de fonctions vérifiant $\sup_{i\geq 1} \|g_i\|_\infty \leq 1$ et $\operatorname{vect}(\{g_i, i \geq 1\})$ est dense dans $\mathcal{U}^d(E,\mathbb{R})$ (prendre pour cela une suite dense de fonctions uniformément continues non-nulles, et les diviser par leur norme uniforme) alors l'application

$$H : \begin{array}{ccc} \mathcal{M}_1(E) & \longrightarrow & [-1,1]^{\mathbb{N}} \\ \mu & \longmapsto & (\mu(g_m))_{m\geq 1} \,, \end{array}$$

est un homéomorphisme de $\mathcal{M}_1(E)$ sur son image ($[-1,1]^{\mathbb{N}}$ étant muni de la topologie produit). Les arguments ont en fait déjà été avancés lors de la preuve de la proposition

1.6, où nous avons déjà vérifié que pour avoir $\mu_n \Rightarrow \mu$, il suffit d'avoir $\mu_n(f) \to \mu(f)$ où f décrit un ensemble dense pour la norme uniforme de fonctions uniformément continues.

Ainsi, $\mathcal{M}_1(E)$ est homéomorphe à un sous-espace du compact $[-1,1]^{\mathbb{N}}$, séparable, et ce sous-espace est donc lui-même séparable (exercice). □

En fait, on peut être plus descriptif et donner un exemple explicite de sous-ensemble dénombrable dense. La proposition suivante est laissée en exercice.

Proposition 1.8 *Soit $(\alpha_i)_{i\geq 1}$ une suite dense de E. L'ensemble des mesures atomiques de la forme*

$$\sum_{j=1}^{n} q_j \delta_{\beta_j} ,$$

où $n \geq 1$, $\beta_1, \ldots, \beta_n \in \{\alpha_1, \alpha_2, \ldots\}$ et les nombres $q_1, \ldots, q_n$ sont des rationnels positifs de somme 1, est un ensemble dénombrable dense dans $\mathcal{M}_1(E)$.

Remarque. La notion de topologie étroite ne dépend que du fait que E est un espace topologique, et non du fait qu'il est polonais. En se plaçant dans cette généralité, les résultats précédents ont les énoncés frappants suivant : « $\mathcal{M}_1(E)$ est séparable (resp. compact, polonais), si E est séparable (resp. compact, polonais) ». Les réciproques sont également vraies ! On consultera par exemple le livre de Parthasarathy [9].

1.5 Le théorème de Prokhorov

Le théorème de Prokhorov, qui est sans doute l'outil le plus fondamental de ce chapitre, donne un critère de relative compacité pour les parties de $\mathcal{M}_1(E)$.

Pour bien comprendre ce théorème, nous commençons par énoncer le résultat fondamental suivant. Rappelons que dans toute la suite, E est un espace polonais, et $\mathcal{M}_1(E)$ sera toujours muni de la topologie étroite.

Théorème 1.5 *Si E est compact, alors $\mathcal{M}_1(E)$ est compact.*

La preuve de ce théorème repose sur un théorème de Riesz que nous rappelons ici (voir par exemple l'ouvrage de Rudin [10] pour une preuve). Rappelons qu'une forme linéaire Λ sur $\mathcal{C}(E,\mathbb{R})$ est dite positive si pour toute fonction positive $f \in \mathcal{C}(E,\mathbb{R})$ on a $\Lambda(f) \geq 0$.

Théorème 1.6 (Riesz) *Soit (E,d) un espace métrique compact et Λ une forme linéaire positive sur $\mathcal{C}(E,\mathbb{R})$, telle que $\Lambda(1) = 1$. Alors il existe une unique mesure $\mu \in \mathcal{M}_1(E)$ telle que $\Lambda(f) = \mu(f)$ pour tout $f \in \mathcal{C}(E,\mathbb{R})$.*

Preuve du théorème 1.5. Par le lemme 1.3, on peut trouver une suite dense $(g_m, m \geq 1)$ de $\mathcal{U}^d(E,\mathbb{R}) = \mathcal{C}(E,\mathbb{R})$ pour la norme uniforme. Sans perte de généralité on peut supposer que $g_1 = 1$.

Soit $(\mu_n)_{n\geq 1}$ une suite de $\mathcal{M}_1(E)$. Comme pour tout $m \geq 1$, la suite $(\mu_n(g_m))_{n\geq 1}$ est bornée, par un argument d'extraction diagonale, on peut alors trouver une sous-suite $(\mu_{\psi(n)})_{n\geq 1}$ telle que $\mu_{\psi(n)}(g_m)$ converge vers une limite notée $\Lambda(g_m)$ lorsque $n \to \infty$.

Par linéarité, on peut alors étendre Λ en une forme linéaire sur le sous-espace vectoriel engendré par $\{g_m, m \geq 1\}$. Plus précisément, si $f \in \{g_m, m \geq 1\}$ s'écrit sous la forme

$$f = \sum_{m=1}^{p} \alpha_m g_m \,,$$

alors on pose

$$\Lambda(f) = \sum_{m=1}^{p} \alpha_m \Lambda(g_m) = \lim_{n\to\infty} \mu_{\psi(n)}(f) \,,$$

la dernière écriture montrant que la première formule ne dépend pas du choix particulier de la combinaison linéaire. Cette forme linéaire est continue puisque $|\Lambda(f)| = \lim |\mu_{\psi_n(f)}| \leq \|f\|_\infty$ pour tout $f \in \mathrm{vect}(\{g_m, m \geq 1\})$. Elle est donc uniformément continue (car 1-lipschitzienne) de $\mathcal{C}(E, \mathbb{R})$ dans $\mathbb{R}$. De là, on en déduit que Λ s'étend de façon unique en une forme linéaire continue sur l'adhérence de $\mathrm{vect}(\{g_m, m \geq 1\})$, c'est-à-dire sur $\mathcal{C}(E, \mathbb{R})$. On appelle encore Λ cette forme linéaire.

Alors $\Lambda(1) = 1$ puisque $g_1 = 1$ et $\Lambda(g_1) = \lim_n \mu_{\psi(n)}(g_1)$. De plus, Λ est positive. En effet, si $f \geq 0$, et $(g'_m)_{n\geq 1}$ est une suite de $\mathrm{vect}(\{g_m, m \geq 1\})$ qui converge vers f, alors pour tout $\varepsilon > 0$ on a que $g'_m + \varepsilon$ est une fonction positive pour tout m assez grand. Donc

$$\Lambda(g'_m + \varepsilon) = \Lambda(g'_m) + \varepsilon \geq 0$$

(c'est la limite de $\mu_{\psi(n)}(g'_m + \varepsilon)$), et en faisant tendre $m \to \infty$ on obtient $\Lambda(f) + \varepsilon \geq 0$, et donc $\Lambda(f) \geq 0$ en faisant tendre ε vers 0.

Par le théorème de Riesz, on en déduit qu'il existe $\mu \in \mathcal{M}_1(E)$ tel que $\Lambda(f) = \mu(f)$ pour tout $f \in \mathcal{C}(E, \mathbb{R})$. Il reste à montrer que $\mu_{\psi(n)} \Rightarrow \mu$. Soit donc $f \in \mathcal{C}(E, \mathbb{R})$ et $\varepsilon > 0$. On choisit m tel que $\|f - g_m\|_\infty \leq \varepsilon/2$ et on obtient

$$\begin{aligned} |\mu_{\psi(n)}(f) - \mu(f)| &\leq |\mu_{\psi(n)}(f - g_m)| + |\mu_{\psi(n)}(g_m) - \mu(g_m)| + |\mu(f - g_m)| \\ &\leq \varepsilon + |\mu_{\psi(n)}(g_m) - \mu(g_m)| \,, \end{aligned}$$

donc $\limsup_n |\mu_{\psi(n)}(f) - \mu(f)| \leq \varepsilon$, comme voulu. □

Il est intéressant de noter que la réciproque est vraie.

Proposition 1.9 *Si $\mathcal{M}_1(E)$ est compact, alors E est compact.*

Preuve. Supposons que l'on puisse trouver une suite $(x_n)_{n\geq 1}$ de E qui n'a pas de sous-suite convergente. On peut supposer sans perte de généralité que les termes de cette suite sont deux-à-deux distincts. Alors l'ensemble $F = \{x_n, n \geq 1\}$ est un fermé, car toute suite convergente de F est forcément stationnaire. Pour la même raison, tout sous-ensemble de F est aussi fermé. On considère alors la suite des mesures $(\delta_{x_n})_{n\geq 1}$. Par compacité de $\mathcal{M}_1(E)$, cette suite admet une sous-suite $(\delta_{x_{\psi(n)}})_{n\geq 1}$ qui converge vers une limite μ. Par le théorème 1.3, on a alors

$$\limsup_{n\to\infty} \delta_{x_{\psi(n)}}(F') \leq \mu(F')$$

pour tout $F' \subseteq F$. Mais, dès lors que F' contient une infinité de termes de la suite $(x_{\psi(n)})_{n\geq 1}$, la limite supérieure $\limsup_{n\to\infty} \delta_{x_{\psi(n)}}(F')$ vaut 1. Donc $\mu(\{x_{\psi(n)} : n \geq p\}) = 1$ pour tout p. C'est impossible car cette quantité tend vers 0 lorsque $p \to \infty$. $\square$

Définition 1.4 *Le sous-ensemble Γ de $\mathcal{M}_1(E)$ est dit* tendu *si pour tout $\varepsilon > 0$, il existe un compact $K_\varepsilon \subseteq E$ tel que*

$$\sup_{\mu\in\Gamma} \mu(E \setminus K_\varepsilon) \leq \varepsilon\,.$$

Autrement dit, une partie de $\mathcal{M}_1(E)$ est tendue si l'on peut trouver un seul et même compact qui concentre la plupart de la masse de tous les éléments de cette partie. On peut voir le théorème suivant comme une généralisation du théorème 1.3.

Théorème 1.7 (Prokhorov) *Soit $\Gamma \subseteq \mathcal{M}_1(E)$. Alors Γ est relativement compacte si et seulement si elle est tendue.*

Preuve. Supposons que Γ est un sous-ensemble de $\mathcal{M}_1(E)$ qui est tendu. Nous allons montrer qu'il est compact en utilisant la caractérisation séquentielle (toute suite de Γ admet une sous-suite qui converge), rendue licite par la proposition 1.6.

Grâce au lemme 1.2, on peut munir E d'une métrique $\hat{d}$ qui le rend précompact. L'espace complété de $(E, \hat{d})$, noté $(\hat{E}, \hat{d})$, est alors compact.

Notons que si $\mu \in \mathcal{M}_1(E)$, alors on peut étendre μ en une mesure $\hat{\mu}$ sur $(\hat{E}, \mathcal{B}_{\hat{E}})$ par la formule

$$\hat{\mu}(A) = \mu(A \cap E)\,, \qquad A \in \mathcal{B}_{\hat{E}}\,.$$

On laissera au lecteur le soin de montrer que

$$\mathcal{B}_E = \Big\{A \cap E : A \in \mathcal{B}_{\hat{E}}\Big\}\,,$$

qui provient du fait que la topologie de E est la restriction de la topologie de $\hat{E}$ à E.

Soit alors $(\mu_n)_{n\geq 1}$ une suite de Γ. La suite de mesures $(\hat{\mu}_n)_{n\geq 1}$ est à valeurs dans $\mathcal{M}_1(\hat{E})$, donc d'après le théorème 1.5, elle admet une sous-suite $(\hat{\mu}_{\psi(n)})_{n\geq 1}$ qui converge vers une limite $\mu' \in \mathcal{M}_1(\hat{E})$.

Montrons que μ' est portée par un borélien inclus dans E. C'est là qu'intervient la tension : pour tout entier $r \geq 1$ on peut en effet trouver un compact $K_r \subseteq E$ tel que

$$\mu_n(K_r) = \hat{\mu}_n(K_r) \geq 1 - 1/r\,, \qquad \text{pour tout} \quad n \geq 1$$

et comme K_r est fermé dans $\hat{E}$, le 5. du théorème 1.3 montre que

$$\mu'(K_r) \geq \limsup_{n\to\infty} \hat{\mu}_{\psi(n)}(K_r) \geq 1 - 1/r\,,$$

et donc que

$$\mu'\bigg(\bigcup_{r\geq 1} K_r\bigg) = 1\,.$$

On définit alors la mesure $\mu \in \mathcal{M}_1(E)$ par restriction :

$$\mu(A \cap E) = \mu'(A) = \mu'\Big(A \cap \bigcup_{r \geq 1} K_r\Big), \qquad A \in \mathcal{B}_{\hat{E}},$$

et il reste à vérifier que $\mu_{\psi(n)} \Rightarrow \mu$. Soit donc $f \in \mathcal{U}_b^{\hat{d}}(E, \mathbb{R})$, et $\hat{f}$ son prolongement continu à $\hat{E}$. Comme $\hat{\mu}_n, n \geq 1$ et μ' sont portées par $\bigcup_{r \geq 1} K_r$, on a

$$\mu_{\psi(n)}(f) = \hat{\mu}_{\psi(n)}(\hat{f}) \underset{n \to \infty}{\longrightarrow} \mu'(\hat{f}) = \mu(f).$$

Donc $\mu_{\psi(n)} \Rightarrow \mu$ par le 3. du théorème 1.3. On en conclut que Γ est relativement compact.

Réciproquement, supposons que Γ est relativement compact. On peut supposer sans perte de généralité que Γ est compact, quitte à considérer son adhérence : il est clair que Γ est tendu si son adhérence est tendue.

Soit $(O_n)_{n \geq 1}$ une suite croissante d'ouverts de réunion E Nous vérifions dans un premier temps que pour tout $\varepsilon > 0$, il existe $N \geq 1$ tel que pour tout $\mu \in \Gamma$, $\mu(O_N) > 1 - \varepsilon$. En effet, si ce n'était pas le cas, on pourrait trouver un $\varepsilon > 0$ et une suite $(\mu_n)_{n \geq 1}$ de Γ telle que $\mu_n(O_n) \leq 1 - \varepsilon$ pour tout $n \geq 1$. Comme Γ est relativement compact, quitte à extraire et à renuméroter les éléments des suites $(O_n)_{n \geq 1}$ et $(\mu_n)_{n \geq 1}$, on peut supposer que $\mu_n \Rightarrow \mu$ pour un $\mu \in \mathcal{M}_1(E)$. Par le 4. du théorème 1.3, on a alors, pour tout $n \geq 1$,

$$\liminf_m \mu_m(O_m) \geq \liminf_m \mu_m(O_n) \geq \mu(O_n),$$

et donc $\mu(O_n) \leq 1 - \varepsilon$ pour tout n, ce qui est absurde puisque $\bigcup_n O_n = E$.

Choisissons maintenant $\varepsilon > 0$, et fixons une suite $(\alpha_i, i \geq 1)$ dense dans E. On munit E d'une distance d telle que (E, d) est complet. Par ce qui précède, pour tout $k \geq 1$, il existe un entier N_k tel que pour tout $\mu \in \Gamma$,

$$\mu\Big(\bigcup_{i=1}^{N_k} B_d(\alpha_i, 1/k)\Big) > 1 - \frac{\varepsilon}{2^k}.$$

Posons alors

$$K = \bigcap_{k \geq 1} \bigcup_{i=1}^{N_k} B_d^f(\alpha_i, 1/k),$$

où $B_d^f(\alpha_i, 1/k)$ est la boule fermée de centre α_i et de rayon $1/k$. Alors K est fermé dans E, donc (K, d) est complet. De plus, il est clairement précompact. Donc K est compact, et si $\mu \in \Gamma$, on a

$$\mu(E \setminus K) \leq \sum_{k \geq 1} \frac{\varepsilon}{2^k} = \varepsilon.$$

On en déduit que Γ est tendu. □

Remarque. Dans la preuve précédente, on pourra noter que le sens réciproque n'utilise que le fait que E est un espace métrique séparable. La complétude n'intervient (et n'est effectivement nécessaire) que dans la preuve du sens direct. Il se trouve que le sens réciproque « tendu $\Longrightarrow$ compact », qui est en pratique le plus utile, est valable dans une généralité bien plus grande.

Dans la suite du cours, nous utiliserons très souvent le résultat suivant, qui découle du fait général, que dans un espace métrique, pour montrer qu'une suite converge, il suffit de montrer qu'elle prend ses valeurs dans un compact, et que toute sous-suite convergente de cette suite converge vers une même limite. On dira qu'une suite de mesures de probabilités $(\mu_n)_{n\geq 1}$ est tendue si l'ensemble $\{\mu_n, n \geq 1\}$ est tendu.

Proposition 1.10 *Soit $(\mu_n)_{n\geq 1}$ une suite de $\mathcal{M}_1(E)$. Supposons que cette suite forme une famille tendue, et qu'il existe $\mu \in \mathcal{M}_1(E)$ telle que tout sous-suite convergente de $(\mu_n)_{n\geq 1}$ ait pour limite μ. Alors $\mu_n \Rightarrow \mu$.*

1.6 Métriques complètes sur $\mathcal{M}_1(E)$

Comme annoncé précédemment au théorème 1.4, l'espace $\mathcal{M}_1(E)$ est lui-même un espace polonais. Pour voir cela, nous allons donner explicitement une distance complète sur $\mathcal{M}_1(E)$.

1.6.1 Distance de Lévy-Prokhorov

Soit d une distance sur E. On définit la *distance de Lévy-Prokhorov* (associée à d) par

$$d_{LP}(\mu, \mu') = \inf\{r > 0 : \mu(F) \leq \mu'(F^r) + r\,,\ \text{pour tout } F \subseteq E \text{ fermé}\}\,,$$

pour tout $\mu, \mu' \in \mathcal{M}_1(E)$, où l'on note $A^r = \{x \in E : d(x, A) < r\}$ le r-épaissi de $A \subseteq E$ (c'est un ouvert).

Proposition 1.11 *La fonction d_{LP} est une distance sur $\mathcal{M}_1(E)$.*

Preuve. La fonction d_{LP} est évidemment positive et prend des valeurs finies (en fait dans $[0, 1]$). Montrons que d_{LP} est symétrique. Pour cela, supposons que r est tel que $\mu(F) \leq \mu'(F^r)+r$ pour tout F fermé. Soit F un fermé donné. Notons alors que l'ensemble $F' = E \setminus F^r$ est un fermé, et que $(F')^r \subseteq E \setminus F$, ou encore, $F \subseteq E \setminus (F')^r$. Ainsi,

$$\mu'(F) \leq 1 - \mu'((F')^r) \leq 1 - \mu(F') + r = \mu(F^r) + r\,.$$

Comme F était un fermé quelconque, on en conclut que si $d_{LP}(\mu, \mu') \leq r$, alors $d_{LP}(\mu', \mu) \leq r$. Donc d_{LP} est bien symétrique.

Par ailleurs, on a clairement que $d_{LP}(\mu, \mu) = 0$. Réciproquement, si $d_{LP}(\mu, \mu') = 0$ alors pour tout F fermé et tout $\varepsilon > 0$ on a $\mu(F) \leq \mu'(F^\varepsilon) + \varepsilon$, et comme F^ε décroît vers F lorsque $\varepsilon \to 0$ (on a utilisé le fait que F est fermé), $\mu(F) \leq \mu'(F)$. Par symétrie, on a donc $\mu(F) = \mu'(F)$ pour tout F fermé, et donc $\mu = \mu'$ par la proposition 1.5.

Enfin, d_{LP} vérifie l'inégalité triangulaire, car si l'on a trois mesures μ, μ', μ'' et deux nombres positifs r, r' tels que pour tout fermé F on a

$$\begin{array}{rcl} \mu(F) & \leq & \mu'(F^r) + r \\ \mu'(F) & \leq & \mu''(F^{r'}) + r' \,, \end{array}$$

alors on a aussi, pour tout F fermé,

$$\mu(F) \leq \mu'(\overline{F^r}) + r \leq \mu''((\overline{F^r})^{r'}) + r + r' \leq \mu''(F^{r+r'}) + r + r' \,,$$

et donc $d_{LP}(\mu, \mu'') \leq r + r'$. On prend alors l'infimum sur r et r' et on conclut. □

Lemme 1.4 *Soit $\mu_n, n \geq 1$ et μ des éléments de $\mathcal{M}_1(E)$. Si $d_{LP}(\mu_n, \mu) \to 0$ lorsque $n \to \infty$, alors $\mu_n \Rightarrow \mu$.*

Preuve. Soit F un fermé de E. Comme $d_{LP}(\mu_n, \mu) \to 0$, il existe une suite $(\varepsilon_n)_{n\geq 1}$ de limite nulle, telle que $\mu_n(F) \leq \mu(F^{\varepsilon_n}) + \varepsilon_n$. Donc $\limsup_n \mu_n(F) \leq \mu(F)$, donc $\mu_n \Rightarrow \mu$ par le 5. du théorème 1.3. □

Proposition 1.12 *La distance d_{LP} induit la topologie étroite sur $\mathcal{M}_1(E)$.*

Preuve. D'après le lemme 1.4, on déduit que les ouverts de $\mathcal{M}_1(E)$ (muni de la topologie étroite) sont aussi des ouverts pour la topologie induite par d_{LP} par le même raisonnement que celui utilisé à la fin de la preuve de la proposition 1.6.

Montrons, réciproquement, que pour tout $\mu \in \mathcal{M}_1(E)$ et $\varepsilon > 0$, la boule $B_{d_{LP}}(\mu, 2\varepsilon)$ est un voisinage de μ pour la topologie étroite. Pour cela, donnons-nous une partition $(A_i)_{i\geq 1}$ d'ensembles de $\mathcal{B}_E$, tels que $\operatorname{diam}(A_i) \leq \varepsilon$ pour tout $i \geq 1$. Par exemple, on peut poser $A_1 = B_d(\alpha_1, \varepsilon/2)$, et par récurrence,

$$A_{k+1} = B_d(\alpha_{k+1}, \varepsilon/2) \setminus \bigcup_{i=1}^{k} A_i \qquad \text{pour tout} \quad k \geq 1 \,,$$

où l'on a choisi une suite $(\alpha_i)_{i\geq 1}$ dense dans E. Si certains des A_i ainsi définis sont vides, on les ignore et on renumérote en fonction.

Soit alors $N \geq 1$ tel que $\mu(\bigcup_{i>N} A_i) < \varepsilon$. Si $I \subseteq \{1, \ldots, N\}$, on note $A_I = \bigcup_{i \in I} A_i$, et on pose

$$f_I(x) = \left(1 - \frac{1}{\varepsilon} d(x, A_I)\right) \vee 0 \,,$$

de sorte que

$$\mathbb{1}_{A_I} \leq f_I \leq \mathbb{1}_{A_I^\varepsilon} \,.$$

Soit alors $\mathcal{V}$ l'ensemble

$$\{\nu \in \mathcal{M}_1(E) : |\nu(f_I) - \mu(f_I)| < \varepsilon \text{ pour tout } I \subseteq \{1, \ldots, N\}\} \,,$$

qui est un voisinage de μ pour la topologie étroite.

Montrons que $\mathcal{V} \subseteq B_{d_{LP}}(\mu, 2\varepsilon)$. Pour cela, on se donne $F \subseteq E$ un fermé. Posons

$$I_F = \{i \in \{1, \ldots, N\} : F \cap A_i \neq \varnothing\}\,.$$

Comme diam $A_i \leq \varepsilon$, pour tout $i \in I_F$, on a $(A_{I_F})^\varepsilon \subseteq F^{2\varepsilon}$. De ce fait

$$\begin{aligned} \mu(F) &\leq \mu(A_{I_F}) + \mu(F \setminus A_{I_F}) \\ &\leq \mu(A_{I_F}) + \mu\Big(\bigcup_{i>N} A_i\Big) \\ &\leq \mu(f_{I_F}) + \varepsilon \\ &\leq \nu(f_{I_F}) + 2\varepsilon \\ &\leq \nu((A_{I_F})^\varepsilon) + 2\varepsilon \\ &\leq \nu(F^{2\varepsilon}) + 2\varepsilon\,. \end{aligned}$$

D'où le résultat. □

Nous montrons maintenant le résultat annoncé plus haut 1.4, stipulant que $\mathcal{M}_1(E)$ polonais. C'est une conséquence immédiate du résultat suivant et de la proposition 1.7.

Proposition 1.13 *Soit d une distance telle que (E, d) est complet, et d_{LP} la métrique de Prokhorov qu'elle induit. Alors $(\mathcal{M}_1(E), d_{LP})$ est complet.*

Preuve. Soit $(\mu_n)_{n\geq 1}$ une suite de Cauchy pour la distance d_{LP}. Pour montrer qu'elle converge, il suffit de montrer qu'elle admet une sous-suite convergente, et par le théorème de Prokhorov il suffit pour cela de montrer qu'elle est tendue. Soit $\varepsilon > 0$. Comme $(\mu_n)_{n\geq 1}$ est de Cauchy, on peut trouver une sous-suite $(\mu_{\psi(n)})_{n\geq 1}$ telle que pour tout $m \geq \psi(n)$, on ait

$$d_{LP}(\mu_m, \mu_{\psi(n)}) \leq \varepsilon 2^{-n-1}\,.$$

Pour tout $n \geq 1$, soit K_n un compact tel que

$$\max_{1\leq i\leq \psi(n)} \mu_i(E \setminus K_n) \leq \frac{\varepsilon}{2^{n+1}}\,.$$

Un tel compact existe car l'ensemble $\{\mu_1, \ldots, \mu_{\psi(n)}\}$ est tendu par le théorème de Prokhorov. Alors on a, pour tout $m \geq \psi(n)$,

$$\mu_m(K_n^{\varepsilon/2^{n+1}}) \geq \mu_{\psi(n)}(K_n) - \frac{\varepsilon}{2^{n+1}} \geq 1 - \frac{\varepsilon}{2^n}\,,$$

et cela est également valable pour $m \in \{1, 2, \ldots, \psi(n)\}$ par définition. Posons alors

$$K = \bigcap_{n\geq 1} \overline{K_n^{\varepsilon/2^{n+1}}}\,.$$

C'est un fermé dans un complet, qui est également précompact puisque pour tout n, l'ensemble $K_n^{\varepsilon/2^{n+1}}$ peut être recouvert par un nombre fini de boules de rayon $\varepsilon/2^n$, qui tend vers 0 lorsque $n \to \infty$. Donc K est compact. De plus, pour tout $m \geq 1$ on a

$$\mu_m(E \setminus K) \leq \sum_{n\geq 1} \mu_m(E \setminus K_n^{\varepsilon/2^{n+1}}) \leq \sum_{n\geq 1} \frac{\varepsilon}{2^n} = \varepsilon\,.$$

Ce qui montre bien que $(\mu_m)_{m\geq 1}$ est tendue. $\square$

Nous terminons cette section en donnant une autre écriture très utile de la distance de Lévy-Prokhorov, sans donner de preuve. On pourra consulter le livre de Dudley [3] ou de Ethier-Kurtz [4] pour plus de détails.

Si $\mu, \nu \in \mathcal{M}_1(E)$, on appelle *couplage* entre μ et ν une mesure $\rho \in \mathcal{M}_1(E \times E)$ telle que, si p_1 et p_2 sont les projections canoniques $p_1(x_1, x_2) = x_1$ et $p_2(x_1, x_2) = x_2$ de $E \times E$ sur E, alors μ est la mesure image de ρ par p_1 et ν est la mesure image de ρ par p_2. On note $\mathrm{Co}(\mu, \nu)$ l'ensemble des couplages entre μ et ν. Noter que $\mathrm{Co}(\mu, \nu)$ n'est pas vide, puisqu'il contient par exemple la mesure produit $\mu \otimes \nu$.

On notera au passage que l'espace $E \times E$ muni de la topologie produit est un espace polonais (exercice).

Proposition 1.14 *Soit $\mu, \nu \in \mathcal{M}_1(E)$, alors*

$$d_{LP}(\mu, \nu) = \inf \Big\{ r \geq 0 : \exists \rho \in \mathrm{Co}(\mu, \nu), \rho\big\{(x, y) : d(x, y) \geq r\big\} \leq r \Big\}.$$

Cette réécriture de la distance de Lévy-Prokhorov est loin d'être triviale dans le sens $\geq$, le sens $\leq$, qui sert beaucoup en pratique, est aisé et laissé en exercice. Elle tire une partie de son utilité du fait qu'elle a une interprétation probabiliste. En effet, on peut interpréter un couplage ρ entre μ et ν comme la loi d'une variable aléatoire (Y, Z), définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, où les lois marginales de Y et de Z sont respectivement μ et ν. Un couplage naturel est le couplage indépendant, où Y et Z sont indépendantes de lois respectives μ et ν. La distance de Prokhorov correspond donc au plus petit r tel qu'il existe un couplage (Y, Z) pour lequel $\mathbb{P}(d(Y, Z) \geq r) \leq r$. Noter qu'en général, le couplage indépendant est loin d'être optimal !

1.6.2 Distance de Fortet-Mourier

Une autre distance utile est la *distance de Fortet-Mourier*, définie par

$$d_{FM}(\mu, \mu') = \sup\{|\mu(f) - \mu'(f)| : f \in \mathrm{Lip}_1^d(E, \mathbb{R}), \|f\|_\infty \leq 1\}.$$

Nous laissons au lecteur le soin de montrer, en s'inspirant éventuellement des preuves précédentes, le résultat suivant.

Proposition 1.15 *La fonction d_{FM} est une distance sur $\mathcal{M}_1(E)$, et l'espace $(\mathcal{M}_1, d_{FM})$ est complet.*

1.6.3 Topologie étroite sur les mesures finies

La plupart des affirmations de cette section sont laissées sans démonstration, et donc en exercice au lecteur.

Rappelons que $\mathcal{M}_f(E)$ est l'ensemble des mesures positives finies sur $(E, \mathcal{B}_E)$. Cet ensemble peut également être muni de la topologie étroite, qui correspond à la topologie faible-* restreinte à $\mathcal{M}_f(E)$, ce dernier étant vu comme sous-espace de $\mathcal{C}_b(E, \mathbb{R})'$. Ainsi, la suite $(\mu_n)_{n\geq 1}$ converge étroitement vers μ si et seulement si $\mu_n(f) \to \mu(f)$ pour tout $f \in \mathcal{C}_b(E, \mathbb{R})$. L'ensemble $\mathcal{M}_f(E)$ est un cône, qui est homéomorphe à

$$(\mathcal{M}_1(E) \times \mathbb{R}_+)/\sim,$$

où l'on a noté $\sim$ la relation définie par $(\mu, r) \sim (\nu, r)$ si et seulement si $\mu = \nu$ ou $r = 0$. En effet, l'application $(\mu, r) \mapsto r\mu$ de $\mathcal{M}_1(E) \times \mathbb{R}_+$ vers $\mathcal{M}_f(E)$ passe au quotient en un tel homéomorphisme. Ainsi, on a convergence étroite de μ_n vers μ si et seulement si

$$\mu_n(1) \to \mu(1) \text{ et, si } \mu(1) > 0\,, \qquad \frac{\mu_n}{\mu_n(1)} \Rightarrow \frac{\mu}{\mu(1)}\,.$$

Il faut noter que $\mathcal{M}_f(E)$ n'est pas compact même lorsque E est compact, par exemple, la suite $(n\delta_x)_{n\geq 1}$, pour un $x \in E$ quelconque, n'a pas de sous-suite convergente. La caractérisation des compacts de $\mathcal{M}_f(E)$ est cependant très similaire au cas de $\mathcal{M}_1(E)$. Comme pour $\mathcal{M}_1(E)$, on dit que l'ensemble $\Gamma \subseteq \mathcal{M}_f(E)$ est tendu si pour tout $\varepsilon > 0$, il existe un compact $K \subseteq E$ tel que

$$\sup_{\mu \in \Gamma} \mu(E \setminus K) \leq \varepsilon\,.$$

Le théorème de Prokhorov se généralise aisément.

Théorème 1.8 *Un sous-ensemble $\Gamma \subseteq \mathcal{M}_f(E)$ est relativement compact si et seulement s'il vérifie les deux conditions suivantes*

1. *Γ est tendu*
2. *$\sup_{\mu\in\Gamma} \mu(1) < \infty$.*

Enfin, les arguments donnés dans les sections précédentes s'adaptent facilement pour montrer que $\mathcal{M}_f(E)$ est polonais. La distance de Fortet-Mourier se généralise sans changement en une distance sur $\mathcal{M}_f(E)$. En revanche, pour généraliser la distance de Lévy-Prokhorov on doit adopter une définition plus symétrique, en posant, pour $\mu, \nu \in \mathcal{M}_f(E)$,

$$d_{LP}(\mu, \nu) = \inf \left\{ r > 0 : \begin{array}{l} \mu(F) \leq \nu(F^r) + r \text{ et} \\ \nu(F) \leq \mu(F^r) + r \end{array} \quad \text{pour tout } F \subseteq E \text{ fermé} \right\}.$$

Les espaces métriques $(\mathcal{M}_f(E), d_{LP})$ et $(\mathcal{M}_f(E), d_{FM})$ sont alors complets.

1.7 La convergence en loi

La convergence en loi est la reformulation de la convergence étroite en termes de variables aléatoires. On note $\mathcal{L}(Y)$ la loi d'une variable aléatoire Y.

Définition 1.5 *Soit $Y_n, n \geq 1$ et Y des variables aléatoires à valeurs dans un espace polonais E. On dit que Y_n converge en loi vers Y si $\mathcal{L}(Y_n) \Rightarrow \mathcal{L}(Y)$.*

Noter que les variables $Y_n, n \geq 1$ et Y peuvent très bien être définies sur des espaces de probabilités tous différents ! Ceci dit, il n'est jamais restrictif de considérer que toutes les variables considérées le sont sur un même espace de probabilités (exercice). Dans la suite, sauf mention explicite de contraire, les variables aléatoires considérées seront définies sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$. La définition précédente est donc équivalente au fait que pour tout $f \in \mathcal{C}_b(E, \mathbb{R})$, on a

$$\mathbb{E}[f(Y_n)] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[f(Y)] .$$

Les différentes notions que nous avons étudiées précédemment ont une traduction en termes de variables aléatoires. On laisse au lecteur le soin de traduire, par exemple, le théorème 1.3 en termes d'une suite de variables aléatoires.

Par abus de langage, on dira parfois que la famille $(Y_i, i \in I)$ de variables aléatoires est tendue si l'ensemble $\{\mathcal{L}(Y_i), i \in I\} \subseteq \mathcal{M}_1(E)$ est tendu. Ceci signifie que pour tout $\varepsilon > 0$, il existe un compact $K_\varepsilon \subseteq E$ tel que

$$\sup_{i\in I} \mathbb{P}(Y_i \notin K_\varepsilon) \leq \varepsilon .$$

Par exemple, le théorème de Prokhorov a la conséquence suivante, dont la preuve est immédiate.

Proposition 1.16 *Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires à valeurs dans un espace polonais. Supposons que pour tout $\varepsilon > 0$, il existe un compact $K_\varepsilon \subseteq E$ tel que*

$$\sup_{n\geq 1} \mathbb{P}(Y_n \notin K_\varepsilon) \leq \varepsilon .$$

Alors il existe une sous-suite $\psi(n)$ et une variable aléatoire Y telle que $Y_{\psi(n)}$ converge en loi vers Y.

Nous donnons quelques conséquences importantes de ces résultats.

Proposition 1.17 *Soit $(Y_n)_{n\geq 1}, (Z_n)_{n\geq 1}$ deux suites de variables aléatoires à valeurs dans des espaces polonais, telles que $Y_n \Rightarrow Y$ et $Z_n \Rightarrow Z$, et telles que Y_n est indépendante de Z_n pour tout n. Alors $(Y_n, Z_n) \Rightarrow (Y', Z')$, où $\mathcal{L}((Y', Z')) = \mathcal{L}(Y) \otimes \mathcal{L}(Z)$, c'est-à-dire que (Y', Z') est un couple de variables aléatoires indépendantes de lois marginales celles de Y et Z.*

Preuve. Comme Y_n et Z_n convergent en loi, et sont à valeurs dans des espaces polonais, on en déduit que les suites $(\mathcal{L}(Y_n))_{n\geq 1}$ et $(\mathcal{L}(Z_n))_{n\geq 1}$ sont tendues. Pour tout $\varepsilon > 0$ on peut donc trouver deux compacts K, K' tels que $\mathbb{P}(Y_n \notin K) \leq \varepsilon/2$ et $\mathbb{P}(Z_n \notin K') \leq \varepsilon/2$ pour tout $n \geq 1$. On en déduit que $\mathbb{P}((Y_n, Z_n) \notin K \times K') \leq \varepsilon$ et donc que la suite $(\mathcal{L}((Y_n, Z_n)))_{n\geq 1}$ est tendue (noter que cela n'utilise pas l'indépendance). Donc de toute sous-suite, on peut réextraire une sous-sous-suite qui converge en loi.

Montrons que la limite de toute sous-suite convergente a pour loi $\mathcal{L}(Y) \otimes \mathcal{L}(Z)$, ce qui permettra de conclure. Soit donc $\psi(n)$ une extraction telle que $(Y_{\psi(n)}, Z_{\psi(n)})$ converge en loi vers une limite (Y'', Z''). Alors pour tout f, g fonctions continues bornées, on a que $\mathbb{E}[f(Y_{\psi(n)})g(Z_{\psi(n)})]$ converge vers $\mathbb{E}[f(Y'')g(Z'')]$, mais par ailleurs,

$$\mathbb{E}[f(Y_{\psi(n)})g(Z_{\psi(n)})] = \mathbb{E}[f(Y_{\psi(n)})]\mathbb{E}[g(Z_{\psi(n)})] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[f(Y)]\mathbb{E}[g(Z)]\,,$$

et on conclut que la loi de (Y'', Z'') est bien la loi produit. □

Un autre résultat important est le lemme de Slutzky. Sa démonstration suit des lignes similaires à celle du résultat précédent, et nous la laissons donc au lecteur.

Proposition 1.18 (Lemme de Slutzky) *Soit $(Y_n)_{n\geq 1}, (Z_n)_{n\geq 1}$ deux suites de variables aléatoires à valeurs dans des espaces polonais, telles que $Y_n \Rightarrow Y$ et $Z_n \to z$ en probabilité, où z est un point donné. Alors $(Y_n, Z_n) \Rightarrow (Y, z)$.*

Remarque. Si la suite $(Y_n)_{n\geq 1}$ de variables aléatoires à valeurs dans (E, d) converge en loi vers une constante y, alors comme $E \setminus B_d(y, \varepsilon)$ est un fermé, on a

$$\limsup_n \mathbb{P}(Y_n \notin B_d(y, \varepsilon)) \leq \mathbb{P}(y \notin B_d(y, \varepsilon)) = 0\,,$$

et donc Y_n converge en probabilités vers y. La réciproque est également vraie, et sa démonstration est laissée au lecteur.

Nous donnons également une propriété évidente mais utile de la convergence en loi.

Proposition 1.19 *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires à valeurs dans E qui converge en loi vers X, et soit f une fonction continue de E dans F, où F est un espace topologique. Alors $f(X_n)$ converge en loi vers $f(X)$.*

C'est évident, puisque pour toute fonction $g \in \mathcal{C}_b(F, \mathbb{R})$, $g \circ f$ est dans $\mathcal{C}_b(E, \mathbb{R})$.

1.8 Le théorème de représentation de Skorokhod

Le théorème de représentation de Skorokhod est un résultat fort utile, qui explique également pourquoi l'on préfère toujours travailler avec des variables aléatoires prenant leurs valeurs dans un espace polonais.

Théorème 1.9 (Théorème de représentation de Skorokhod) *Soit $(Y_n)_{n\geq 1}$ et Y_∞ des variables aléatoires à valeurs dans un espace polonais E. On suppose que Y_n converge en loi vers Y_∞. Alors il existe un espace de probabilités $(\Omega', \mathcal{F}', \mathbb{P}')$, et des variables aléatoires $(Z_n)_{n\geq 1}$ et Z_∞ définies sur cet espace, telles que*

1. *$\mathcal{L}(Y_n) = \mathcal{L}(Z_n)$ pour tout $n \in \mathbb{N} \cup \{\infty\}$, et*
2. *la suite $(Z_n)_{n\geq 1}$ converge presque-sûrement vers Z_∞.*

D'après ce théorème, la convergence en loi (pour des variables aléatoires à valeurs dans un espace polonais) peut donc toujours être vue comme une convergence presque-sûre dans un espace de probabilités idoine (une « représentation » des variables aléatoires considérées).

Preuve. Nous proposons une construction où l'espace $(\Omega', \mathcal{F}', \mathbb{P}')$ est $(]0,1], \mathcal{B}_{]0,1]}, \lambda)$, où λ est la mesure de Lebesgue sur $]0,1]$. Soit d une distance sur E telle que (E,d) est complet, et soit $(\gamma_i)_{i\geq 1}$ une suite dense de E. Dans cette preuve on notera $\mu_n = \mathcal{L}(Y_n)$ pour tout $n \in \mathbb{N} \cup \{\infty\}$.

Pour tout $i \geq 1$ et $m \geq 1$, soit $r(i,m) \in (1/2^{m+1}, 1/2^m)$ tel que $\mu_\infty(\partial B_d(\gamma_i, r(i,m))) = 0$. Noter qu'un tel $r(i,m)$ existe car l'ensemble $\{r > 0 : \mu_\infty(\partial B_d(\gamma_i, r)) > 0\}$ est au plus dénombrable.

On définit alors par récurrence une famille $A^m_{i_1,\ldots,i_m}$ indexée par $m \geq 1$ et un multi-indice $i_1, \ldots, i_m$ d'entiers strictement positifs, de sorte que pour tout $m \geq 1$ et $i_1, \ldots, i_m$,
- l'ensemble $A^m_{i_1,\ldots,i_m}$ est mesurable
- $\operatorname{diam}(A^m_{i_1,\ldots,i_m}) \leq 2^{-m}$
- $\mu_\infty(\partial A^m_{i_1,\ldots,i_m}) = 0$
- la famille $(A^m_{i_1,\ldots,i_{m-1},j})_{j\geq 1}$ est une partition de $A^{m-1}_{i_1,\ldots,i_{m-1}}$,

où par convention, on note $A^0_\varnothing = E$ pour que le dernier point soit bien défini, et où la définition de partition diffère un peu de la définition standard (on a ordonné les termes et autorisé des termes vides). On définit ces sous-ensembles par récurrence sur m, partant de $A^0_\varnothing = E$ et en posant

$$A^{m+1}_{i_1,\ldots,i_m,1} = A^m_{i_1,\ldots,i_m} \cap B_d(\gamma_1, r(1, m+1))\,,$$

puis par récurrence,

$$A^{m+1}_{i_1,\ldots,i_m,j+1} = A^m_{i_1,\ldots,i_m} \cap \left(B_d(\gamma_{j+1}, r(j+1, m+1)) \setminus \bigcup_{k=1}^{j} A^{m+1}_{i_1,\ldots,i_m,k} \right)\,.$$

On se convaincra facilement que $\partial A^m_{i_1,\ldots,i_m}$ est inclus dans la réunion des ensembles $\partial B_d(\gamma_i, r(i,m))$ pour $i, m \geq 1$, et est donc de μ_∞-mesure nulle. De plus, $A^m_{i_1,\ldots,i_m}$ est inclus dans $B_d(\gamma_{i_m}, r(i_m, m))$, et est donc de diamètre au plus 2^{-m}. Les deux autres propriétés sont faciles. Dans la suite, pour plus de simplicité, on notera $\mathbf{i}$ un multi-indice $i_1, \ldots, i_m$, la longueur de l'indice devant être claire selon le contexte.

Pour chaque $m \geq 1$ et $n \in \mathbb{N} \cup \{\infty\}$, définissons alors une partition de $]0,1]$, notée $(B^{m,n}_{\mathbf{i}})$ et indexée par le multi-indice $\mathbf{i} = i_1, \ldots, i_m$, de sorte que
- pour tout $m, n, \mathbf{i}$ on a $\lambda(B^{m,n}_{\mathbf{i}}) = \mu_n(A^m_{\mathbf{i}})$, et
- pour tout $m, n, \mathbf{i}, j$ on a $B^{m+1,n}_{\mathbf{i},j} \subseteq B^{m,n}_{\mathbf{i}}$.

On pose pour cela $B^{m,n}_{\mathbf{i}} =]\alpha^{m,n}_{\mathbf{i}}, \beta^{m,n}_{\mathbf{i}}]$, où

$$\alpha^{m,n}_{\mathbf{i}} = \sum_{\mathbf{j}<\mathbf{i}} \mu_n(A^m_{\mathbf{j}})\,, \qquad \beta^{m,n}_{\mathbf{i}} = \sum_{\mathbf{j}\leq\mathbf{i}} \mu_n(A^m_{\mathbf{j}})\,, \tag{5}$$

et où $\mathbf{j} \leq \mathbf{i}$ correspond à l'ordre lexicographique sur les multi-indices (de longueur m). On laisse au lecteur la vérification des propriétés avancées.

On fixe maintenant une fois pour toutes un $x^m_{\mathbf{i}}$ dans chaque $A^m_{\mathbf{i}}$, et pour tout $m \geq 1$ et $n \in \mathbb{N} \cup \{\infty\}$ on définit une variable aléatoire Z^m_n sur l'espace de probabilités $(]0,1], \mathcal{B}_{]0,1]}, \lambda)$ en posant

$$Z^m_n(y) = x^m_{\mathbf{i}}\,, \qquad \text{si} \quad y \in B^{m,n}_{\mathbf{i}}\,.$$

Pour tout $y \in]0,1]$ et $n \geq 1$ fixés, et $m,p \geq 1$, on a alors que $Z_n^m(y), Z_n^{m+p}(y)$ sont dans le même ensemble $A_{\mathbf{i}}^m$, puisque les $B_{\mathbf{i}}^{m,n}$, comme les $A_{\mathbf{i}}^m$, sont emboîtés. Par conséquent,

$$d(Z_n^m(y), Z_n^{m+p}(y)) \leq 2^{-m}\,,$$

et on déduit que la suite $(Z_n^m(y))_{m\geq 1}$ est de Cauchy. On note $Z_n(y)$ sa limite, et comme pour tout $y \in]0,1]$ on a, par passage à la limite,

$$d(Z_n^m(y), Z_n(y)) \leq 2^{-m}\,, \tag{6}$$

on en déduit que $(Z_n^m(y))_{y\in]0,1]}$ converge uniformément vers $(Z_n(y))_{y\in]0,1]}$.

Montrons que pour tout $n \in \mathbb{N}\cup\{\infty\}$, la variable aléatoire Z_n a pour loi μ_n. Pour cela, donnons-nous une fonction $f \in \mathrm{Lip}_1^d(E,\mathbb{R})$, et écrivons

$$\left|\mu_n(f) - \int_{]0,1]} f(Z_n)\mathrm{d}\lambda\right| \leq \left|\mu_n(f) - \int_{]0,1]} f(Z_n^m)\mathrm{d}\lambda\right| + \int_{]0,1]} |f(Z_n^m) - f(Z_n)|\mathrm{d}\lambda\,.$$

Par la convergence uniforme de Z_n^m vers Z_n, le second terme à droite de l'inégalité converge vers 0 lorsque $m\to\infty$. Par ailleurs, on réécrit le premier terme à l'intérieur des valeurs absolues sous la forme

$$\begin{aligned}
\sum_{\mathbf{i}}\left(\mu_n(f\mathbb{1}_{A_{\mathbf{i}}^m}) - \int_{B_{\mathbf{i}}^{m,n}} f(Z_n^m)\mathrm{d}\lambda\right) &= \sum_{\mathbf{i}}\left(\mu_n(f\mathbb{1}_{A_{\mathbf{i}}^m}) - f(x_{\mathbf{i}}^m)\lambda(B_{\mathbf{i}}^{m,n})\right)\\
&= \sum_{\mathbf{i}}\left(\mu_n(f\mathbb{1}_{A_{\mathbf{i}}^m}) - f(x_{\mathbf{i}}^m)\mu_n(A_{\mathbf{i}}^m)\right)\\
&= \sum_{\mathbf{i}}\int_{A_{\mathbf{i}}^m}(f(x) - f(x_{\mathbf{i}}^m))\mathrm{d}\mu_n(x)\,,
\end{aligned}$$

et comme f est 1-Lipschitzienne et $\mathrm{diam}\,(A_{\mathbf{i}}^m) \leq 2^{-m}$, la valeur absolue du terme précédent est bornée par

$$\sum_{\mathbf{i}}\mu_n(A_{\mathbf{i}}^m)2^{-m} = 2^{-m}\underset{m\to\infty}{\longrightarrow} 0\,.$$

On en déduit finalement que $\mu_n(f) = \int_{]0,1]} f(Z_n)\mathrm{d}\lambda$, et donc que la loi de Z_n est μ_n, par la proposition 1.5.

Montrons enfin que Z_n converge p.s. vers Z_∞. Comme $\mu_\infty(\partial A_{\mathbf{i}}^m) = 0$ pour tout $m,\mathbf{i}$, on en déduit par le 6. du théorème 1.3 que pour tout $m,\mathbf{i}$,

$$\mu_n(A_{\mathbf{i}}^m)\underset{n\to\infty}{\longrightarrow}\mu_\infty(A_{\mathbf{i}}^m)\,.$$

Le lemme de Fatou montre alors que, pour tout $m,\mathbf{i}$.

$$\alpha_{\mathbf{i}}^{m,\infty} \leq \liminf_n \alpha_{\mathbf{i}}^{m,n} \qquad \text{et} \qquad \beta_{\mathbf{i}}^{m,\infty} \leq \liminf_n \beta_{\mathbf{i}}^{m,n},. \tag{7}$$

Par ailleurs, comme $\sum_{\mathbf{j}} \mu_n(A^m_{\mathbf{j}}) = 1$, on a pour tout $m, n, \mathbf{i}$,

$$1 - \alpha^{m,n}_{\mathbf{i}} = \sum_{\mathbf{j} \geq \mathbf{i}} \mu_n(A^m_{\mathbf{j}}) \,, \qquad 1 - \beta^{m,n}_{\mathbf{i}} = \sum_{\mathbf{j} > \mathbf{i}} \mu_n(A^m_{\mathbf{j}}) \,,$$

d'où l'on tire

$$1 - \alpha^{m,\infty}_{\mathbf{i}} \leq \liminf_n (1 - \alpha^{m,n}_{\mathbf{i}}) \quad \text{et} \quad 1 - \beta^{m,\infty}_{\mathbf{i}} \leq \liminf_n (1 - \beta^{m,n}_{\mathbf{i}}) \,. \tag{8}$$

Les inégalités (7) et (8) montrent finalement que

$$\lim_n \alpha^{m,n}_{\mathbf{i}} = \alpha^{m,\infty}_{\mathbf{i}} \quad \text{et} \quad \lim_n \beta^{m,n}_{\mathbf{i}} = \beta^{m,\infty}_{\mathbf{i}} \,.$$

Soit alors y un point dans l'intérieur de $B^{m,\infty}_{\mathbf{i}}$, c'est-à-dire dans l'intervalle $]\alpha^{m,\infty}_{\mathbf{i}}, \beta^{m,\infty}_{\mathbf{i}}[$. Par ce qui précède, pour tout n assez grand, disons $n \geq N(m,y)$, on a alors que y est également dans l'intérieur de $B^{m,n}_{\mathbf{i}}$. De ce fait, par définition, on a $Z^m_n(y) = Z^m_\infty(y)$ pour tout $n \geq N(m,y)$. Notons

$$D =]0,1] \setminus \{\alpha^{m,\infty}_{\mathbf{i}}, \beta^{m,\infty}_{\mathbf{i}} : m \geq 1, \mathbf{i} \in \mathbb{N}^m\} \,.$$

Alors $\lambda(D) = 1$. Fixons $y \in D$, et donnons-nous un $\varepsilon > 0$. Soit également $m \geq 1$ tel que $2^{-m} \leq \varepsilon/2$, alors, pour tout $n \geq N(m,y)$ on a $Z^m_n(y) = Z^m_\infty(y)$ et donc

$$\begin{aligned} d(Z_n(y), Z_\infty(y)) &\leq d(Z_n(y), Z^m_n(y)) + d(Z^m_\infty(y), Z_\infty(y)) \\ &\leq 2^{-m} + 2^{-m} \leq \varepsilon \,, \end{aligned}$$

où l'on a utilisé (6). Ceci signifie bien que $Z_n(y)$ converge vers $Z_\infty(y)$ pour tout $y \in D$, c'est-à-dire que Z_n converge p.s. vers Z_∞. □

Une preuve beaucoup plus simple dans le cas particulier $E = \mathbb{R}$ est présentée en exercice plus bas. Voici une conséquence de ce résultat.

Corollaire 1.1 *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}$, qui converge en loi vers X. On suppose que les $(X_n)_{n\geq 1}$ forment une famille uniformément intégrable, c'est-à-dire que*

$$\lim_{a\to\infty} \sup_{n\geq 1} \mathbb{E}[|X_n| \mathbb{1}_{\{|X_n|\geq a\}}] = 0 \,.$$

Alors $E[X_n]$ converge vers $E[X]$.

Preuve. Par le théorème de représentation de Skorokhod, on peut supposer que la suite $(X_n)_{n\geq 1}$ converge p.s. vers sa limite X. Comme l'uniforme intégrabilité est une propriété des lois individuelles des X_n, on en déduit que faire cette supposition conserve le caractère uniformément intégrable de la suite. Or il est bien connu qu'une suite de variables aléatoires qui converge p.s. converge également dans L^1 si et seulement si elle est uniformément intégrable. □

1.9 Exercices pour le chapitre 1

Exercice 1 — Mesures atomiques, mesures diffuses
Soit $(\mu_n, n \geq 0)$ une suite de mesures de probabilités sur $\mathbb{R}$. On suppose que μ_n converge étroitement vers une mesure de probabilités μ. Y-a-t'il des implications entre les propositions suivantes ?

1. μ_n est à densité pour tout n
2. μ est à densité
3. μ_n est atomique pour tout n
4. μ est atomique

Rappelons qu'une mesure atomique est une mesure qui s'écrit $\sum a_i \delta_{b_i}$ pour des suites $a_i \in \mathbb{R}_+, b_i \in \mathbb{R}$.

Exercice 2 — Théorème de Lévy
Montrer le théorème de Lévy dans $\mathbb{R}^k$, en s'inspirant de la preuve donnée pour $k = 1$.

Exercice 3 — Exposant caractéristique
On rappelle que si $f : \mathbb{R}^d \to \mathbb{C}^*$ est une fonction continue et vérifiant $f(0) = 1$, alors il existe une unique fonction continue $g : \mathbb{R}^d \to \mathbb{C}$ telle que $g(0) = 0$ et $f = e^g$.

Si μ est une loi de probabilités dont la fonction caractéristique ϕ_μ ne s'annule pas, on note ψ_μ le relèvement de ϕ_μ nul en 0, et on l'appelle *exposant caractéristique* de μ.

(i) Soit $f_n : \mathbb{R}^d \to \mathbb{C}^*, n \geq 1$ une suite de fonctions continues avec $f_n(0) = 1$, et g_n le relèvement de f_n nul en 0. On suppose que f_n converge vers $f : \mathbb{R}^d \to \mathbb{C}^*$ uniformément sur les compacts de $\mathbb{R}^d$. Montrer que g_n converge vers g, le relèvement de f nul en 0, uniformément sur les compacts.

(ii) Soit μ une loi de probabilités sur $\mathbb{R}^d$, montrer que pour tout $K > 0$ et $u, v \in \mathbb{R}^d$,

$$|\phi_\mu(u) - \phi_\mu(v)| \leq 2\mu(\{x : |x| > K\}) + K|u - v| \,.$$

En déduire que si la suite $(\mu_n, n \geq 1)$ de lois de probabilités sur $\mathbb{R}^d$ est uniformément tendue, alors $(\phi_{\mu_n}, n \geq 1)$ est équicontinue.

(iii) En déduire que si $(\mu_n, n \geq 1)$ est une suite de lois de probabilités sur $\mathbb{R}^d$ telle que μ_n converge étroitement vers la loi μ, et si $\phi_{\mu_n}, n \geq 1$ et ϕ_μ ne s'annulent pas, alors les exposants caractéristiques ψ_{μ_n} convergent uniformément sur les compacts vers l'exposant caractéristique ψ_μ.

Exercice 4 — Transformée de Laplace
Si $\mu \in \mathcal{M}_1(\mathbb{R}_+)$, on note

$$L_\mu(t) = \int_{\mathbb{R}_+} \exp(-tx)\mu(\mathrm{d}x) \,, \qquad t \geq 0$$

sa transformée de Laplace. On justifiera que la fonction L_μ est de classe $\mathcal{C}^\infty$ sur $]0,\infty[$.

(i) Montrer que pour tout $x \geq 0$, on a

$$\mu([0,x[) + \frac{1}{2}\mu(\{x\}) = \lim_{\lambda\to\infty} \sum_{k=0}^{\lfloor \lambda x\rfloor} \frac{(-\lambda)^k}{k!} \frac{d^k L_\mu}{\mathrm{d}t^k}(\lambda)\,.$$

On pourra penser à faire intervenir la loi de Poisson. En déduire que si μ et ν sont deux mesures de probabilités sur $\mathbb{R}_+$ telles que $L_\mu = L_\nu$, alors on a $\mu = \nu$. On pourra aussi donner une autre preuve à l'aide du théorème de Stone-Weierstraß.

(ii) En s'inspirant de la preuve du théorème de Lévy, montrer l'énoncé suivant. Soit $(\mu_n, n \geq 1)$ une suite de $\mathcal{M}_1(\mathbb{R}_+)$. Si $L_{\mu_n}(t)$ converge vers $L(t)$ pour tout $t \geq 0$, où la fonction L est continue à droite en 0, alors il existe $\mu \in \mathcal{M}_1(\mathbb{R}_+)$ telle que $L = L_\mu$, et μ_n converge étroitement vers μ lorsque $n \to \infty$.

Exercice 5 — Lemme de Scheffé

Soit $(X_n, n \geq 0)$ une suite de variables aléatoires dans $\mathbb{R}^k$, dont les lois respectives $f_n(x)\mathrm{d}x$ admettent une densité par rapport à la mesure de Lebesgue. On suppose que $f_n(x) \to f(x)$ pour Lebesgue-presque-tout $x \in \mathbb{R}^k$, où $\int_{\mathbb{R}^k} f(x)\mathrm{d}x = 1$. Montrer que X_n converge en loi vers une variable aléatoire X de loi $f(x)\mathrm{d}x$. On commencera par montrer que f_n converge vers f dans $L^1(\mathbb{R}^k)$.

Exercice 6 — Norme de variation totale

Cet exercice nécessite quelques notions sur les mesures signées, rappelées dans l'énoncé. Pour plus de détails, consulter par exemple le livre de Rudin, *Real and Complex Analysis*.

Si m est une mesure signée (différence de deux mesures positives finies) sur un espace mesurable $(E, \mathcal{E})$, on note $|m|$ la mesure de variation totale associée :

$$|m|(A) = \sup \sum_{i\geq 1} |m(A_i)|\,,$$

le supremum portant sur toutes les partitions dénombrables de A en ensembles de $\mathcal{E}$. On rappelle que si m est une mesure signée, la décomposition de Jordan de m est l'écriture unique $m = m_+ - m_-$ en différence de deux mesures positives telles que $|m| = m_+ + m_-$. Elle est donnée par $m_+ = \mathbb{1}_A|m|$ et $m_- = \mathbb{1}_{A^c}|m|$ pour un ensemble mesurable A unique à un ensemble de $|m|$-mesure nulle près. On note $\|m\|_{\mathrm{VT}} = |m|(E)$, définissant une norme sur l'espace vectoriel des mesures signées sur E.

(i) Montrer que

$$\|m\|_{\mathrm{VT}} = \sup_{\|f\|_\infty \leq 1} |m(f)|\,,$$

où le sup porte sur toutes les fonctions mesurables bornés par 1 en valeur absolue.

(ii) En déduire (dans le cas où E est un espace métrique muni de sa tribu borelienne) que si l'on se donne une suite $(\mu_n, n \geq 0)$ de mesures de probabilité sur E, telle que $\mu_n \to \mu$

au sens le la variation totale (i.e. $\|\mu_n - \mu\|_{\mathrm{VT}} \to 0$), alors μ_n converge étroitement vers μ. Donner un contre-exemple pour la réciproque de cet énoncé.

(iii) On suppose que $E = \mathbb{R}^k$ est muni de la tribu borelienne et que $\mu_n(\mathrm{d}x) = f_n(x)\mathrm{d}x$ est une mesure à densité par rapport à la mesure de Lebesgue, ainsi que $\mu(\mathrm{d}x) = f(x)\mathrm{d}x$. Montrer que μ_n converge en variation totale vers μ si et seulement si f_n converge vers f dans $L^1(\mathbb{R}^k)$. Déduire que l'on a une conclusion plus forte dans l'exercice précédent.

(iv) On suppose que $E = \mathbb{Z}_+$. Montrer que si m est une mesure signée sur E, alors

$$\|m\|_{\mathrm{VT}} = \sum_{k \geq 0} |m(\{k\})| \,.$$

En déduire que la topologie étroite coïncide avec la topologie induite par $\|\cdot\|_{\mathrm{VT}}$.

(v) Montrer que si $E = \mathbb{R}$, alors l'espace $(\mathcal{M}_1(E), \|\cdot\|_{\mathrm{VT}})$ n'est pas séparable.

Exercice 7 — Principe des lois accompagnantes

Sur un même espace de probabilités, on suppose donnée une famille de variables aléatoires $(Y_{n,k}, n \geq 1, k \geq 1)$ ainsi qu'une autre suite de variables aléatoires $(X_n, n \geq 1)$, toutes étant à valeurs dans le même espace métrique (E, d), supposé séparable et complet.

On suppose que pour tout $k \geq 1$, la suite $(Y_{n,k}, n \geq 1)$ converge en loi vers une limite Y_k. Enfin, on suppose que pour tout $\varepsilon > 0$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} P(d(X_n, Y_{n,k}) > \varepsilon) = 0\,.$$

Montrer alors que Y_k converge en loi vers une variable aléatoire Y lorsque $k \to \infty$, puis que X_n converge en loi vers Y lorsque $n \to \infty$. (On pourra utiliser avec profit la distance de Lévy-Prokhorov.)

Ce résultat est très utile en pratique.

Exercice 8 — Compacité de E et de $\mathcal{M}_1(E)$

Soit (E, d) un espace métrique séparable.

(i) Montrer que l'application $x \mapsto \delta_x$ est un homéomorphisme de E sur $E_0 = \{\delta_y : y \in E\}$, où ce dernier est vu comme sous-espace de $\mathcal{M}_1(E)$ muni de la topologie étroite.

(ii) Soit $\mu_n = \delta_{x_n}$ une suite de E_0 qui converge étroitement vers μ. On suppose par l'absurde que $(x_n, n \geq 1)$ n'admet aucune sous-suite convergente. Montrer que $\{x_n, n \geq 1\}$ est un fermé, ainsi que toutes ses parties (finies ou infinies). Montrer que cela est contradictoire avec la convergence étroite de μ_n. En déduire que E_0 est un fermé de $\mathcal{M}_1(E)$.

(iii) En déduire que si $\mathcal{M}_1(E)$ est compact, alors E est compact.

Exercice 9 — Le théorème de représentation de Skorokhod pour $\mathbb{R}$

Soit F la fonction de répartition d'une variable aléatoire X à valeurs dans R. On note

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}\,.$$

(i) Montrer que F^{-1} est une fonction croissante continue à gauche de $]0,1[$ dans $\mathbb{R}$.

(ii) Montrer que $F(x) \geq u$ si et seulement si $F^{-1}(u) \leq x$. En déduire que, si U est une variable aléatoire uniforme sur $]0,1[$, alors $F^{-1}(U)$ a même loi que X.

(iii) Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}$, qui converge en loi vers X. On note F_n la fonction de répartition de X_n. Montrer que pour tout $u \in]0,1[$, on a

$$F^{-1}(u) \leq \liminf_n F_n^{-1}(u) \leq \limsup_n F_n^{-1}(u) \leq F^{-1}(u+)\,,$$

où bien sûr $F^{-1}(u+)$ est la limite à droite de F^{-1} en u. En déduire une preuve du théorème de représentation de Skorokhod, dans le cas $E = \mathbb{R}$.

Exercice 10 — Stabilité du caractère polonais

Soit $E_1, E_2, \ldots$ une suite d'espaces polonais.

(i) Montrer que le produit $E = E_1 \times E_2 \times \ldots$ (muni de la topologie produit) est également un espace polonais. On pourra munir le produit d'une métrique s'inspirant de celles intervenant dans la section 1.4.

(ii) Montrer que la famille $\Gamma \subseteq \mathcal{M}_1(E)$ est tendue si et seulement si pour tout $i \geq 1$, $\{\pi^i_*\mu : \mu \in \Gamma\} \subseteq \mathcal{M}_1(E_i)$ est tendue, où $\pi^i : E \to E_i$ est la projection canonique.

(iii) Montrer que si E est polonais, alors l'espace $E^{[0,1]}$ des fonctions de $[0,1]$ dans E (muni de la topologie produit) n'est pas polonais en général. On pourra prendre $E = \mathbb{R}$, et montrer l'existence d'une suite de fonctions qui ne converge pas ponctuellement, mais telle que de toute sous-suite on peut réextraire une sous-sous-suite qui converge ponctuellement. On en conclura que $\mathbb{R}^{[0,1]}$ n'est pas métrisable.

On verra dans le chapitre suivant que l'espace $\mathcal{C}([0,1],E)$ des fonctions continues de $[0,1]$ dans E, muni de la topologie uniforme, est polonais.

Exercice 11 — Distance de Wasserstein

On fixe un entier $d \geq 1$. Soit $\mathcal{P}_2$ l'ensemble des mesures de probabilités μ boreliennes sur $\mathbb{R}^d$, qui vérifient de plus

$$\int_{\mathbb{R}^d} |x|^2 \mu(\mathrm{d}x) < \infty\,,$$

où $|\cdot|$ est une norme quelconque sur $\mathbb{R}^d$. Si $\mu, \nu \in \mathcal{P}_2$, on note $\Pi(\mu,\nu)$ l'ensemble des mesures π boreliennes sur $\mathbb{R}^d \times \mathbb{R}^d$ dont les lois marginales par rapport à la première et seconde coordonnée sont respectivement μ, ν, c'est-à-dire que pour tout borelien $A \subseteq \mathbb{R}^d$

$$\pi(A \times \mathbb{R}^d) = \mu(A)\,, \qquad \pi(\mathbb{R}^d \times A) = \nu(A)\,.$$

On note alors

$$T_2(\mu,\nu) = \inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \pi(\mathrm{d}x,\mathrm{d}y) : \pi \in \Pi(\mu,\nu)\right\}\,.$$

1. Montrer que $T_2(\mu,\nu) < \infty$ pour tout $\mu,\nu \in \mathcal{P}_2$ (on n'oubliera pas de montrer que $\Pi(\mu,\nu)$ n'est pas vide), et que $T_2(\mu,\mu) = 0$ pour tout $\mu \in \mathcal{P}_2$.

2. Soit $\mu, \nu \in \mathcal{P}_2$.
a. Montrer que pour tout $\pi \in \Pi(\mu, \nu)$ et pour A, B deux boreliens de $\mathbb{R}^d$, on a $\pi(A \times B) \leq \mu(A) + \nu(B)$.
b. En déduire que l'ensemble $\Pi(\mu, \nu)$ est uniformément tendu dans l'ensemble $\mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}^d)$ des mesures de probabilités boreliennes sur $\mathbb{R}^d \times \mathbb{R}^d$.
c. Justifier que $\Pi(\mu, \nu)$ est fermé dans $\mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}^d)$, pour la topologie étroite.

3. On se donne toujours $\mu, \nu \in \mathcal{P}_2$.
a. Montrer que l'on peut trouver une suite $(\pi_n, n \geq 0)$ qui converge vers un élément $\pi \in \Pi(\mu, \nu)$, telle que

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \pi_n(\mathrm{d}x, \mathrm{d}y) \underset{n \to \infty}{\longrightarrow} T_2(\mu, \nu)\,.$$

b. Montrer que

$$\liminf_{n \to \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \pi_n(\mathrm{d}x, \mathrm{d}y) \geq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \pi(\mathrm{d}x, \mathrm{d}y)$$

(on pourra introduire la fonction $|x-y|^2 \wedge R$ avec $R > 0$) et en déduire que

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \pi(\mathrm{d}x, \mathrm{d}y) = T_2(\mu, \nu)\,.$$

On dira que π ainsi défini est un couplage optimal entre μ et ν.

4. Soit $(\mu_n, n \geq 0)$ une suite de $\mathcal{P}_2$, et $\mu \in \mathcal{P}_2$. On suppose désormais que $T_2(\mu_n, \mu) \to 0$.
a. Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction lipschitzienne. Montrer qu'il existe $K \in]0, \infty[$ tel que

$$|\mu_n(f) - \mu(f)| \leq K\, T_2(\mu_n, \mu)^{1/2}\,.$$

(On pourra commencer par introduire un couplage optimal entre μ_n et μ.) En déduire que μ_n converge étroitement vers μ.
b. Déduire de la question précédente que

$$\liminf_{n \to \infty} \int_{\mathbb{R}^d} |x|^2 \mu_n(\mathrm{d}x) \geq \int_{\mathbb{R}^d} |x|^2 \mu(\mathrm{d}x)\,.$$

c. Justifier que pour tout $\varepsilon > 0$ il existe $C_\varepsilon \in]0, \infty[$ tel que $|x|^2 \leq (1+\varepsilon)|y|^2 + C_\varepsilon |x-y|^2$. En déduire, en utilisant un couplage optimal, que

$$\limsup_{n \to \infty} \int_{\mathbb{R}^d} |x|^2 \mu_n(\mathrm{d}x) \leq \int_{\mathbb{R}^d} |x|^2 \mu(\mathrm{d}x)\,.$$

2 Convergence en loi des processus continus

Dans ce chapitre, on s'intéressera à la notion de convergence en loi pour des processus aléatoires continus, c'est-à-dire des variables aléatoires à valeurs dans un espace de fonctions continues indexées par un sous-intervalle I de $\mathbb{R}_+$. Comme dans le chapitre précédent, l'espace E désignera un espace polonais fixé une fois pour toutes (dans la plupart des applications discutées ici, $E = \mathbb{R}^k$), c'est l'espace dans lequel les processus considérés prennent leurs valeurs.

Dans un premier temps, on se concentrera sur le cas plus simple où I est un intervalle compact, et sans perte de généralité, on supposera $I = [0,1]$. On aura recours occasionnellement à d'autres intervalles compacts de définition que $[0,1]$, les résultats décrits plus bas pour $I = [0,1]$ se généralisant sans effort.

Le cas d'un intervalle I non-compact est plus délicat, en particulier parce qu'il faut placer une topologie idoine sur l'espace $\mathcal{C}(I,E)$ des fonctions continues de I dans E. Ceci sera discuté dans le cas $I = \mathbb{R}_+$ dans la section 2.4. Les cas plus généraux d'intervalles non-compacts sont bien sûr analogues au cas $I = \mathbb{R}_+$, et nous laissons au lecteur le soin de reformuler les résultats ci-dessous dans ces situations.

2.1 L'espace $\mathcal{C} = \mathcal{C}([0,1],E)$

2.1.1 Caractère polonais

Dans la suite, on notera $\mathcal{C}$ l'ensemble $\mathcal{C}([0,1],E)$ des fonctions de $[0,1]$ dans E qui sont continues. Un élément générique de $\mathcal{C}$ sera noté $x = (x(t), 0 \leq t \leq 1)$. On munit $\mathcal{C}$ de la distance uniforme

$$D(x,y) = \sup\{d(x(t),y(t)) : 0 \leq t \leq 1\} ,$$

où d est une distance sur E. On notera que, si D dépend du choix de d, la topologie induite par D sur $\mathcal{C}$, elle, n'en dépend pas (exercice).

Proposition 2.1 *L'espace $\mathcal{C}$ est polonais.*

Preuve. Il est facile et bien connu que $(\mathcal{C}, D)$ est complet dès que l'on a choisi pour d une distance telle que (E,d) est complet. Il reste à justifier qu'il est séparable. Soit $\hat{d}$ une distance sur E telle que $(E,\hat{d})$ soit précompact. Soit $(\hat{E},\hat{d})$ son complété, qui est compact. Nous considérons pour tout $x \in \hat{E}$ l'application

$$\psi_x : \begin{array}{ccc} \hat{E} & \longrightarrow & \mathbb{R} \\ y & \longmapsto & \hat{d}(x,y) . \end{array}$$

L'application ψ_x est évidemment 1-lipschitzienne pour tout $x \in \hat{E}$, donc continue. De plus, l'application $x \mapsto \psi_x$ est une isométrie de $\hat{E}$ dans l'espace $(\mathcal{C}(\hat{E},\mathbb{R}), \|\cdot\|_\infty)$, puisque pour tout $x,y,z \in \hat{E}$,

$$|\psi_x(z) - \psi_y(z)| \leq \hat{d}(x,y) ,$$

la borne supérieure étant atteinte pour $z = x$, ce qui implique $\|\psi_x - \psi_y\|_\infty = \hat{d}(x,y)$.

On déduit de cela que $\mathcal{C}([0,1],\hat{E})$ est isométrique à une partie de l'espace de Banach $\mathcal{C}([0,1],\mathcal{C}(\hat{E},\mathbb{R}))$ des fonctions continues de $[0,1]$ dans $\mathcal{C}(\hat{E},\mathbb{R})$. On a alors recours au résultat suivant

Lemme 2.1 *Soit $(V,\|\cdot\|)$ un espace vectoriel normé séparable. Alors l'espace $\mathcal{C}([0,1],V)$ muni de la norme $\|f\|_\infty = \sup_{0\le t\le 1}\|f(t)\|$ est séparable.*

Preuve. Soit A un sous-ensemble dénombrable dense de V. On considère l'ensemble des fonctions $g : [0,1] \to V$ telles qu'il existe un $n \geq 1$ pour lequel, pour tout $i \in \{0,1,\dots,n-1\}$, on a $g(i/n) \in A$, et de plus,

$$g(t) = (i+1-nt)g(i/n) + (nt-i)g((i+1)/n)\,, \quad i/n \le t \le (i+1)/n\,.$$

Cet ensemble est alors dénombrable, vérifions qu'il est dense dans $\mathcal{C}([0,1],V)$. Pour cela, soit $f \in \mathcal{C}([0,1],V)$. Alors f est uniformément continue sur $[0,1]$, et donc pour tout $\varepsilon > 0$ on peut trouver $n > 0$ tel que $\|f(s)-f(t)\| \le \varepsilon/2$ dès que $|s-t| \le 1/n$. Pour tout $i \in \{0,1,\dots,n\}$, choisissons alors $v_i \in A$ tel que $\|f(i/n)-v_i\| \le \varepsilon/2$, et soit g la fonction interpolatrice vérifiant $g(i/n) = v_i$, définie comme ci-dessus. Si $t \in [0,1]$, soit alors $i \in \{0,1,\dots,n-1\}$ tel que $s \in [i/n,(i+1)/n]$. On a alors

$$\begin{aligned}
\|f(t)-g(t)\| &\le \|(i+1-nt)(f(t)-g(i/n))\| + \|(nt-i)(f(t)-g((i+1)/n))\| \\
&\le (i+1-nt)(\|f(t)-f(i/n)\| + \|f(i/n)-v_i\|) \\
&\quad +(nt-i)(\|f(t)-f((i+1)/n)\| + \|f((i+1)/n)-v_{i+1}\|) \\
&\le (i+1-nt)(\varepsilon/2+\varepsilon/2) + (nt-i)(\varepsilon/2+\varepsilon/2) \le \varepsilon\,,
\end{aligned}$$

donc $\|f-g\|_\infty \le \varepsilon$, d'où le résultat. □

Nous terminons la preuve de la proposition 2.1. Comme $(\mathcal{C}(\hat{E},\mathbb{R}),\|\cdot\|_\infty)$ est séparable par le théorème de Stone-Weierstraß (voir le lemme 1.3), le lemme précédent implique que $\mathcal{C}([0,1],\mathcal{C}(\hat{E},\mathbb{R}))$ est séparable. Donc $\mathcal{C}([0,1],\hat{E})$, qui est isométrique à un de ses sous-espaces, est également séparable. Comme $\mathcal{C}([0,1],E)$ est un sous-espace de ce dernier, on en déduit qu'il est aussi séparable. □

Remarque. On peut conclure également sans utiliser le lemme 2.1, en raisonnant de la façon suivante. Nous laissons au lecteur le soin d'en compléter les détails, en exercice. On considère l'application Φ qui à une fonction

$$F : \begin{array}{ccc} [0,1] & \longrightarrow & \mathcal{C}(\hat{E},\mathbb{R}) \\ t & \longmapsto & F_t \end{array}$$

associe la fonction

$$\Phi(F) : \begin{array}{ccc} [0,1]\times\hat{E} & \longrightarrow & \mathbb{R} \\ (t,x) & \longmapsto & F_t(x) \end{array}.$$

On vérifie alors que Φ envoie $\mathcal{C}([0,1],\mathcal{C}(\hat{E},\mathbb{R}))$ dans $\mathcal{C}([0,1]\times\hat{E},\mathbb{R})$, puis que cette application réalise un homéomorphisme (et même une isométrie bijective pour des choix de distances naturels) de $\mathcal{C}([0,1],\mathcal{C}(\hat{E},\mathbb{R}))$ sur $\mathcal{C}([0,1]\times\hat{E},\mathbb{R})$. Ce dernier est séparable par le lemme 1.3.

2.1.2 Compacité dans $\mathcal{C}$

Dans l'optique de caractériser la tension pour des familles de mesures sur $\mathcal{C}$, nous rappelons la caractérisation bien connue des sous-ensembles compacts de $\mathcal{C}$, donnée par le théorème d'Arzelà-Ascoli.

Rappelons qu'une partie $\mathcal{K}$ de $\mathcal{C}$ est dite *équicontinue* si pour tout $\varepsilon > 0$, il existe un $\eta > 0$ tel que pour tout $x \in \mathcal{K}$,

$$\forall s, t \in [0, 1], |t - s| \leq \eta \implies d(x(s), x(t)) \leq \varepsilon\,.$$

Nous notons

$$\omega(x, h) = \sup\{d(x(t), x(s)) : s, t \in [0, 1], |s - t| \leq h\}$$

le module de continuité de x évalué en h. L'équicontinuité de $\mathcal{K}$ se réécrit donc plus simplement

$$\forall \varepsilon > 0, \exists \eta > 0, \sup_{x \in \mathcal{K}} \omega(x, \eta) \leq \varepsilon\,.$$

Théorème 2.1 (Arzelà-Ascoli) *Une partie $\mathcal{K}$ de $\mathcal{C}$ est relativement compacte si et seulement si*

- *pour tout $t \in [0, 1]$, l'ensemble $\{x(t) : x \in \mathcal{K}\}$ est relativement compact, et*
- *la famille $\mathcal{K}$ est équicontinue.*

Preuve. Supposons que $\mathcal{K}$ satisfasse les deux hypothèses de l'énoncé, et montrons qu'il est relativement compact. Pour cela, soit $(x_n)_{n\geq 1}$ une suite de $\mathcal{K}$. Alors pour tout $t \in [0, 1] \cap \mathbb{Q}$, la suite $(x_n(t))_{n\geq 1}$ est à valeurs dans un compact par hypothèse. Par conséquent, par extraction diagonale, on peut trouver une extraction $(\psi(n))$ telle que $x_{\psi(n)}(t)$ converge vers une limite $x(t)$ pour tout $t \in [0, 1] \cap \mathbb{Q}$.

La fonction x ainsi définie est alors uniformément continue sur $[0, 1] \cap \mathbb{Q}$, car, si η est tel que $\omega(y, \eta) \leq \varepsilon$ pour tout $y \in \mathcal{K}$, alors pour tout $s, t \in [0, 1] \cap \mathbb{Q}$ tels que $|t - s| \leq \eta$, on a $d(x_{\psi(n)}(s), x_{\psi(n)}(t)) \leq \varepsilon$ et donc $d(x(s), x(t)) \leq \varepsilon$. Comme E est complet, on en déduit que x s'étend d'une unique façon en une fonction de $\mathcal{C}$, toujours notée x.

Il reste à montrer que x est limite uniforme de $y_n = x_{\psi(n)}$. Pour cela, soit $\varepsilon > 0$ fixé, et choisissons $\eta > 0$ tel que $\sup_{y\in\mathcal{K}} \omega(y, \eta) \leq \varepsilon/3$. Soit $N = \lfloor \eta^{-1} \rfloor + 1$. Enfin, choisissons n_0 tel que pour tout $n \geq n_0$ on ait

$$\max_{0\leq i\leq N} d(y_n(i/N), x(i/N)) \leq \varepsilon/3\,,$$

ce qui est possible puisque i/N est un rationnel de $[0, 1]$ pour tout $i \in \{0, 1, \ldots, N\}$. On a alors, pour tout $n \geq n_0$,

$$\begin{aligned} &d(y_n(t), x(t)) \\ &\quad\leq\ d(y_n(t), y_n(\lfloor Nt \rfloor/N)) + d(y_n(\lfloor Nt \rfloor/N), x(\lfloor Nt \rfloor/N)) + d(x(\lfloor Nt \rfloor/N), x(t)) \\ &\quad\leq\ \varepsilon/3 + \varepsilon/3 + \varepsilon/3\,, \end{aligned}$$

d'où le résultat.

Réciproquement, supposons que $\mathcal{K}$ est relativement compact. Quitte à considérer l'adhérence, on peut supposer que $\mathcal{K}$ est en fait compact. Alors l'image de $\mathcal{K}$ par la projection π^t, continue, est un compact de E, ce qui donne le premier point. Supposons enfin que $\mathcal{K}$ n'est pas équicontinue, et donc qu'il existe un $\varepsilon > 0$ tel que pour tout $n \geq 1$, on puisse trouver x_n dans $\mathcal{K}$ et s_n, t_n avec $|s_n - t_n| \leq 1/n$ mais $d(x_n(s_n), x_n(t_n)) > \varepsilon$. On voit alors facilement que x_n n'admet aucune sous-suite qui converge uniformément. Donc $\mathcal{K}$ n'est pas compact. $\square$

En fait, quitte à adapter légèrement la preuve précédente, en extrayant diagonalement non plus pour t dans $[0,1] \cap \mathbb{Q}$ mais dans un ensemble dénombrable dense quelconque, on montre que l'on peut améliorer l'énoncé précédent, de la façon suivante.

Corollaire 2.1 *Une partie $\mathcal{K}$ de $\mathcal{C}$ est relativement compacte si et seulement si*
- *il existe un sous-ensemble Q dénombrable de $[0,1]$, dense dans ce dernier, tel que pour tout $t \in Q$, l'ensemble $\{x(t) : x \in \mathcal{K}\}$ est relativement compact, et*
- *la famille $\mathcal{K}$ est équicontinue.*

Par ailleurs, dans le cas d'un espace où les boules fermées sont compactes, comme $\mathbb{R}$ ou $\mathbb{R}^k$, on a l'amélioration suivante.

Corollaire 2.2 *Soit E un espace (polonais) dont les fermés bornés sont compacts. Alors une partie $\mathcal{K}$ de $\mathcal{C}$ est relativement compacte si et seulement si*
- *l'ensemble $\{x(0) : x \in \mathcal{K}\}$ est relativement compact, et*
- *la famille $\mathcal{K}$ est équicontinue.*

Preuve. Par équicontinuité, il existe $\eta > 0$ tel que $\omega(x, \eta) \leq 1$ pour tout $x \in \mathcal{K}$, et donc

$$d(x(0), x(t)) \leq \sum_{i=1}^{\lfloor t/\eta \rfloor} d(x((i-1)\eta), x(i\eta)) + d(x(\lfloor t/\eta \rfloor \eta), x(t)) \leq \lfloor 1/\eta \rfloor + 1 , \quad t \in [0,1] ,$$

d'où l'on déduit, partant du fait que $\{x(0) : x \in \mathcal{K}\}$ est borné, que $\{x(t) : x \in \mathcal{K}\}$ est également borné pour tout $t \in [0,1]$, et donc relativement compact. $\square$

2.2 Convergence en loi dans $\mathcal{C}$

2.2.1 Tribu borélienne, tribu produit

Pour appliquer les résultats du chapitre précédents à l'espace polonais $\mathcal{C}$, il faut que cet espace soit muni de la tribu borélienne $\mathcal{B}_{\mathcal{C}}$. Néanmoins, il existe sur $\mathcal{C}$ une autre tribu, également naturelle. En effet, on peut voir $\mathcal{C}$ comme sous-ensemble de $E^{[0,1]}$, ensemble des fonctions de $[0,1]$ dans E, et cet espace est naturellement muni de la tribu produit $\mathcal{B}'_{\mathcal{C}}$, définie comme la plus petite tribu rendant mesurables les applications de projection

$$\pi^t : \begin{array}{ccc} \mathcal{C} & \longrightarrow & E \\ x & \longmapsto & x(t) \end{array} , \qquad t \in [0,1] .$$

Fort heureusement, l'existence de ces deux tribus ne pose pas de problème grâce au résultat suivant.

Proposition 2.2 *Les tribus $\mathcal{B}_{\mathcal{C}}$ et $\mathcal{B}'_{\mathcal{C}}$ coïncident.*

Preuve. Tout d'abord, on note que l'application π^t est continue de $(\mathcal{C}, D)$ dans E, et par conséquent, elle est mesurable pour la tribu borélienne $\mathcal{B}_{\mathcal{C}}$. On en déduit que $\mathcal{B}'_{\mathcal{C}} \subseteq \mathcal{B}_{\mathcal{C}}$.

Réciproquement, montrons que la boule fermée $B^f_D(x,r)$ est dans $\mathcal{B}'_{\mathcal{C}}$ pour tout $x \in \mathcal{C}$ et $r > 0$. En effet, par continuité,

$$\begin{aligned} B^f_D(x,r) &= \{y \in \mathcal{C} : d(x(t), y(t)) \leq r \text{ pour tout } t \in [0,1]\} \\ &= \{y \in \mathcal{C} : d(x(t), y(t)) \leq r \text{ pour tout } t \in [0,1] \cap \mathbb{Q}\} \\ &= \bigcap_{t \in [0,1] \cap Q} (\pi^t)^{-1}(B^f_d(x(t), r)) \,. \end{aligned}$$

Comme l'intersection est dénombrable, on voit bien que $B^f_D(x,r) \in \mathcal{B}'_{\mathcal{C}}$. De ce fait, toute boule ouverte $B_D(x,r)$ est également dans $\mathcal{B}'_{\mathcal{C}}$, puisque c'est la réunion des boules fermées $B^f_D(x, r(1-1/n))$ pour $n \geq 1$.

Enfin, comme l'espace $\mathcal{C}$ est séparable, tout ouvert est réunion dénombrable de boules ouvertes, et est donc dans $\mathcal{B}'_{\mathcal{C}}$. On en déduit que $\mathcal{B}'_{\mathcal{C}}$ contient $\mathcal{B}_{\mathcal{C}}$. □

On en déduit immédiatement une propriété fondamentale des mesures de probabilités sur $\mathcal{C}$. Si $\mu \in \mathcal{M}_1(\mathcal{C})$, et si $0 \leq t_1 < \ldots < t_k \leq 1$, on note

$$\mu^{t_1,\ldots,t_k} = (\pi^{t_1}, \ldots, \pi^{t_k})_* \mu \,,$$

ce qui définit une mesure sur E^k. Cette mesure est aussi la loi de $(X(t_1), \ldots, X(t_k))$, où $X = (X(t), 0 \leq t \leq 1)$ est une variable aléatoire à valeurs dans $\mathcal{C}$, de loi μ.

Définition 2.1 *Les mesures $\{\mu^{t_1,\ldots,t_k} : 0 \leq t_1 < \ldots < t_k \leq 1)$ sont appelées les* marginales de dimension k *de la mesure $\mu \in \mathcal{M}_1(\mathcal{C})$ (ou* lois marginales de dimension k *d'une variable aléatoire X de loi μ). La famille formée de toutes les marginales de dimension k, pour tout $k \geq 1$, est appelée famille des* marginales de dimension finie *de μ.*

On a alors le résultat suivant.

Proposition 2.3 *La famille des marginales de dimension finie d'une mesure $\mu \in \mathcal{M}_1(\mathcal{C})$ caractérise cette dernière.*

Preuve. Soient μ, μ' deux mesures ayant les mêmes marginales de dimensions finies. Alors on a $\mu(A) = \mu'(A)$ pour tout ensemble A de la forme

$$A = \{x \in \mathcal{C} : x(t_1) \in A_1, \ldots, x(t_k) \in A_k\} \,, \qquad 0 \leq t_1 < \ldots < t_k \leq 1 \,,\ A_1, \ldots, A_k \in \mathcal{B}_E \,.$$

Ces ensembles forment une classe stable par intersection finie, et qui engendre la tribu produit, donc la tribu borélienne $\mathcal{B}_{\mathcal{C}}$ par la proposition 2.2. On en déduit que $\mu = \mu'$ par le théorème de classe monotone. □

2.2.2 Un critère général de convergence étroite dans $\mathcal{M}_1(\mathcal{C})$

Nous nous intéressons maintenant à la notion de convergence étroite dans $\mathcal{M}_1(\mathcal{C})$. Il est clairement difficile de montrer la convergence étroite d'une suite $(\mu_n)_{n\geq 1}$ de $\mathcal{M}_1(\mathcal{C})$ en revenant à la définition, c'est-à-dire en vérifiant que $\mu_n(f)$ converge pour toute fonction f dans une classe de fonctions assez grande. Par ailleurs, les considérations qui précèdent suggèrent d'introduire un nouveau mode de convergence.

Définition 2.2 *Soit $(\mu_n)_{n\geq 1}$ une suite de $\mathcal{M}_1(\mathcal{C})$, et $\mu \in \mathcal{M}_1(\mathcal{C})$. On dit que μ_n converge vers μ au sens des marginales de dimension finie si pour tout $0 \leq t_1 < t_2 < \ldots < t_k \leq 1$ on a que $\mu_n^{t_1,\ldots,t_k}$ converge étroitement vers $\mu^{t_1,\ldots,t_k}$.*

De même, une suite $(X_n)_{n\geq 1}$ de variables aléatoires à valeurs dans $\mathcal{C}$ converge vers X au sens des marginales de dimension finie si pour tout $0 \leq t_1 < t_2 < \ldots < t_k$, $(X_n(t_1), \ldots, X_n(t_k))$ converge en loi vers $(X(t_1), \ldots, X(t_k))$.

La convergence au sens des marginales de dimension finie est plus facile à vérifier que la convergence étroite sur $\mathcal{M}_1(\mathcal{C})$, car c'est évidemment une notion plus faible.

Proposition 2.4 *Si une suite $(\mu_n)_{n\geq 1}$ de $\mathcal{M}_1(\mathcal{C})$ converge étroitement, elle converge également au sens des marginales de dimension finie.*

Ceci est évident, puisque les projections π^t sont continues. En revanche, la réciproque est fausse. Pour s'en convaincre, soit U une variable aléatoire uniforme sur $[0,1]$, et soit $X_n(t) = \max(0, 1-n|U-t|)$. Comme pour tout t fixé, on a que p.s. $U \neq t$, on voit que X_n converge au sens des marginales de dimension finie vers le processus identiquement égal à 0. Néanmoins, $\|X_n\|_\infty$ vaut 1 p.s. et ne converge donc pas vers 0, d'où l'on conclut que X_n ne converge pas en loi (puisque sa seule limite potentielle serait alors 0). Néanmoins, nous avons le résultat suivant.

Proposition 2.5 *Soit $(\mu_n)_{n\geq 1}$ une suite de $\mathcal{M}_1(\mathcal{C})$. Alors μ_n converge étroitement vers μ si et seulement si*

- *μ_n converge vers μ au sens des marginales de dimension finie, et*
- *la suite $(\mu_n)_{n\geq 1}$ est tendue.*

Preuve. Un sens est facile : si μ_n converge vers μ étroitement alors $\{\mu_n, n \geq 1\} \cup \{\mu\}$ est compact et donc tendu par le théorème de Prokhorov (car $(\mathcal{C}, D)$ est complet), et de plus μ_n converge vers μ au sens des marginales de dimension finie, par la proposition 2.4.

Réciproquement, supposons que $(\mu_n)_{n\geq 1}$ est tendue et converge vers μ au sens des marginales de dimension finie. Alors l'ensemble $\{\mu_n, n \geq 1\}$ est relativement compact par le théorème de Prokhorov. Par ailleurs, si $(\mu_{\psi(n)})_{n\geq 1}$ et une sous-suite convergente vers une limite μ', alors μ' est également limite de $\mu_{\psi(n)}$ au sens des marginales de dimension finie. Donc μ' et μ ont les mêmes marginales de dimension finie, et sont donc égales par la proposition 2.3. On en déduit que μ est la limite de μ_n par la proposition 1.10. □

Cette proposition est extrêmement utile, et nous incite à trouver des critères pour qu'une suite de $\mathcal{M}_1(\mathcal{C})$ soit tendue. C'est l'objectif de ce qui suit.

2.3 Relative compacité dans $\mathcal{M}_1(\mathcal{C})$

Nous voulons maintenant déterminer des conditions pratiques impliquant qu'un sous-ensemble de $\mathcal{M}_1(\mathcal{C})$ soit tendu. Voici un premier critère de tension pour les processus continus.

Proposition 2.6 *Le sous-ensemble $\Gamma \subset \mathcal{M}_1(\mathcal{C})$ est tendu si et seulement si*
- *pour tout $t \in [0,1]$, la famille $\{\pi_*^t \mu : \mu \in \Gamma\}$ est tendue, et*
- *pour tout $\varepsilon > 0, \eta > 0$, il existe un $\delta > 0$ tel que*

$$\sup_{\mu \in \Gamma} \mu(\{x : \omega(x,\delta) \geq \eta\}) \leq \varepsilon \,.$$

Si E est un espace dont les fermés bornés sont compacts, alors on peut changer la première hypothèse par le fait que la famille $\{\pi_*^0 \mu : \mu \in \Gamma\}$ est tendue.

Preuve. Fixons $\varepsilon > 0$. Soit $\{t_1, t_2, \ldots\}$ une énumération des rationnels de $[0,1]$. Par hypothèse, pour tout $i \geq 1$, on peut trouver un compact K_i de E tel que, pour tout $\mu \in \Gamma$,

$$\pi_*^{t_i} \mu(E \setminus K_i) \leq \frac{\varepsilon}{2^{i+1}} \,.$$

Par ailleurs, on peut également trouver, pour tout $k \geq 1$, un $\delta_k > 0$ tel que pour tout $\mu \in \Gamma$,

$$\sup_{\mu \in \Gamma} \mu(\{x : \omega(x,\delta_k) > 2^{-k}\}) \leq \frac{\varepsilon}{2^k} \,.$$

On définit alors

$$\mathcal{K}_\varepsilon = \bigcap_{i \geq 1} (\pi^{t_i})^{-1}(K_i) \cap \bigcap_{k \geq 1} \{x : \omega(x,\delta_k) \leq 2^{-k}\} \,.$$

Le corollaire 2.1 montre que $\mathcal{K}_\varepsilon$ est relativement compacte, et en fait compacte car elle est clairement fermée. De plus, par définition des K_i et des δ_k, on a, pour tout $\mu \in \Gamma$,

$$\mu(\mathcal{C} \setminus \mathcal{K}_\varepsilon) \leq \sum_{i \geq 1} \frac{\varepsilon}{2^{i+1}} + \sum_{k \geq 1} \frac{\varepsilon}{2^{k+1}} = \varepsilon \,.$$

On en déduit que Γ est tendu. On laisse au lecteur le soin de vérifier le cas où E est tel que ses fermés bornés sont compacts.

La réciproque est aisée, et également laissée au lecteur. $\square$

On peut immédiatement traduire la proposition précédente pour caractériser la tension d'une suite de processus aléatoires continus. À partir de maintenant, c'est essentiellement dans ce langage que nous allons raisonner.

Corollaire 2.3 *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans $\mathcal{C}$. Cette suite est uniformément tendue si et seulement si*
- *pour tout $t \in [0,1]$, la suite $(X_n(t))_{n \geq 1}$ est tendue, et*

– *pour tout* $\varepsilon > 0, \eta > 0$*, il existe un* $\delta > 0$ *tel que*

$$\limsup_{n\to\infty} \mathbb{P}(\omega(X_n,\delta) \geq \eta) \leq \varepsilon\,.$$

Si E *est un espace dont les fermés bornés sont compacts, alors on peut changer la première hypothèse par le fait que la suite* $(X_n(0))_{n\geq 1}$ *est tendue.*

Preuve. La seule chose à prouver est qu'il suffit bien de considérer la limite supérieure dans le second point, là où l'on aurait attendu un supremum en n.

Soit $\varepsilon, \eta > 0$, par hypothèse il existe un δ tel que

$$\limsup_{n\to\infty} \mathbb{P}(\omega(X_n,\delta) \geq \eta) \leq \varepsilon/2\,,$$

et on en déduit qu'il existe n_0 tel que $\sup_{n\geq n_0} \mathbb{P}(\omega(X_n,\delta) \geq \eta) \leq \varepsilon$. Par ailleurs, quitte à choisir δ plus petit, on peut supposer que l'on a également

$$\max_{1\leq n\leq n_0} \mathbb{P}(\omega(X_n,\delta) \geq \eta) \leq \varepsilon\,.$$

On en déduit que $\sup_{n\geq 1} \mathbb{P}(\omega(X_n,\delta) \geq \eta) \leq \varepsilon$, et donc le résultat. □

Le problème des caractérisations précédentes de la tension est qu'elles font intervenir les modules de continuité $\omega(X_n,\delta)$, qui sont typiquement des variables aléatoires malaisées à étudier. L'obtention de critères de tension utiles passe donc par des estimées sur le module de continuité $\omega(X_n,\delta)$ qui font intervenir la loi de X_n de la façon la moins compliquée possible, par exemple, une fonction de ses marginales de dimension finie. Un premier résultat en ce sens est le suivant.

Proposition 2.7 *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires à valeurs dans* $\mathcal{C}$*. Alors* $(X_n)_{n\geq 1}$ *est tendue si*

– *pour tout* $t \in [0,1]$*, la suite* $(X_n(t))_{n\geq 1}$ *est tendue, et*
– *pour tout* $\varepsilon > 0, \eta > 0$*, il existe un* $k \in \mathbb{N}$ *tel que*

$$\limsup_{n\to\infty} k \max_{0\leq i\leq k-1} \mathbb{P}\left(\sup_{0\leq r\leq 1/k} d(X_n(r+i/k), X_n(i/k)) \geq \eta\right) \leq \varepsilon\,.$$

Si E *est un espace dont les fermés bornés sont compacts, alors on peut changer la première hypothèse par le fait que la suite* $(X_n(0))_{n\geq 1}$ *est tendue.*

Preuve. Nous remarquons d'abord que pour tout $x \in \mathcal{C}$ et $k \in \mathbb{N}$,

$$\omega(x, 1/k) \leq 3 \max_{0\leq i\leq k-1} \sup_{r\in[0,1/k]} d(x(r+i/k), x(i/k))\,.$$

En effet, soit $s,t \in [0,1]$ tels que $|t-s| \leq 1/k$. Supposons sans perte de généralité que $s \leq t$. On a alors deux possibilités.

Ou bien il existe $i \in \{0, 1, \dots, k-1\}$ tel que $s, t \in [i/k, (i+1)/k]$. Dans ce cas, on a

$$\begin{aligned} d(x(s), x(t)) &\leq d(x(t), x(i/k)) + d(x(s), x(i/k)) \\ &\leq 2 \max_{0 \leq i \leq k-1} \sup_{r \in [0,1/k]} d(x(r + i/k), x(i/k)) \,. \end{aligned}$$

Ou bien il existe $i \in \{0, 1, \dots, k-2\}$ tel que $s \in [i/k, (i+1)/k]$ et $t \in [(i+1)/k, (i+2)/k]$. Dans ce cas,

$$\begin{aligned} d(x(s), x(t)) &\leq d(x(s), x(i/k)) + d(x(i/k), x((i+1)/k)) + d(x(t), x((i+1)/k)) \\ &\leq 3 \max_{0 \leq i \leq k-1} \sup_{r \in [0,1/k]} d(x(r + i/k), x(i/k)) \,. \end{aligned}$$

Ceci donne la majoration demandée sur $\omega(x, 1/k)$. Nous en déduisons que, pour tout $n \geq 1$, et $\eta > 0$,

$$\begin{aligned} \mathbb{P}(\omega(X_n, \delta) \geq \eta) &\leq \mathbb{P}\left(3 \max_{0 \leq i \leq k-1} \sup_{r \in [0,1/k]} d(x(r + i/k), x(i/k)) \geq \eta\right) \\ &\leq k \max_{0 \leq i \leq k-1} \mathbb{P}\left(\sup_{r \in [0,1/k]} d(x(r + i/k), x(i/k)) \geq \eta/3\right) \,. \end{aligned}$$

Sous nos hypothèses, pour tout $\varepsilon > 0$, on en déduit l'existence de k tel que la limite supérieure en n du majorant est plus petite que ε. Les hypothèses du corollaire 2.3 sont donc vérifiées, ce qui conclut la preuve. □

Un autre critère très important est dû à Kolmogorov : il ne fait intervenir que les lois marginales de dimension 2 de X_n (!) Dans l'énoncé suivant, la quantité

$$N_\alpha(x) = \sup_{h>0} \frac{\omega(x, h)}{h^\alpha} = \sup_{s,t \in [0,1], s \neq t} \frac{d(x(s), x(t))}{|t - s|^\alpha} \,,$$

est appelée la *norme de Hölder d'exposant* α de x.

Proposition 2.8 (Critère de Kolmogorov) *On considère une suite $(X_n)_{n \geq 1}$ de variables aléatoires à valeurs dans $\mathcal{C}$. On suppose*

- *que les suites $(X_n(t))_{n \geq 1}$ sont tendues pour tout $t \in [0, 1]$, et*
- *qu'il existe $\beta, p, C > 0$ tel que*

$$\sup_{n \geq 1} \mathbb{E}[d(X_n(s), X_n(t))^p] \leq C|t - s|^{1+\beta} \,, \qquad \textit{pour tout } s, t \in [0, 1] \,.$$

Alors $(X_n)_{n \geq 1}$ est tendue. Plus précisément, pour tout $\alpha \in]0, \beta/p[$ fixé et pour tout $\varepsilon > 0$, il existe $M > 0$ tel que

$$\sup_{n \geq 1} \mathbb{P}\left(N_\alpha(X_n) \geq M\right) \leq \varepsilon \,. \tag{9}$$

Pour montrer ce résultat, nous avons recours à un lemme intermédiaire. Notons $Q_k = \{i/2^k : 0 \leq i \leq 2^k\}$ les rationnels dyadiques de niveau k dans $[0, 1]$.

Lemme 2.2 *Soit $x : [0,1] \to E$ une fonction. Supposons qu'il existe $K, \alpha > 0$ pour lesquels, pour tout $k \geq 0$ et tout $s, t \in Q_k$* consécutifs*, on ait $d(x(s), x(t)) \leq K|t-s|^\alpha$. Alors pour tout $s, t \in \bigcup_{n\geq 0} Q_n$ on a*

$$d(x(s), x(t)) \leq \frac{2K}{1-2^{-\alpha}}|t-s|^\alpha .$$

En particulier, si $x \in \mathcal{C}$, on a

$$N_\alpha(x) \leq \frac{2K}{1-2^{-\alpha}} .$$

Preuve. Soit s, t deux nombres dans $\bigcup_{k\geq 0} Q_k$. Alors il existe un unique entier $r \geq 0$ tel que $2^{-r-1} < |t-s| \leq 2^{-r}$. Sans perte de généralité, on peut supposer que $s \leq t$. On distingue deux cas.

Dans un premier cas, il existe $i \in \{0, 1, \ldots, 2^r - 1\}$ tel que $i2^{-r} \leq s, t \leq (i+1)2^{-r}$. Dans ce cas, les écritures dyadiques de s et t sont de la forme

$$s = \frac{i}{2^r} + \sum_{k\geq r+1} \frac{i_k(s)}{2^k}, \qquad t = \frac{i}{2^r} + \sum_{k\geq r+1} \frac{i_k(t)}{2^k}$$

où les suites $(i_k(s))_{k\geq r+1}$ et $(i_k(t)_{k\geq r+1})$ sont à valeurs dans $\{0,1\}$, et nulles à partir d'un certain rang. On a alors, en posant

$$s_l = \frac{i}{2^r} + \sum_{k=r+1}^{r+l} \frac{i_k(s)}{2^k},$$

que s_l et s_{l+1} sont ou bien des nombres dyadiques consécutifs dans Q_{r+l+1}, ou bien égaux (ce qu'ils sont pour tout l assez grand, puisqu'alors $s_l = s$). On a donc par hypothèse

$$\begin{aligned} d(x(s), x(i2^{-r})) &\leq \sum_{l\geq 0} d(x(s_l, s_{l+1})) \leq \sum_{l\geq 0} K 2^{-(l+r+1)\alpha} \\ &= \frac{K2^{-(r+1)\alpha}}{1-2^{-\alpha}} \leq \frac{K}{1-2^{-\alpha}}|t-s|^\alpha \end{aligned}$$

La même majoration est vraie pour t à la place de s, et on déduit que

$$d(x(s), x(t)) \leq \frac{2K}{1-2^{-\alpha}}|t-s|^\alpha .$$

Dans un second cas, il existe $i \in \{1, 2, \ldots, 2^r - 1\}$ tel que $(i-1)2^{-r} \leq s \leq i2^{-r} \leq t \leq (i+1)2^{-r}$. On écrit alors plutôt

$$s = \frac{i}{2^r} - \sum_{k\geq r+1} \frac{i'_k(s)}{2^k}, \qquad t = \frac{i}{2^r} + \sum_{k\geq r+1} \frac{i_k(t)}{2^k},$$

où de la même manière que précédemment, $(i'_k(s))_{k\geq r+1}$ est une suite de $\{0,1\}$ nulle à partir d'un certain rang. Le même raisonnement que ci-dessus permet alors de conclure à la même majoration pour $d(x(s), x(t))$.

Comme l'ensemble $\bigcup_n Q_n$ est dense dans $[0,1]$, si $x \in \mathcal{C}$, nous en déduisons que la majoration reste valable pour tout $s,t \in [0,1]$. Ceci termine la preuve du lemme. □

Nous prouvons à présent la proposition 2.8. Posons

$$Z_{n,k} = \max_{0 \leq i \leq 2^k - 1} d(X_n((i+1)2^{-k}), X_n(i2^{-k}))\,.$$

On a alors, par inégalité de Markov,

$$\begin{aligned}\mathbb{P}(Z_{n,k} \geq K2^{-k\alpha}) &\leq 2^k \max_{0 \leq i \leq 2^k-1} \mathbb{P}(d(X_n((i+1)2^{-k}), X_n(i2^{-k})) \geq K2^{-k\alpha}) \\ &\leq 2^k \max_{0 \leq i \leq 2^k-1} \frac{\mathbb{E}[d(X_n((i+1)2^{-k}), X_n(i2^{-k}))^p]}{K^p 2^{-pk\alpha}}\end{aligned}$$

Par hypothèse, cette dernière quantité est majorée par $CK^{-p}2^{-(\beta - p\alpha)k}$. Si l'on choisit $\alpha \in]0, \beta/p[$, on en déduit que ces majorants sont sommables sur k. Le lemme 2.2 implique que

$$\begin{aligned}\mathbb{P}\left(N_\alpha(X_n) > \frac{2K}{1-2^{-\alpha}}\right) &\leq \mathbb{P}(\exists k \geq 0 : Z_{n,k} > K2^{-k\alpha}) \\ &\leq \frac{C}{K^p} \sum_{k \geq 0} 2^{-(\beta - p\alpha)k} \leq \frac{C'}{K^p}\,,\end{aligned}$$

pour une constante $C' \in]0,\infty[$, qui ne dépend que de α, β, p, C mais non de n. Le dernier majorant converge vers 0 lorsque $K \to \infty$, d'où l'on déduit que pour tout $\varepsilon > 0$, il existe $M > 0$ tel que (9) est vraie. Comme enfin on a $\omega(x,h) \leq N_\alpha(x)h^\alpha$ pour tout $x \in \mathcal{C}$ et $h > 0$, on déduit que

$$\mathbb{P}(\omega(X^n, \delta) \geq \eta) \leq \mathbb{P}(N_\alpha(X_n) \geq \eta/\delta^\alpha)\,,$$

ce qui, conjointement à la première hypothèse de la proposition 2.8, implique la tension de $(X_n)_{n \geq 1}$ par la proposition 2.7. □

2.4 L'espace $\mathcal{C}(\mathbb{R}_+, E)$

En pratique, les processus aléatoires considérés en probabilités sont souvent indicés par les réels positifs plutôt que par $[0,1]$. Il faut cependant être un peu prudent dans la manipulation de l'espace $\mathcal{C}(\mathbb{R}_+, E)$. Par exemple, muni de la distance uniforme $D'(x,y) = \sup_{t \geq 0} d(x(t), y(t))$, pour d une distance sur E, l'espace topologique $\mathcal{C}(\mathbb{R}_+, E)$ n'est pas séparable, et donc pas polonais.

La topologie que l'on placera *toujours* sur $\mathcal{C}(\mathbb{R}_+, E)$ est donc une autre topologie, dite compact-ouverte, ou «topologie de la convergence uniforme sur les compacts». C'est la topologie dont une base de voisinage est donnée par les ensembles

$$\{x \in \mathcal{C}(\mathbb{R}_+, E) : x(K) \subset O\}\,,$$

où K est un compact de $\mathbb{R}_+$ et O est un ouvert de E. C'est la topologie induite, par exemple, par la distance

$$D(x,y) = \sum_{n\geq 1} \frac{\sup_{0\leq t\leq n} d(x(t),y(t)) \wedge 1}{2^n}\,,$$

où d est une distance sur E. On voit qu'une suite de fonctions converge au sens compact-ouvert si et seulement si elle converge uniformément sur les compacts.

Il est alors facile de voir que $\mathcal{C}(\mathbb{R}_+, E)$ est un espace polonais. La notion de marginales de dimension finie se généralise pour cet espace, et ces lois marginales caractérisent les mesures de $\mathcal{M}_1(\mathcal{C}(\mathbb{R}_+, E))$. En effet, la tribu borélienne $\mathcal{B}_{\mathcal{C}(\mathbb{R}_+,E)}$ coïncide ici encore avec la trace sur $\mathcal{C}(\mathbb{R}_+, E)$ de la tribu produit de $E^{\mathbb{R}_+}$.

L'avantage important de cet espace est que la relative compacité est une propriété «locale», c'est-à-dire qu'il suffit de la vérifier en restriction à des intervalles compacts arbitraires. Plus précisément, on laisse au lecteur le soin de montrer les résultats suivants.

Proposition 2.9 *Une partie $\mathcal{K}$ de $\mathcal{C}(\mathbb{R}_+, E)$ est relativement compacte si et seulement si les parties $\mathcal{K}_N = \{x|_{[0,N]} : x \in \mathcal{K}\}$ sont relativement compactes dans $\mathcal{C}([0,N], E)$ pour tout $N \geq 1$, où $x|_F$ est la restriction de x à l'ensemble $F \subseteq \mathbb{R}_+$.*

Proposition 2.10 *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires à valeurs dans $\mathcal{C}(\mathbb{R}_+, E)$. La suite $(X_n)_{n\geq 1}$ est tendue si et seulement si, pour tout $N \geq 1$, la suite des restrictions $((X_n)|_{[0,N]})_{n\geq 1}$, vues comme variables aléatoires à valeurs dans $\mathcal{C}([0,N], E)$, est tendue.*

Ainsi, la convergence en loi dans $\mathcal{C}(\mathbb{R}_+, E)$ se ramène par restriction à la convergence en loi dans les espaces $\mathcal{C}(I, E)$, où I est un intervalle compact, et on peut donc appliquer les résultats des sections précédentes pour caractériser cette dernière (tension et identification des lois marginales de dimension finie).

2.5 Le théorème de Donsker

Nous en arrivons au résultat central de la première partie du cours, le théorème de Donsker, déjà mentionné dans l'introduction. Fixons quelques notations. Soit $\xi_1, \xi_2, \ldots$ une suite i.i.d. de variables aléatoires à valeurs dans $\mathbb{R}^d$, pour un entier $d \geq 1$ fixé. On suppose que

$$\mathbb{E}[\xi_1] = 0\,, \qquad \mathrm{Cov}\,(\xi_1) = I_d\,,$$

où I_d est la matrice identité de $\mathbb{R}^d$.

On pose $S_0 = 0$, et pour $n \geq 1$, $S_n = \xi_1 + \ldots + \xi_n$. On étend alors S en une fonction continue de $\mathbb{R}_+$ dans $\mathbb{R}^d$ par la formule

$$S_t = (1 - \{t\})S_{\lfloor t\rfloor} + \{t\}S_{\lfloor t\rfloor+1}\,, \qquad t \geq 0\,,$$

où $\{t\} = t - \lfloor t\rfloor$ est la partie fractionnaire de t. Enfin, pour tout $n \geq 1$, on pose

$$S_t^{(n)} = \frac{S_{nt}}{\sqrt{n}}\,, \qquad t \geq 0\,.$$

On voit ainsi $S^{(n)}$ comme une variable aléatoire à valeurs dans $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$.

Théorème 2.2 (Donsker) *La suite* $(S^{(n)})_{n\geq 1}$ *converge en loi dans* $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ *vers un processus* $(B_t)_{t\geq 0}$*, appelé* mouvement brownien standard de dimension d.

La loi de ce dernier est caractérisée par le fait que pour tout $0 = t_0 \leq t_1 \leq \ldots \leq t_k$*, les variables* $B_{t_i} - B_{t_{i-1}}, 1 \leq i \leq k$ *sont indépendantes, respectivement de loi gaussienne* $\mathcal{N}(0, (t_i - t_{i-1})I_d)$.

La deuxième partie du théorème est une conséquence de la proposition 2.3, puisque les lois marginales de dimension finie sont déterminées par les propriétés d'indépendance des accroissements de $(B_t)_{t\geq 0}$, et le fait qu'ils soient de lois prescrites. Ces considérations ont été déjà faites dans l'introduction, où nous avions remarqué que l'existence du mouvement brownien n'est pas garantie. En fait, la preuve du théorème de Donsker va montrer au passage ce résultat d'existence. En effet, nous allons montrer que la suite des lois des $S^{(n)}$ converge étroitement vers une limite $\mathcal{W}$ dont les lois marginales de dimension finie sont celles du mouvement brownien, et donc que la loi du mouvement brownien est $\mathcal{W}$, et existe donc bien !

Pour l'instant, seules les lois marginales $\mathcal{W}^{t_1,\ldots,t_k}$ sont bien définies : un vecteur aléatoire $(Y_1, \ldots, Y_k)$ a pour loi $\mathcal{W}^{t_1,\ldots,t_k}$ si et seulement si les accroissements $(Y_i - Y_{i-1}, 1 \leq i \leq k)$ sont indépendants et respectivement de loi $\mathcal{N}(0, (t_i - t_{i-1})I_d)$, où l'on a noté $Y_0 = 0$.

Étape 1 : convergence des lois marginales de dimension finie. Nous fixons des instants $0 = t_0 < t_1 < \ldots < t_k$, et nous montrons que $(S^{(n)}_{t_1}, \ldots, S^{(n)}_{t_k})$ converge en loi. Pour cela, remarquons que les variables aléatoires $(S_{\lfloor nt_i \rfloor} - S_{\lfloor nt_{i-1} \rfloor}, 1 \leq i \leq k)$ sont indépendantes, respectivement de même loi que $S_{\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor}, 1 \leq i \leq k$. De cela, on déduit que

$$\left(\frac{S_{\lfloor nt_i \rfloor} - S_{\lfloor nt_{i-1} \rfloor}}{\sqrt{n}}, 1 \leq i \leq k \right) \xrightarrow[n\to\infty]{\text{(loi)}} (N_1, N_2, \ldots, N_k)\,,$$

où les variables aléatoires $N_1, \ldots, N_k$ sont indépendantes, respectivement de loi $\mathcal{N}(0, (t_i - t_{i-1})I_d)$. Ceci provient d'une application du théorème centrale limite vectoriel. Par image continue, on en déduit que

$$\left(\frac{S_{\lfloor nt_i \rfloor}}{\sqrt{n}}, 1 \leq i \leq k \right) \xrightarrow[n\to\infty]{\text{(loi)}} (N_1, N_1 + N_2, \ldots, N_1 + \ldots + N_k)\,.$$

Enfin, remarquons que

$$\left| S^{(n)}_t - \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \right| \leq \frac{|\xi_{\lfloor nt \rfloor}|}{\sqrt{n}}\,,$$

qui converge vers 0 en probabilité pour tout t fixé. Pour cette raison, on a la convergence en probabilité :

$$\left(S^{(n)}_{t_i} - \frac{S_{\lfloor nt_i \rfloor}}{\sqrt{n}}, 1 \leq i \leq k \right) \xrightarrow[n\to\infty]{\mathbb{P}} 0\,.$$

Par le lemme de Slutzky (proposition 1.18), on en déduit que

$$\left(S^{(n)}_{t_i}, 1 \leq i \leq k \right) \xrightarrow[n\to\infty]{\text{(loi)}} (N_1, N_1 + N_2, \ldots, N_1 + \ldots + N_k)\,.$$

On en déduit que $S^{(n)}$ converge bien au sens des marginales de dimension finie, et que les limites des lois marginales sont les $\mathcal{W}^{t_1,\dots,t_k}$.

Étape 2 : une inégalité maximale. Afin de montrer que la suite $(S^{(n)})_{n\geq 1}$ est tendue, nous avons recours à un lemme intermédiaire.

Lemme 2.3 *Soit $\lambda \geq 0$. Alors on a, pour tout $n \geq 1$,*

$$\mathbb{P}(\max_{0\leq i\leq n} |S_i| > \lambda\sqrt{n}) \leq 2\mathbb{P}(|S_n| > (\lambda - \sqrt{2d})\sqrt{n})\,.$$

Preuve. Le lemme est évidemment vérifié pour $\lambda \in [0, \sqrt{2d}[$, donc nous supposons que $\lambda \geq \sqrt{2d}$. Notons $S_n^* = \max_{0\leq i\leq n} |S_i|$, et $T = \inf\{k \geq 0 : |S_k| > \lambda\sqrt{n}\}$. On a alors

$$\begin{aligned}
\mathbb{P}(S_n^* > \lambda\sqrt{n}) &= \mathbb{P}(T \leq n) \\
&= \mathbb{P}(T \leq n, |S_n| > (\lambda - \sqrt{2d})\sqrt{n}) + \mathbb{P}(T \leq n, |S_n| \leq (\lambda - \sqrt{2d})\sqrt{n}) \\
&\leq \mathbb{P}(|S_n| > (\lambda - \sqrt{2d})\sqrt{n}) + \sum_{j=0}^{n} \mathbb{P}(T = j, |S_n| \leq (\lambda - \sqrt{2d})\sqrt{n})\,.
\end{aligned}$$

Sur l'événement $\{T = j\}$ (avec $j \leq n$) on a forcément $|S_j| > \lambda\sqrt{n}$, et donc si on suppose en plus que $|S_n| < (\lambda - \sqrt{2d})\sqrt{n}$, on a $|S_n - S_j| > \sqrt{2dn}$ par inégalité triangulaire. On remarque que ce dernier événement est indépendant de $\sigma(\xi_1, \dots, \xi_j)$ puisque $S_n - S_j = \xi_{j+1} + \dots + \xi_n$, tandis que l'événement $\{T = j\}$ est clairement dans $\sigma(\xi_1, \dots, \xi_j)$. On en déduit que

$$\begin{aligned}
\mathbb{P}(T = j, |S_n| \leq (\lambda - \sqrt{2d})\sqrt{n}) &\leq \mathbb{P}(T = j, |S_n - S_j| > \sqrt{2dn}) \\
&= \mathbb{P}(T = j)\mathbb{P}(|S_n - S_j| > \sqrt{2dn}) \\
&\leq \mathbb{P}(T = j)\frac{(n-j)\mathrm{Var}\,(|\xi_1|)}{2dn} \\
&\leq \frac{d}{2d}\mathbb{P}(T = j)\,,
\end{aligned}$$

où l'on a utilisé l'inégalité de Bienaymé-Chebychev pour l'avant-dernière inégalité, ainsi que le fait que $\mathrm{Var}\,(|\xi_1|) \leq \mathbb{E}[|\xi_1|^2] = d$. Finalement, on en déduit que

$$\mathbb{P}(T \leq n) \leq \mathbb{P}(|S_n| > (\lambda - \sqrt{2d})\sqrt{n}) + \frac{1}{2}\mathbb{P}(T \leq n)\,,$$

d'où l'on obtient immédiatement le résultat. □

Étape 3 : tension. Nous montrons enfin que la suite $(S^{(n)})_{n\geq 1}$ est tendue dans $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$. Du fait de la discussion de la section 2.4, nous savons qu'il suffit de montrer que les suites des restrictions $(S^{(n)}|_{[0,N]})_{n\geq 1}$ sont tendues pour tout entier $N \geq 1$. Pour simplifier les notations, nous le faisons dans le cas $N = 1$, le cas général étant identique.

La première chose à remarquer est que pour tout $\delta \in]0, 1[$ et $t \in [0, 1 - \delta]$, on a

$$\sup_{t\leq s\leq t+\delta} |S_{ns} - S_{nt}| \leq 2 \max_{0\leq i\leq \lfloor n\delta\rfloor + 2} |S_{i+\lfloor nt\rfloor} - S_{\lfloor nt\rfloor}|\,.$$

En effet, pour tout $s \in [t, t+\delta]$, en utilisant le fait que S est interpolée linéairement entre ses valeurs aux entiers,

$$\begin{aligned}|S_{ns} - S_{nt}| &\leq |S_{ns} - S_{\lfloor nt\rfloor}| + |S_{nt} - S_{\lfloor nt\rfloor}| \\ &\leq \max\left(|S_{\lfloor ns\rfloor} - S_{\lfloor nt\rfloor}|, |S_{\lfloor ns\rfloor+1} - S_{\lfloor nt\rfloor}|\right) + |S_{\lfloor nt\rfloor+1} - S_{\lfloor nt\rfloor}|,\end{aligned}$$

et on en déduit la majoration voulue du fait que $\lfloor ns\rfloor + 1 \leq \lfloor nt\rfloor + \lfloor n\delta\rfloor + 2$. On en déduit que pour tout $\eta > 0$, pour tout $\delta \in]0,1[$, et enfin pour tout $t \in [0, 1-\delta]$,

$$\begin{aligned}\mathbb{P}\left(\sup_{0\leq r\leq\delta} |S^{(n)}_{r+t} - S^{(n)}_t| > \eta\right) &\leq \mathbb{P}\left(\max_{0\leq i\leq\lfloor n\delta\rfloor+2} |S_{i+\lfloor nt\rfloor} - S_{\lfloor nt\rfloor}| > \eta\sqrt{n}/2\right) \\ &= \mathbb{P}\left(\max_{0\leq i\leq\lfloor n\delta\rfloor+2} |S_i| > \frac{\eta}{2}\sqrt{\frac{n}{\lfloor n\delta\rfloor+2}}\sqrt{\lfloor n\delta\rfloor+2}\right) \\ &\leq 2\mathbb{P}\left(\frac{|S_{\lfloor n\delta\rfloor+2}|}{\sqrt{\lfloor n\delta\rfloor+2}} > \frac{\eta}{2}\sqrt{\frac{n}{\lfloor n\delta\rfloor+2}} - \sqrt{2d}\right),\end{aligned}$$

où l'on a utilisé le lemme 2.3. Le majorant ne dépend pas de t, et le théorème central limite (conjointement au fait que le module d'une variable gaussienne est à densité) montre qu'il tend lorsque $n \to \infty$ vers $2\mathbb{P}(|N| > \eta/2\sqrt{\delta} - \sqrt{2d})$, où N est de loi $\mathcal{N}(0, I_d)$. On utilise alors le fait que, par un changement de variables en coordonnées sphériques dans $\mathbb{R}^d$, en notant σ_{d-1} le volume de la sphère unité de $\mathbb{R}^d$,

$$\mathbb{P}(|N| > x) = \frac{2\sigma_{d-1}}{(2\pi)^{d/2}} \int_x^\infty r^{d-2} e^{-r^2/2} \mathrm{d}r = \frac{2\sigma_{d-1}}{(2\pi)^{d/2}} x^{d-3} e^{-x^2/2}(1 + o(1)),$$

cette dernière estimée lorsque $x \to \infty$ s'obtenant par une intégration par parties. Finalement, nous avons obtenu, en prenant $\delta = 1/k$, avec $k \geq 1$ entier,

$$\begin{aligned}&\limsup_{n\to\infty} k \max_{0\leq i\leq k-1} \mathbb{P}\left(\sup_{0\leq r\leq 1/k} |S^{(n)}_{i/k+r} - S^{(n)}_{i/k}| > \eta\right) \\ &\leq Ck\left(\frac{\eta\sqrt{k}}{2} - \sqrt{2d}\right)^{d-3} \exp\left(-\left(\frac{\eta\sqrt{k}}{2} - \sqrt{2d}\right)^2\right),\end{aligned}$$

pour une constante $C > 0$. Comme le majorant tend vers 0 lorsque $k \to \infty$, la proposition 2.7 entraîne que $(S^{(n)})_{n\geq1}$ est tendue — l'hypothèse que $(S^{(n)}_0)_{n\geq1}$ est tendue étant évidente, puisque toutes les variables aléatoires de cette suite sont nulles.

Conclusion. Nous avons montré que la suite $(S^{(n)})_{n\geq1}$ est tendue, et converge au sens des marginales de dimension finie vers les lois marginales $\mathcal{W}^{t_1,\ldots,t_k}$. Toute mesure limite de la loi de $S^{(n)}$ le long d'une sous-suite est donc la loi $\mathcal{W}$ du mouvement brownien, qui existe donc en particulier. La proposition 2.5 montre finalement que $S^{(n)}$ converge en loi vers un mouvement brownien standard, ce qui termine la preuve du théorème.

2.6 Quelques mots sur des processus discontinus

De nombreux processus naturels en probabilités, comme le processus de Poisson que nous allons introduire dans le chapitre suivant, ne sont pas continus, mais ont des sauts. Par ailleurs, même si l'on s'intéresse à la convergence des marches aléatoires vers le mouvement brownien, comme dans le théorème de Donsker, il peut paraître naturel de travailler avec le processus

$$S_t^{[n]} = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}}, \qquad t \geq 0,$$

plutôt qu'avec la version interpolée $S^{(n)}$ que nous avons considérée.

L'espace $\mathcal{C}(\mathbb{R}_+, E)$ est évidemment inadapté à l'étude de la convergence en loi de tels processus. Néanmoins, il est naturel de penser que le processus $S^{[n]}$ est proche du processus continu $S^{(n)}$, et par conséquent, converge en un sens naturel vers le mouvement brownien.

On peut par exemple considérer que $S^{[n]}$ est un processus à valeurs dans l'espace $\mathcal{D}(\mathbb{R}_+, E)$ des fonctions continues à droite admettant des limites à gauche (ce que l'on abrège en « càdlàg »), muni de la topologie compact-ouverte. Malheureusement, cette topologie n'est pas séparable en général. Par exemple, la famille de fonctions $\mathbb{1}_{[a,\infty[} : a \geq 0$ de $\mathcal{D}(\mathbb{R}_+, \mathbb{R})$ est sans point d'accumulation. Ceci n'empêche pas *a priori* de considérer la convergence étroite sur l'espace $\mathcal{D}(\mathbb{R}_+, \mathbb{R})$, mais il convient de la manipuler avec précaution.

Proposition 2.11 *On se place sous les hypothèses de la section 2.5. Alors pour tout $T > 0$ et $\varepsilon > 0$ on a*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{0 \leq t \leq T} |S_t^{[n]} - S_t^{(n)}| \geq \varepsilon \right) = 0.$$

Par conséquent, $S^{[n]}$ converge en loi dans $\mathcal{D}(\mathbb{R}_+, \mathbb{R}^d)$, muni de la topologie compact-ouverte, vers le mouvement brownien standard.

Preuve. Montrons la première partie de l'énoncé pour $T = 1$, le résultat général étant similaire. On constate que

$$\sup_{0 \leq t \leq 1} |S_t^{[n]} - S_t^{(n)}| = \sup_{0 \leq t \leq 1} \frac{\{t\}}{\sqrt{n}} |S_{\lfloor nt \rfloor + 1} - S_{\lfloor nt \rfloor}| \leq \omega\left(S^{(n)}, \frac{1}{n} \right),$$

puisque $S_i = \sqrt{S_{i/n}^{(n)}}$ pour tout entier $i \geq 1$. Comme on sait déjà par le théorème de Donsker que $(S^{(n)}, n \geq 1)$ est une suite de variables aléatoires dans $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ qui est tendue, la proposition 2.6 montre que pour tout $\varepsilon > 0$, on a

$$\lim_{\delta \downarrow 0} \sup_{n \geq 1} \mathbb{P}(\omega(S^{(n)}, \delta) \geq \varepsilon) = 0,$$

ce qui implique que $\mathbb{P}(\omega(S^{(n)}, 1/n) > \varepsilon)$ converge vers 0 lorsque $n \to \infty$. Ceci démontre la première partie de l'énoncé.

Notons $M_n = \sup_{0\leq t\leq 1}|S^{[n]}_t - S^{(n)}_t|$, et considérons alors la suite $(S^{(n)}, M_n)_{n\geq 1}$ à valeurs dans $\mathcal{C}([0,1],\mathbb{R}^d)\times\mathbb{R}_+$, polonais. Le lemme de Slutzky, conjointement à ce que nous venons de montrer, implique que $(S^{(n)}, M_n)$ converge en loi vers $(B, 0)$ où $B = (B_t, 0\leq t\leq 1)$ est un mouvement brownien standard restreint à l'intervalle $[0,1]$. Par le théorème de représentation de Skorokhod, on peut supposer sans perte de généralité que cette convergence a lieu au sens presque-sûr. On a alors que p.s. (la norme infinie désigne en fait la restriction aux fonctions indexées par $[0,1]$)

$$\begin{aligned} \|S^{[n]} - B\|_\infty &\leq \|S^{(n)} - B\|_\infty + \|S^{[n]} - S^{(n)}\|_\infty \\ &= \|S^{(n)} - B\|_\infty + M_n \\ &\underset{n\to\infty}{\longrightarrow} 0\,. \end{aligned}$$

Comme la convergence p.s. implique la convergence en loi (dans un espace métrique quelconque), on en déduit le résultat, au moins en restriction à l'intervalle $[0,1]$. Mais le résultat sur l'intervalle $[0,T]$ est vrai pour les mêmes raisons, et est donc valable pour la topologie compact-ouverte (on laisse au lecteur le soin de formaliser proprement la fin du raisonnement). $\square$

Remarque. Le recours au théorème de représentation de Skorokhod (en faisant attention de considérer des variables à valeurs dans un espace polonais !) n'est pas essentiel, mais assez commode pour conclure rapidement. Le lecteur intéressé pourra essayer de trouver une preuve qui n'utilise pas ce résultat.

Comme on le voit, l'incursion dans des topologies non polonaises est assez malcommode et oblige à quelques contorsions. En fait, on a en général recours à une topologie «polonaise» dite *topologie J_1 de Skorokhod* (il y a en tout 4 topologies de Skorokhod, dites J_1, M_1, J_2 et M_2) sur l'ensemble $\mathcal{D}(\mathbb{R}_+, E)$. L'espace de Skorokhod $\mathcal{D}(\mathbb{R}_+, E)$ est sensiblement plus compliqué que l'espace $\mathcal{C}(\mathbb{R}_+, E)$, et nous ne rentrerons pas davantage dans les détails de sa définition. Le lecteur intéressé pourra consulter par exemple [1, 4, 7]. Cette topologie devient pratiquement indispensable lorsque l'on s'intéresse à des théorèmes limites pour lesquels les processus intervenant à la limite sont discontinus, comme par exemple les processus de Lévy que nous allons rencontrer au chapitre 4. C'est le cas également pour de nombreux *processus de Markov*, qui sont l'objet d'étude de [4], et pour les *semimartingales*, qui sont traitées dans [7] (le mouvement brownien, ou plus généralement les processus de Lévy, se situent au carrefour de ces deux classes d'objets, puisqu'ils sont à la fois des processus de Markov et des semimartingales).

2.7 Les théorèmes d'existence et de régularité de Kolmogorov

Cette partie dévie légèrement de notre propos principal. La preuve que nous avons donnée précédemment de l'existence de la loi $\mathcal{W}$ du mouvement brownien peut paraître très indirecte, et difficile à généraliser. En effet, il n'est pas dit qu'un processus continu donné est nécessairement limite de modèles discrets aisés à étudier, comme les marches aléatoires !

2.7.1 Théorème d'existence de Kolmogorov

Il existe un résultat très général, dû à Kolmogorov, qui tranche la question d'existence mentionnée ci-dessus. Pour l'énoncer, nous devons d'abord définir la notion de *famille projective* de mesures. On notera $I \subset_f \mathbb{R}_+$ pour dire que I est une partie finie non-vide de $\mathbb{R}_+$. Si $J \subset I \subset_f \mathbb{R}_+$, on a une application de projection naturelle π^{IJ} de E^I vers E^J.

Définition 2.3 *Pour tout $I \subset_f \mathbb{R}_+$, soit μ^I une mesure de probabilités sur E^I. La famille $(\mu^I, I \subset_f \mathbb{R}_+)$ est dite projective (ou cohérente) si pour tout $J \subseteq I \subset_f \mathbb{R}_+$ on a $\mu^J = \pi_*^{IJ}\mu^I$.*

Notons que si μ est une mesure de probabilités sur $E^{\mathbb{R}_+}$ muni de la tribu produit, alors la famille $(\mu^{t_1,\ldots,t_k} : \{t_1, \ldots, t_k\} \subset_f \mathbb{R}_+)$ des marginales de μ de dimension finie (dans la notation, on a ordonné systématiquement $t_1 < t_2 < \ldots < t_k$) est évidemment cohérente, puisque $\pi^{IJ} \circ \pi^I = \pi^J$ pour $J \subset_f I \subset_f \mathbb{R}_+$, où π^I est la projection naturelle de $E^{\mathbb{R}_+}$ sur E^I. Le théorème de Kolmogorov est une forme de réciproque à cette constatation. Même si nous supposons presque toujours tacitement dans ce texte que E est un espace polonais, nous insistons sur le fait que cette hypothèse est fondamentale pour le résultat suivant, et nous l'incluons donc dans l'énoncé.

Théorème 2.3 (Théorème d'existence de Kolmogorov) *Soit E un espace polonais. Soit $(\mu^I, I \subset_f \mathbb{R}_+)$ une famille projective de mesures de probabilités. Alors il existe une mesure de probabilités μ sur l'espace produit $E^{\mathbb{R}_+}$, telle que $\mu^{\{t_1,\ldots,t_k\}} = \mu^{t_1,\ldots,t_k}$ est la marginale de μ en $t_1, \ldots, t_k$, pour tout $t_1 < t_2 < \ldots < t_k$ positifs.*

La preuve est omise.

2.7.2 Critère de continuité de Kolmogorov

Il faut noter que le dernier théorème porte sur l'existence d'une mesure μ sur l'espace E^I des fonctions de I dans E, mais ne dit rien sur la potentielle régularité d'un processus aléatoire de loi μ — et pour cause, puisque par exemple, la continuité n'est pas une propriété mesurable pour la tribu produit (voir les exercices ci-dessous). En particulier, rien ne dit *a priori* que la mesure μ est en fait portée par l'espace $\mathcal{C}(\mathbb{R}_+, E)$.

Le «mieux» que l'on puisse espérer est donc de trouver un processus stochastique *le plus régulier possible*, par exemple continu, et qui admet μ pour loi.

Pour mieux comprendre ce qui se passe, considérons le processus nul $(X_t = 0)_{0 \leq t \leq 1}$, et le processus $(Y_t = \mathbb{1}_{\{t=U\}})_{0 \leq t \leq 1}$, où U est une variable aléatoire uniforme dans $[0, 1]$. Comme pour tout t fixé on a $U \neq t$ presque sûrement, on en déduit que les marginales de dimension finie de X et de Y sont les mêmes, donc que X et Y ont la même loi, si on les voit tous deux comme variables aléatoires à valeurs dans $\mathbb{R}^{[0,1]}$ muni de la tribu produit. Pourtant, on voit que X est continu alors que Y ne l'est pas. Ceci donne d'ailleurs une autre preuve du fait que $\mathcal{C}([0, 1], \mathbb{R})$ n'est pas un sous-ensemble mesurable de $\mathbb{R}^{[0,1]}$, puisque sinon, en notant μ la loi de X, on aurait $\mu(\mathcal{C}([0, 1], \mathbb{R})) = \mathbb{P}(X \in \mathcal{C}([0, 1], \mathbb{R})) = 1$, mais aussi $\mu(\mathcal{C}([0, 1], \mathbb{R})) = \mathbb{P}(Y \in \mathcal{C}([0, 1], \mathbb{R})) = 0$.

Définition 2.4 *Soit* $(X_t, t \in I)$ *et* $(Y_t, t \in I)$ *deux processus stochastiques, c'est-à-dire deux familles de variables aléatoires indexées par* t*, définies sur un même espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$*. On dit que* X *est une modification (ou une version) de* Y *si pour tout* $t \in I$ *on a* $\mathbb{P}(X_t = Y_t) = 1$*.*

En particulier si X est une modification de Y, alors X et Y ont la même loi, c'est-à-dire ont les mêmes lois marginales de dimension finie.

Théorème 2.4 (Théorème de régularité de Kolmogorov) *Soit* $(X_t, t \geq 0)$ *un processus stochastique à valeurs dans un espace polonais* E*. On suppose qu'il existe* $\beta, p, C > 0$ *tel que*

$$\mathbb{E}[d(X_s, X_t)^p] \leq C|t-s|^{1+\beta}, \qquad \text{pour tout } s, t \geq 0 .$$

Alors il existe une modification Y *de* X *qui est continue, et même p.s. localement Hölder-continue d'exposant* α *pour tout* $\alpha \in]0, \beta/p[$*.*

Preuve. On procède comme dans la preuve de la proposition 2.8. Raisonnons d'abord sur le processus $(X_t, 0 \leq t \leq 1)$. Soit $\alpha \in]0, \beta/p[$ fixé. Rappelons qu'une étape intermédiaire de la preuve est de montrer que

$$\mathbb{P}(Z_k \geq K 2^{-k\alpha}) \leq \frac{C}{K^p 2^{(\beta - p\alpha)k}},$$

où $Z_k = \max_{0 \leq i \leq 2^k - 1} d(X_{(i+1)2^{-k}}, X_{i 2^{-k}})$. Comme les majorants forment une suite sommable, le lemme de Borel-Cantelli montre que p.s. pour tout k à partir d'un certain rang, on a $Z_k \leq K 2^{-k\alpha}$. Donc p.s. il existe un $K' > K$ (aléatoire) fini tel que $Z_k \leq K' 2^{-k\alpha}$ pour tout $k \geq 0$. Par le lemme 2.2, ceci implique que pour toute paire de rationnels dyadiques $s, t \in \bigcup_{n \geq 0} Q_n$, on a

$$d(X_t, X_s) \leq \frac{2K'}{1 - 2^{-\alpha}} |t-s|^{\alpha} .$$

Par conséquent, le processus $(X_t, t \in \bigcup_{n \geq 0} Q_n)$ est uniformément continu, ce qui implique qu'il admet un unique prolongement continu $(Y_t, 0 \leq t \leq 1)$. Nous montrons finalement que ce dernier est une modification de $(X_t, 0 \leq t \leq 1)$. Soit donc $t \in [0,1]$ fixé, et t_n une suite de rationnels dyadiques de limite t. Alors on a

$$\mathbb{E}[d(X_{t_n}, X_t)^p] \leq C|t_n - t|^{1+\beta} .$$

Comme X_{t_n} converge p.s. vers Y_t par définition de Y, le lemme de Fatou montre alors que $\mathbb{E}[d(Y_t, X_t)^p] = 0$, et donc que $Y_t = X_t$ presque sûrement.

Le résultat général se déduit du cas des processus indexés par $[0,1]$ en considérant les processus $(X_{At}, 0 \leq t \leq 1)$ pour $A > 0$. □

Ainsi, sous les hypothèses du théorème précédent, on déduit que la loi μ de X peut être vue comme une loi sur $\mathcal{C}(\mathbb{R}_+, E)$, quitte à considérer la modification continue Y.

2.8 Exercices pour le chapitre 2

Exercice 1 — Une convergence jointe
Soit $(X^n, n \geq 0)$ une suite de processus dans $\mathcal{C} = \mathcal{C}([0,1], \mathbb{R}^d)$. On suppose que X^n converge en loi vers X pour la topologie uniforme sur $\mathcal{C}$. Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction continue. On pose

$$Y_n = \int_0^1 f(X_s^n)\mathrm{d}s\,, \qquad Y = \int_0^1 f(X_s)\mathrm{d}s\,.$$

Montrer que Y_n converge en loi vers Y. Montrer que l'on a même que le couple (X^n, Y_n) converge en loi vers (X, Y).

Exercice 2 — Temps d'atteinte
(i) L'application $x \mapsto t_a(x) = \inf\{t \geq 0 : x(t) \geq a\}$, définie sur $\mathcal{C} = \mathcal{C}(\mathbb{R}_+, \mathbb{R})$ et à valeurs dans $\mathbb{R}_+ \cup \{\infty\}$, est-elle continue ? Sinon, a-t-elle des points de continuité, et si oui, lesquels ?

(ii) Soit $(X^n, n \geq 0)$ une suite de processus dans $\mathcal{C}$. On suppose que X^n converge en loi vers X pour la topologie uniforme sur les compacts. On suppose également que $X_0^n = 0$ pour tout n. On note, pour $a \geq 0$,

$$T_a^n = t_a(X^n)\,, \qquad T_a = t_a(X)\,.$$

Discuter la convergence (ou non) de T_a^n vers T_a. Discuter le cas particulier où X est le mouvement brownien standard (cette dernière question se traite mieux avec la notion de *propriété de Markov forte* pour le mouvement brownien, voir le cours de calcul stochastique).

Exercice 3 — Processus croissants
Soit $X^n, n \geq 0$ et X des processus croissants, continus, de $[0,1]$ dans $\mathbb{R}$.

Montrer que X^n converge en loi vers X pour la topologie uniforme sur $\mathcal{C}$ si et seulement si X^n converge vers X au sens des marginales de dimension finie.

Exercice 4 — Martingales continues et convergence
Dans l'exercice suivant, on supposera que l'espace de probabilités ambiant $(\Omega, \mathcal{F}, \mathbb{P})$ est muni d'une filtration, c'est-à-dire d'une famille $(\mathcal{F}_t, t \geq 0)$ de tribus telle que $\mathcal{F}_s \subseteq \mathcal{F}_t$ pour tout $s \leq t$. Une martingale est une famille de variables aléatoires $(X_t, t \geq 0)$ telle que $X_t \in L^1$ pour tout $t \geq 0$, et telle que pour tout $s \leq t$,

$$\mathbb{E}[X_t \,|\, \mathcal{F}_s] = X_s\,.$$

Soit $(X^n, n \geq 0)$ une suite de processus de $\mathcal{C} = \mathcal{C}([0,1], \mathbb{R})$, intégrables et adaptés par rapport à des filtrations $(\mathcal{F}_t^n)$, c'est-à-dire que X_t^n est mesurable par rapport à $\mathcal{F}_t^n$ pour tout $t \geq 0$. On suppose que X^n est une $\mathcal{F}^n$-martingale, et que X^n converge en loi vers X dans $\mathcal{C}$.

(i) En supposant que X^n est uniformément bornée, $\sup_n \|X^n\|_\infty \leq C$ pour un $C < \infty$ déterministe, montrer que $(X_t, 0 \leq t \leq 1)$ est une $(\mathcal{F}_t)$-martingale où $\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t)$.

(ii) Même question en supposant $(X^n_t, n \geq 0)$ uniformément intégrable pour tout $t \in [0,1]$.

Exercice 5 — $\mathcal{C}([0,1],E)^2$ ***vs.*** $\mathcal{C}([0,1],E^2)$

Pour tout $n \geq 0$, on se donne deux processus X^n, Y^n dans l'espace $\mathcal{C}([0,1],\mathbb{R})$, définis sur un même espace de probabilités (qui peut dépendre éventuellement de n). On suppose que $X^n \to X$ et $Y^n \to Y$ en loi.

(i) Montrer que $((X^n, Y^n), n \geq 0)$ est une suite tendue dans $\mathcal{C}([0,1],\mathbb{R})^2$.

(ii) On voit maintenant le couple (X^n, Y^n) comme un unique processus

$$t \longmapsto (X^n_t, Y^n_t)\,, \qquad t \in [0,1]$$

dans $\mathcal{C}([0,1],\mathbb{R}^2)$. Est-il vrai que la suite $((X^n, Y^n), n \geq 0)$ est tendue dans $\mathcal{C}([0,1],\mathbb{R}^2)$?

(iii) On voit maintenant le couple (X^n, Y^n) comme une fonction $(X^n(s), Y^n(t))_{s,t\in[0,1]}$ dans $\mathcal{C}([0,1]^2,\mathbb{R}^2)$. Est-il vrai que la suite $((X^n, Y^n), n \geq 0)$ est tendue dans $\mathcal{C}([0,1]^2,\mathbb{R}^2)$?

Exercice 6 — **Une autre preuve de l'existence du mouvement brownien**

En utilisant les résultats de la section 2.7, donner une autre preuve du fait que la loi du mouvement brownien standard $\mathcal{W} \in \mathcal{M}_1(\mathcal{C}(\mathbb{R}_+,\mathbb{R}^d))$ existe.

Exercice 7 — **Théorème de Donsker avec hypothèse de moments**

Soit $(X_1, X_2, \ldots)$ une suite de variables aléatoires i.i.d. centrées à valeurs dans $\mathbb{R}$. On suppose que $E[X_1^4] < \infty$, et on note $\sigma = \mathrm{Var}\,(X_1)$. On pose

$$S_n = \sum_{k=1}^{n} X_k\,, \qquad n \geq 0\,,$$

et $S_t = (1 - \{t\})S_{\lfloor t\rfloor} + \{t\}S_{\lfloor t\rfloor+1}, t \geq 0$ l'interpolation continue affine par morceaux associée.

(i) Montrer qu'il existe $C \in \mathbb{R}_+$ telle que pour tout $n \geq 0$, on a $E[|S_n|^4] \leq Cn^2$.

(ii) En déduire qu'il existe une constante $C' \in \mathbb{R}_+$ telle que pour tout $s, t \in \mathbb{R}_+$,

$$E[|S_{nt} - S_{ns}|^4] \leq C'n^2|t-s|\,.$$

(iii) Montrer que $(S_{nt}/(\sigma\sqrt{n}), t \geq 0)$ converge en loi dans $\mathcal{C}(\mathbb{R}_+,\mathbb{R})$ vers un mouvement brownien standard.

Exercice 8 — Autour du théorème de Donsker
Soit $(\xi_i, i \geq 1)$ une suite de variables réelles i.i.d. centrées et de variance 1. On pose $S_0 = 0$ et $S_n = \xi_1 + \ldots + \xi_n$ pour tout $n \geq 1$. Puis, on note $A_0 = 0$ et $A_n = S_1 + \ldots + S_n$ pour tout $n \geq 1$. Les processus S, A sont interpolées linéairement entre leurs valeurs aux entiers, et déterminent deux processus continus. Enfin, on note

$$S_t^{(n)} = \frac{S_{nt}}{\sqrt{n}}, \qquad A_t^{(n)} = \frac{A_{nt}}{n^{3/2}}, \qquad 0 \leq t \leq 1\,.$$

Montrer que $(S^{(n)}, A^{(n)})$ converge en loi dans $\mathcal{C}([0,1], \mathbb{R})^2$ vers une limite dont on précisera la loi.

Exercice 9 — Invariance par changement d'échelle du mouvement brownien, propriété de Markov simple
En revenant à la définition du mouvement brownien par ses marginales de dimension finie, ou en utilisant le théorème de Donsker, montrer que si $(B_t, t \geq 0)$ est un mouvement brownien standard dans $\mathbb{R}^d$, alors pour tout $\lambda > 0$, le processus

$$\left(\frac{1}{\sqrt{\lambda}} B_{\lambda t}, t \geq 0\right)$$

est aussi un mouvement brownien standard. Montrer également que pour tout $t > 0$ fixé, le processus

$$B^{(t)} = (B_{t+s} - B_t, s \geq 0)$$

est un mouvement brownien standard, et que ce processus est indépendant du processus $(B_u, 0 \leq u \leq t)$ (propriété de Markov simple).

Ces résultats servent de façon cruciale dans l'exercice suivant.

Exercice 10 — Un mouvement brownien peut en cacher un autre
Dans cet exercice, on notera $(B_t, t \geq 0)$ un mouvement brownien standard à valeurs dans $\mathbb{R}$. On pose également, pour tout $c > 0$,

$$\beta_t^{(c)} = \frac{1}{\sqrt{c}} B_{ct}, \qquad t \geq 0\,.$$

(i) Montrer que la famille $(\beta^{(c)}, c > 0)$ de variables aléatoires à valeurs dans $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ est uniformément tendue.

(ii) On suppose que $c \in (0,1)$. Montrer l'égalité en loi

$$((\beta_{c+t}, 0 \leq t \leq 1-c), \beta^{(c)}) \overset{\text{(loi)}}{=} ((\sqrt{c}\beta_1' + \beta_t, 0 \leq t \leq 1-c), \beta')\,,$$

où β' est un processus indépendant de β, et de même loi.

(iii) On se donne des réels positifs $0 < t_1 < t_2 < \ldots < t_k \leq 1$ et $0 < t'_1 < \ldots < t'_{k'} \leq 1$. Montrer la convergence en loi

$$(\beta_{t_1}, \ldots, \beta_{t_k}, \beta^{(c)}_{t'_1}, \ldots, \beta^{(c)}_{t'_{k'}}) \xrightarrow[c\to 0]{\text{(loi)}} (\beta_{t_1}, \ldots, \beta_{t_k}, \beta'_{t'_1}, \ldots, \beta'_{t'_{k'}}),$$

avec les mêmes notations que dans la question précédente. Montrer que cela reste vrai si $t_1 = 0$ ou $t'_1 = 0$.

(iv) En conclure, toujours avec les mêmes notations, que $(\beta, \beta^{(c)})$ converge en loi dans $\mathcal{C}(\mathbb{R}_+, \mathbb{R})^2$ vers (β, β') lorsque $c \to 0$. Expliquer pourquoi ce résultat peut paraître surprenant.

(v) Que devient la question précédente, si cette fois l'on fait tendre $c \to \infty$?

Exercice 11 — Non-mesurabilité du caractère continu

(i) Sur l'ensemble $E^{[0,1]}$ des fonctions de $[0,1]$ dans E, on place la tribu produit $\mathcal{B}_E^{\otimes[0,1]}$, c'est-à-dire la plus petite tribu rendant les applications de projection $x \mapsto x(t)$ mesurables. De façon équivalente, c'est la tribu engendrée par les cylindres mesurables

$$\prod_{t\in[0,1]} A_t,$$

où $A_t \in \mathcal{B}_E$ pour tout $t \in [0,1]$, et $\operatorname{Card}\{t \in [0,1] : A_t \neq E\} < \infty$. Justifier que $\mathcal{B}_{\mathcal{C}} = \{\mathcal{C} \cap A : A \in \mathcal{B}_{\mathbb{R}}^{\otimes[0,1]}\}$ est la restriction de la tribu produit à $\mathcal{C}$.

(ii) Pour tout $I \subset [0,1]$, on note $\mathcal{B}_E(I)$ la plus petite tribu sur $E^{[0,1]}$ rendant mesurables les applications $\pi^t, t \in I$. On note $\mathscr{D}$ l'ensemble des parties dénombrables de $[0,1]$. Montrer que la réunion

$$\bigcup_{I\in\mathscr{D}} \mathcal{B}_E(I)$$

est une tribu. Montrer que cette tribu coïncide avec la tribu produit $\mathcal{B}_E^{\otimes[0,1]}$.

(iii) En déduire que $\mathcal{C}$ n'est pas en général un élément de $\mathcal{B}_E^{\otimes[0,1]}$.

Exercice 12 — Modèle des vols aléatoires de Rayleigh

Ce problème utilise quelques rudiments sur le processus de Poisson homogène, qui sont exposés dans le chapitre suivant.

Partie I.

On considère un processus de Poisson standard $(N(t), t \geq 0)$, de sorte que $\mathrm{d}N(t)$ est une mesure de Poisson aléatoire d'intensité la mesure de Lebesgue sur $\mathbb{R}$, et $N(0) = 0$. On note $T_0 = 0$ puis $T_1 < T_2 < \ldots$ les instants de saut de N.

1. Montrer que l'on a presque-sûrement

$$N(t) \underset{t\to\infty}{\longrightarrow} \infty, \qquad \frac{T_n}{n} \underset{n\to\infty}{\longrightarrow} 1.$$

En déduire que $T_{N(t)}/N(t)$ converge presque-sûrement vers 1 lorsque $t \to \infty$.

2. Montrer que $T_{N(t)} \leq t < T_{N(t)+1}$ pour tout $t \geq 0$. En déduire que presque-sûrement on a

$$\frac{N(t)}{t} \underset{t\to\infty}{\longrightarrow} 1\,.$$

3. Pour $\varepsilon > 0$ on note

$$N_\varepsilon(t) = \varepsilon N\left(\frac{t}{\varepsilon}\right)\,, \qquad t \geq 0\,.$$

Montrer que pour tout $t \geq 0$, p.s. on a que $N_\varepsilon(t)$ converge vers t lorsque $\varepsilon \to 0$. Puis montrer que le processus $(N_\varepsilon(t), t \geq 0)$ converge p.s. vers l'identité uniformément sur les compacts.

Partie II. Soit $(X_i, i \geq 1)$ une suite de variables réelles indépendantes et de même loi, dans L^2. On suppose que $E[X_1] = 0$ et $\mathrm{Var}\,(X_1) = 1$. On pose

$$S(k) = \sum_{i=1}^{k} X_i\,, \qquad k \geq 0\,,$$

et $(S(t), t \geq 0)$ est l'interpolation linéaire continue de $(S(k), k \geq 0)$ définie par

$$S(t) = (1 - t + \lfloor t \rfloor)S(\lfloor t \rfloor) + (t - \lfloor t \rfloor)S(\lfloor t \rfloor + 1)\,, \qquad t \geq 0\,.$$

Pour tout $n \geq 1$ on note

$$S_n(t) = \frac{S(nt)}{\sqrt{n}}\,, \qquad t \geq 0\,.$$

Enfin, on fixe un nombre réel $T > 0$.

1. Justifier que la suite des lois des variables aléatoires $(\sup_{0\leq t\leq T} |S_n(t)|, n \geq 1)$ est uniformément tendue.

2. Justifier que pour tout $\eta > 0$, on a

$$\lim_{\delta\to 0} \sup_{n\geq 1} P\left(\sup_{0\leq s,t\leq T+1, |t-s|\leq\delta} |S_n(t) - S_n(s)| \geq \eta\right) = 0$$

3. Pour tout $\varepsilon > 0$ on pose (la notation $N(t)$ est celle de la partie I)

$$Y_\varepsilon(t) = \sqrt{\varepsilon} \sum_{i=1}^{N(t/\varepsilon)} X_i = \sqrt{\varepsilon} S(N(t/\varepsilon))\,.$$

On se donne $n \geq 1$ et $\eta > 0$. Montrer que pour tout $\delta \in (0, 1)$, on a

$$\begin{aligned} P\left(\sup_{0\leq t\leq T} |Y_\varepsilon(t) - S_n(t)| \geq \eta\right) \quad \leq \quad & P\left(\sup_{0\leq t\leq T} \left|\frac{N(t/\varepsilon)}{n} - t\right| \geq \delta\right) \\ & + P\left((\sqrt{n\varepsilon} - 1) \sup_{0\leq t\leq T+1} |S_n(t)| \geq \frac{\eta}{2}\right) \\ & + P\left(\sup_{0\leq s,t\leq T+1, |t-s|\leq\delta} |S_n(t) - S_n(s)| \geq \frac{\eta}{2}\right)\,. \end{aligned}$$

(Remarquer d'abord que $Y_\varepsilon(t)-S_n(t) = (\sqrt{n\varepsilon}-1)S_n(N(t/\varepsilon)/n)+S_n(N(t/\varepsilon)/n)-S_n(t)$.)

4. En posant $n_\varepsilon = \lfloor \varepsilon^{-1} \rfloor$, pour $\varepsilon \in]0,1[$, en déduire que

$$\lim_{\varepsilon\to 0} P\left(\sup_{0\le t\le T} |Y_\varepsilon(t) - S_{n_\varepsilon}(t)| \ge \eta \right) = 0 \,.$$

5. Posons

$$R(t) = \frac{t - T_{N(t)}}{T_{N(t)+1} - T_{N(t)}} X_{N(t)+1} + \sum_{i=1}^{N(t)} X_i \,,$$

de sorte que R est l'interpolation linéaire continue de $(S(N(t)), t \ge 0)$ entre ses instants de sauts. Finalement, posons

$$R_\varepsilon(t) = \sqrt{\varepsilon} R(t/\varepsilon) \,, \qquad t \ge 0 \,.$$

Montrer que pour tout $t \ge 0$, on a

$$|R_\varepsilon(t) - Y_\varepsilon(t)| \le \sqrt{\varepsilon} |X_{N(t/\varepsilon)+1}| \,,$$

puis que pour tout $\eta > 0$, on a

$$\lim_{\varepsilon\to 0} P\left(\sup_{0\le t\le T} \sqrt{\varepsilon} |X_{N(t/\varepsilon)+1}| \ge \eta \right) = 0 \,.$$

(Considérer les événements $\{N(T/\varepsilon) \ge (T+1)/\varepsilon\}$ et $\{\max\{|X_i| : 1 \le i \le (T+1)/\varepsilon\} \ge \eta/\sqrt{\varepsilon}\}$, puis justifier que $P(|X_1| \ge t) \le t^{-2} E[X_1^2 \mathbb{1}_{|X_1|\ge t}] = o(t^{-2})$.)

6. À l'aide des deux questions précédentes, montrer que le processus continu $(R_\varepsilon(t), t \ge 0)$ converge en loi lorsque $\varepsilon \to 0$ vers un mouvement brownien réel standard, pour la topologie de la convergence uniforme sur les compacts.

Le processus $(R(t), t \ge 0)$ est appelé *processus de Rayleigh*, on l'interprète en imaginant le vol d'un oiseau qui, au n-ième instant de saut T_n du processus de Poisson, choisit de voler à vitesse constante $X_n/(T_{n+1} - T_n)$ jusqu'au prochain instant où il change d'*avis*.

3 Mesures aléatoires de Poisson

L'objet de ce chapitre est de construire des mesures ponctuelles aléatoires, correspondant à l'occurrence d'événements « rares et indépendants ». En quelque sorte, la mesure qui gouverne ce genre de phénomènes est la mesure de Poisson, à travers le résultat bien connu d'approximation Poisson-binômiale suivant : si X_n est une variable aléatoire de loi binômiale de paramètres (n, p_n) où la suite (p_n) est telle que np_n converge vers $\lambda \geq 0$ lorsque $n \to \infty$, alors X_n converge en loi vers une variable aléatoire X de loi de Poisson de paramètre λ :

$$\mathbb{P}(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \qquad n \geq 0 .$$

On notera $\mathcal{P}(\lambda)$ cette loi. Si $\lambda = 0$ il s'agit de la loi δ_0. Par convention, si $\lambda = \infty$ on pose $\mathcal{P}(\infty) = \delta_\infty$. Ceci est cohérent avec le fait que $\mathcal{P}(\lambda)$, vue comme mesure sur $\mathbb{Z}_+ \cup \{\infty\}$, converge étroitement vers δ_∞ lorsque $\lambda \to \infty$. Rappelons deux propriétés cruciales de la loi de Poisson.

Proposition 3.1 *La loi de Poisson vérifie la propriété de* **superposition** *suivante : pour tout $\lambda, \lambda' \in \mathbb{R}_+ \cup \{\infty\}$, on a*

$$\mathcal{P}(\lambda) * \mathcal{P}(\lambda') = \mathcal{P}(\lambda + \lambda') .$$

Autrement dit, si X a pour loi $\mathcal{P}(\lambda)$ et Y a pour loi $\mathcal{P}(\lambda')$, et si X et Y sont indépendantes, alors $X + Y$ a pour loi $\mathcal{P}(\lambda + \lambda')$.

La loi de Poisson vérifie la propriété de **raréfaction** *suivante : soit X une variable aléatoire de loi $\mathcal{P}(\lambda)$ avec $\lambda \in \mathbb{R}_+ \cup \{\infty\}$. Indépendamment de X, soit $\xi_1, \xi_2, \ldots$ une suite de variables aléatoires i.i.d. à valeurs dans un ensemble dénombrable I, et notons $\mathbb{P}(\xi_1 = i) = p_i$ pour tout $i \in I$. Alors les variables aléatoires*

$$\sum_{j=1}^{X} \mathbb{1}_{\{\xi_j = i\}}, \qquad i \in I$$

sont indépendantes, respectivement de loi $\mathcal{P}(\lambda p_i), i \in I$.

Nous laissons au lecteur la démonstration de ces propriétés. On notera qu'elles sont en quelque sorte réciproques l'une de l'autre. Bien sûr, les cas où $\lambda \in \{0, \infty\}$ sont des cas triviaux de cette proposition.

3.1 Définition et premières propriétés des mesures aléatoires de Poisson

Soit $(E, \mathcal{E})$ un espace mesurable. Soit E^* l'ensemble des mesures atomiques σ-finies sur E de la forme $\sum_{i \in I} \delta_{x_i}$, où I est dénombrable.

On munit l'ensemble E^* de la tribu produit $\mathcal{E}^* = \sigma(\pi^A, A \in \mathcal{E})$, où $\pi^A(m) = m(A)$ pour $m \in E^*$, et $A \in \mathcal{E}$. L'application π^A est bien une application de projection, analogue aux projections π^t du chapitre 2, si l'on voit un élément de E^* comme une fonction de $\mathcal{E}$ dans $\mathbb{Z}_+ \cup \{\infty\}$. Ainsi, pour tout $A \in \mathcal{E}$, l'application $m \mapsto m(A)$ de E^* dans $\mathbb{Z}_+ \cup \{\infty\}$ est mesurable par rapport à $\mathcal{E}^*$.

Définition 3.1 *Soit μ une mesure positive, σ-finie, sur $(E, \mathcal{E})$. Une mesure aléatoire de Poisson sur $(E, \mathcal{E})$ d'intensité μ est une variable aléatoire M à valeurs dans $(E^*, \mathcal{E}^*)$, telle que si $(A_k, k \geq 1)$ est une suite de parties disjointes de $\mathcal{E}$, telles que $\mu(A_k) < \infty$ pour tout $k \geq 1$, on a que*

1. *les variables aléatoires $M(A_k), k \geq 1$ sont indépendantes, et*
2. *pour tout $k \geq 1$, la loi de $M(A_k)$ est $\mathcal{P}(\mu(A_k))$.*

Cette définition étant posée, il est naturel de se demander s'il existe une mesure de Poisson M d'intensité μ donnée sur $(E, \mathcal{E})$, et si sa loi est unique. Nous allons commencer par répondre à la seconde question.

Proposition 3.2 *La loi d'une mesure aléatoire de Poisson d'intensité μ sur $(E, \mathcal{E})$ est déterminée de façon unique par les propriétés 1. et 2. de la définition 3.1.*

Preuve. Soit M est une mesure aléatoire de Poisson d'intensité μ, définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. Nous constatons d'abord que les propriétés 1. et 2. de la définition restent valables pour $A_1, \ldots, A_k$ deux-à-deux disjoints mais plus nécessairement tels que $\mu(A_k) < \infty$, la variable $M(A_k)$ étant de loi δ_∞ si $\mu(A_k) = \infty$. Pour le voir, remarquons que si $A \in \mathcal{E}$ est tel que $\mu(A) = \infty$, alors on peut partitionner A en sous-ensembles $A_{(n)}, n \geq 1$ tels que $\mu(A_{(n)}) < \infty$. Par la définition 3.1, les variables aléatoires $M(A_{(n)}), n \geq 1$ sont indépendantes de loi respectives $\mathcal{P}(\mu(A_{(n)}))$, et leur somme est donc p.s. infinie puisque $\sum_{n \geq 1} \mu(A_{(n)}) = \infty$ et par des propriétés évidentes de la loi de Poisson. On en déduit que

$$\mathbb{P}(M(A_1) = i_1, \ldots, M(A_k) = i_k) = \prod_{j=1}^{k} e^{-\mu(A_j)} \frac{\mu(A_j)^{i_j}}{n_j!} \mathbb{1}_{\{\mu(A_i) < \infty\}}, \tag{10}$$

pour tout $A_1, \ldots, A_k \in \mathcal{E}^*$ deux-à-deux disjoints et $i_1, \ldots, i_k \in \mathbb{Z}_+$.

On remarque ensuite que les événements de la forme

$$\{m \in E^* : m(B_1) \in C_1, \ldots, m(B_k) \in C_k\},$$

où $B_1, \ldots, B_k \in \mathcal{E}$ et $C_1, \ldots, C_k$ sont des parties de $\mathbb{Z}_+$, forment un π-système $\mathcal{A}$ qui engendre la tribu $\mathcal{E}^*$. Par ailleurs, notons que de tels événements peuvent se réécrire comme réunion disjointe des événements $\{m \in E^* : m(B_1) = i_1, \ldots, m(B_k) = i_k\}$ avec $i_1 \in C_1, \ldots, i_k \in C_k$. Si maintenant $\alpha \in \{0,1\}^k$, notons

$$B_\alpha = \bigcap_{j:\alpha_j = 1} B_j \setminus \bigcup_{j:\alpha_j = 0} B_j,$$

de sorte que les ensembles $(B_\alpha, \alpha \in \{0,1\}^k)$ sont deux-à-deux disjoints, et les ensembles non vides parmi les $\{B_\alpha : \alpha_j = 1\}$ forment une partition de B_j. De ce fait, on peut écrire $\{m \in E^* : m(B_1) = i_1, \ldots, m(B_k) = i_k\}$ comme la réunion disjointe

$$\bigcup_{i_\alpha : \sum_{\alpha:\alpha_j = 1} i_\alpha = i_j} \{m \in E^* : m(B_\alpha) = i_\alpha, \alpha \in \{0,1\}^k\},$$

et les ensembles qui apparaissent dans la dernière union sont tels que les B_α sont deux-à-deux disjoints.

Nous avons donc montré que les éléments du π-système $\mathcal{A}$ sont des réunions disjointes d'événements de la forme $\{m \in E^* : m(A_1) = i_1, \dots, m(A_k) = i_k\}$ où les $A_i, 1 \leq i \leq k$ sont dans $\mathcal{E}$ et deux-à-deux disjoints. Comme la probabilité que M appartienne à un tel événement est donnée par (10), on déduit que pour tout $A \in \mathcal{A}$, la probabilité $\mathbb{P}(M \in A)$ est entièrement déterminée par les propriétés 1. et 2. de la définition 3.1. Comme $\mathcal{A}$ est un π-système qui engendre la tribu $\mathcal{E}^*$, on en déduit que la loi de M est elle-même uniquement déterminée. □

Nous nous tournons maintenant vers le problème de l'existence.

Proposition 3.3 *Pour toute mesure μ positive σ-finie sur $(E, \mathcal{E})$, il existe une mesure aléatoire de Poisson d'intensité μ.*

Preuve. La démonstration consiste à donner une construction explicite d'une mesure de Poisson. On partage la preuve en deux cas.

Cas où $\mu(E) < \infty$. Sur un certain espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, soit N une variable aléatoire de Poisson de paramètre $\mu(E)$, et $X_1, X_2, \dots$ une suite de variables aléatoires indépendantes, et indépendantes de N, de loi commune $\mu/\mu(E)$. Posons $M = \sum_{i=1}^N \delta_{X_i}$. Alors M est une variable aléatoire dans $(E^*, \mathcal{E}^*)$.

Nous affirmons alors que M est une mesure de Poisson d'intensité μ sur $(E, \mathcal{E})$. En effet, si $A_1, \dots, A_k$ sont dans $\mathcal{E}$ et deux-à-deux disjoints, posons $\xi_j = i$ si $X_j \in A_i$. Les variables $(\xi_j, j \geq 1)$ sont alors indépendantes, indépendantes de N, et à valeurs dans $\{1, \dots, k\}$, avec $P(\xi_j = i) = \mu(A_i)/\mu(E)$. Par la propriété de raréfaction énoncée à la proposition 3.1, les variables aléatoires $M(A_i), 1 \leq i \leq k$ sont indépendantes, de lois respectives $\mathcal{P}(\mu(E)\mu(A_j)/\mu(E)), 1 \leq j \leq k$. Donc M est bien une mesure aléatoire de Poisson d'intensité μ.

Cas où $\mu(E) = \infty$. Comme μ est σ-finie, on peut trouver une partition de E par des ensembles mesurables $E_n, n \geq 1$, et tous de mesure $\mu(E_n)$ finie. Par le cas précédent, on peut alors se donner une famille de mesures de Poisson indépendantes M_n sur E d'intensités respectives finies $\mu(\cdot \cap E_n)$, pour $n \geq 1$. Nous afirmons alors que la mesure

$$M = \sum_{n \geq 1} M_n$$

est une mesure aléatoire de Poisson d'intensité μ. En effet, constatons tout d'abord que $M_n(E \setminus E_n) = 0$ p.s. puisque sa loi est $\mathcal{P}(0)$. Si maintenant $A_1, \dots, A_k$ sont dans $\mathcal{E}$ et sont deux-à-deux disjoints, alors les ensembles $(A_i \cap E_n, 1 \leq i \leq k, n \geq 1)$ sont également dans $\mathcal{E}$ et deux-à-deux disjoints. Comme M_n est une mesure aléatoire de Poisson d'intensité $\mu(\cdot \cap E_n)$, les variables aléatoires $M_n(A_i) = M_n(A_i \cap E_n), 1 \leq i \leq k$ sont donc indépendantes, respectivement de loi $\mathcal{P}(\mu(A_i \cap E_n))$, et comme les mesures $M_n, n \geq 1$ sont indépendantes, on en déduit que les variables aléatoires $M_n(A_i), 1 \leq i \leq k, n \geq 1$ sont indépendantes. Finalement, la propriété de conservation de l'indépendance par regroupement par paquets, groupée avec le propriété de superposition de la loi de Poisson énoncée en proposition 3.1, implique que les variables

$$M(A_i) = \sum_{n\geq 1} M_n(A_i)\,, \qquad 1 \leq i \leq k$$

sont indépendantes et respectivement de loi $\mathcal{P}(\sum_{n\geq 1} \mu(A_i \cap E_n)) = \mathcal{P}(\mu(A_i))$. Ceci montre que M est une mesure aléatoire de Poisson d'intensité μ, comme voulu. □

En utilisant la construction de la preuve précédente, ainsi que l'unicité en loi d'une mesure de Poisson d'intensité donnée, nous obtenons une propriété très utile des mesures de Poisson.

Proposition 3.4 *Soit M une mesure aléatoire de Poisson sur E d'intensité μ, et soit $A \in \mathcal{E}$ tel que $\mu(A) < \infty$. Alors $M(A)$ a pour loi $\mathcal{P}(\mu(A))$, et sachant $M(A) = k$, la restriction $M|_A$ a même loi que $\sum_{i=1}^k \delta_{X_i}$, où $(X_1, X_2, \dots, X_k)$ sont des variables aléatoires indépendantes de loi $\mu(\cdot \mid A) = \mu(\cdot \cap A)/\mu(A)$.*

De plus, si $A_1, \dots, A_k \in \mathcal{E}$ sont disjoints, alors les restriction $M|_{A_i}, 1 \leq i \leq k$ sont indépendantes.

Enfin, toute mesure aléatoire de Poisson peut s'écrire sous la forme

$$M(\mathrm{d}x) = \sum_{i\in I} \delta_{x_i}(\mathrm{d}x)\,,$$

où I est un ensemble d'indices fini ou dénombrable, et les $x_i, i \in I$ sont des variables aléatoires à valeurs dans E.

3.2 Fonctionnelle de Laplace d'une mesure aléatoire de Poisson

La formule de la fonctionnelle de Laplace (parfois appelée formule de Campbell, par exemple dans [8]) donne une caractérisation extrêmement utile des mesures aléatoires de Poisson. Pour pouvoir l'énoncer, nous devons faire une remarque préliminaire sur l'intégrale d'une fonction par rapport à une mesure de Poisson. Par défaut, dans toute la suite, les mesures considérées sur l'espace $(E, \mathcal{E})$ seront supposées σ-finies, mais nous omettrons de le mentionner.

Lemme 3.1 *Soit M une variable aléatoire à valeurs dans $(E^*, \mathcal{E}^*)$. Alors, pour toute fonction mesurable $f : E \to \mathbb{R}_+$, la quantité*

$$M(f) = \int_E f(x) M(\mathrm{d}x)$$

définit une variable aléatoire.

Si M est une mesure aléatoire de Poisson d'intensité μ sur $(E, \mathcal{E})$, alors

$$\mathbb{E}[M(f)] = \mu(f)\,.$$

De plus, pour tout $f \in L^1(E, \mathcal{E}, \mu)$, on a que $f \in L^1(E, \mathcal{E}, M)$ presque sûrement, et $M(f)$ est une variable aléatoire d'espérance $\mu(f)$.

Preuve. Si la fonction f est de la forme $f = \mathbb{1}_A$ avec $A \in \mathcal{E}$, alors $M(f) = M(A) = \pi^A \circ M$ est une variable aléatoire par définition de la tribu $\mathcal{E}^*$. Donc ceci reste valable pour toute fonction f étagée, par linéarité. Si f est maintenant mesurable positive quelconque, soit f_n une suite croissante de fonctions étagées convergeant vers f ponctuellement. Alors par convergence monotone $M(f_n)$ converge vers $M(f)$, qui est donc une variable aléatoire (comme limite de variables aléatoires).

Si maintenant M est une mesure aléatoire de Poisson d'intensité μ alors si $f = \mathbb{1}_A$ on a que $M(A)$ a pour loi $\mathcal{P}(\mu(A))$ par définition, et son espérance est bien $\mu(A) = \mu(f)$. De plus, par convergence monotone, on obtient que $\mathbb{E}[M(f)] = \mu(f)$ en passant à la limite dans $\mathbb{E}[M(f_n)] = \mu(f_n)$. Enfin, si $f \in L^1(E, \mathcal{E}, \mu)$, on a $\mathbb{E}[M(|f|)] = \mu(|f|) < \infty$, d'où l'on conclut que $f \in L^1(E, \mathcal{E}, M)$ presque surement. Donc $M(f) = M(f_+) - M(f_-)$ est une variable aléatoire d'espérance $\mu(f_+) - \mu(f_-) = \mu(f)$. Ceci conclut la preuve. □

Remarques. 1. Si M est une variable aléatoire à valeurs dans $(E^*, \mathcal{E}^*)$, on définit une mesure m sur $(E, \mathcal{E})$ par la formule $m(A) = \mathbb{E}[M(A)]$. Cette mesure s'appelle *mesure d'intensité* de M. On voit qu'avec cette définition, la mesure d'intensité d'une mesure aléatoire de Poisson d'intensité μ est... la mesure μ, et la terminologie est cohérente. L'énoncé précédent reste alors vrai si l'on remplace μ par m dans l'énoncé.

2. Remarquons que l'argument de limite monotone donné dans la preuve montre que pour tout $f : E \to \mathbb{R}$ mesurable positive, ou dans $L^1(E, \mathcal{E}, \mu)$, la variable aléatoire $M(f)$ est mesurable par rapport à la tribu $\sigma(M)$. En particulier, si M est une mesure aléatoire de Poisson et $A_1, A_2, \dots, A_k$ sont des ensembles de $\mathcal{E}$ disjoints, les familles $\{M|_{A_i}(f) : f : E \to \mathbb{R}_+\}$ sont indépendantes, puisque les mesures aléatoires $M|_{A_i}, 1 \le i \le k$ sont indépendantes, par la proposition 3.4.

Nous énonçons maintenant la formule de la fonctionnelle de Laplace d'une mesure aléatoire de Poisson.

Proposition 3.5 *Soit M une mesure aléatoire de Poisson d'intensité μ sur $(E, \mathcal{E})$. Alors pour toute fonction $f : E \to \mathbb{R}_+$ mesurable, on a*

$$E[\exp(-M(f))] = \exp\left(-\int_E \mu(\mathrm{d}x)(1 - \exp(-f(x)))\right).$$

De plus, si $f : E \to \mathbb{R}$ est dans $L^1(E, \mathcal{E}, \mu)$, alors

$$E[\exp(\mathrm{i}M(f))] = \exp\left(\int_E \mu(\mathrm{d}x)(\exp(\mathrm{i}f(x)) - 1)\right).$$

Preuve. On peut montrer cette proposition en commençant par le cas des fonctions étagées et par limite monotone, un peu comme pour le lemme 3.1, ce que nous laissons comme exercice au lecteur (voir aussi la preuve de le proposition 3.6 ci-dessous).

On peut également raisonner de la façon suivante. Supposons dans un premier temps que f est mesurable et positive. Soit $\{E_n, n \ge 1\}$ une partition mesurable de E telle que $\mu(E_n) < \infty$ pour tout $n \ge 1$. Le nombre $N_n = M(E_n)$ d'atomes de M appartement à E_n a pour loi $\mathcal{P}(\mu(E_n))$, et par la proposition 3.4, conditionnellement à $N_n = k$, on peut supposer que ces atomes sont donnés par k variables aléatoires indépendantes de loi $\mu(\cdot \mid E_n)$. De ce fait,

$$\begin{aligned} E[\exp(-M(f\mathbb{1}_{E_n}))] &= \sum_{k=0}^{\infty} e^{-\mu(E_n)} \frac{\mu(E_n)^k}{k!} \left(\int_{E_n} \frac{\mu(\mathrm{d}x)}{\mu(E_n)} e^{-f(x)} \right)^k \\ &= \exp\left(- \int_{E_n} \mu(\mathrm{d}x)(1 - \exp(-f(x))) \right) \end{aligned}$$

Comme les variables $M(f\mathbb{1}_{E_n}) = M|_{E_n}(f)$ sont indépendantes (voir la remarque après la preuve du lemme 3.1), on peut prendre le produit sur n et obtenir le résultat voulu puisque $\sum_{n\geq 1} \mathbb{1}_{E_n} = 1$ et par convergence monotone.

Dans le cas où $f \in L^1(E, \mathcal{E}, \mu)$, on raisonne de façon similaire, en établissant le formule pour le fonction $f\mathbb{1}_{E_n}$ au lieu de f. Cette partie du raisonnement se traite exactement de la même façon. Pour conclure, nous devons montrer que $\int_{A_n} \mu(\mathrm{d}x)(e^{\mathrm{i}f(x)} - 1)$ converge vers $\int_E \mu(\mathrm{d}x)(e^{\mathrm{i}f(x)} - 1)$, où $A_n = E_0 \cup \cdots \cup E_n$. Mais $|e^{\mathrm{i}f(x)} - 1| \leq |f(x)|$, et la fonction dominante est intégrable par rapport à μ. On applique cette fois le théorème de convergence dominée, qui donne le résultat. $\square$

Remarquons que, quitte à remplacer f par af dans l'une des deux formules (avec $a > 0$ pour la première), puis en dérivant par rapport à a et en faisant tendre a vers 0, on réobtient la formule du premier moment

$$E[M(f)] = \int_E f(x)\mu(\mathrm{d}x)\,,$$

valable pour f mesurable positive, ou pour $f \in L^1(E, \mathcal{E}, \mu)$. Si l'on dérive une seconde fois avant de faire tendre a vers 0, on obtient la formule du second moment :

$$\mathrm{Var}\, M(f) = \int_E f(x)^2 \mu(\mathrm{d}x)\,,$$

valable pour toute fonction $f : E \to \mathbb{R}$ mesurable.

Nous faisons maintenant une remarque facile mais importante.

Proposition 3.6 *Si M est une variable aléatoire à valeurs dans $(E^*, \mathcal{E}^*)$ qui vérifie la première formule de la proposition 3.5 (pour toute fonction f mesurable positive), alors M est une mesure aléatoire de Poisson d'intensité μ.*

Preuve. En prenant f de la forme $\sum_{i=1}^k \lambda_i \mathbb{1}_{A_i}$, où les λ_i sont des réels positifs et les A_i sont des ensembles dans $\mathcal{E}$ deux-à-deux disjoints, et de μ-mesure finie. La formule donne

$$\mathbb{E}\left[\exp\left(- \sum_{i=1}^k \lambda_i M(A_i) \right) \right] = \prod_{i=1}^k \exp\left(-\mu(A_i)(1 - e^{-\lambda_i}) \right)\,.$$

On reconnaît à droite le produit des transformées de Laplace (évaluées en λ_i respectivement) de variables aléatoires de loi de Poisson de paramètres $\mu(A_i)$ respectifs. D'où le résultat. $\square$

3.3 Propriétés de stabilité des mesures aléatoires de Poisson

Soit μ une mesure σ-finie sur un espace mesurable $(E, \mathcal{E})$. On se propose de montrer quelques propriétés de stabilité des mesures de Poisson.

$\boxed{\text{Image}}$ Soit M une mesure de Poisson d'intensité μ. Soit $(F, \mathcal{F})$ une autre espace mesurable, et $f : E \to F$ une fonction mesurable. Alors l'image f_*M de M par f est une mesure de Poisson d'intensité $f_*\mu$.

Preuve. On a $f_*M(A) = M(f^{-1}(A))$. Si les ensembles $A_1, \ldots, A_k$ sont dans $\mathcal{F}$ et deux-à-deux disjoints, alors $f^{-1}(A_1), \ldots, f^{-1}(A_k)$ sont dans $\mathcal{E}$ et deux-à-deux disjoints, de plus, $f^*\mu(A_i) = \mu(f^{-1}(A_i))$ par définition. Donc le résultat est évident par définition d'une mesure de Poisson. □

$\boxed{\text{Superposition}}$ Soit μ' une autre mesure σ-finie sur $(E, \mathcal{E})$, et M, M' deux mesures de Poisson d'intensités respectives μ, μ'. On suppose M, M' indépendantes. Alors $M + M'$ est une mesure de Poisson d'intensité $\mu + \mu'$.

Preuve. Soit $f : E \to \mathbb{R}_+$ mesurable. Alors en utilisant l'indépendance et en appliquant la formule de la fonctionnelle de Laplace,

$$\begin{aligned}
\mathbb{E}[\exp(-(M+M')(f))] &= \mathbb{E}[\exp(-M(f))]\mathbb{E}[\exp(-M'(f))] \\
&= \exp\left(\int_E \mu(\mathrm{d}x)(1-e^{-f(x)})\right) \\
&\quad \times \exp\left(\int_E \mu'(\mathrm{d}x)(1-e^{-f(x)})\right) \\
&= \exp\left(\int_E (\mu+\mu')(\mathrm{d}x)(1-e^{-f(x)})\right),
\end{aligned}$$

et l'on conclut par la proposition 3.6. □

$\boxed{\text{Restriction}}$ Soit $A_1, A_2, \ldots, A_k$ dans $\mathcal{E}$, deux-à-deux disjoints. Alors les mesures $M|_{A_1}, \ldots, M|_{A_k}$ sont des mesures aléatoires de Poisson indépendantes d'intensités respectives $\mu(\cdot \cap A_i), 1 \le i \le k$.

Preuve. Nous l'avons vue en proposition 3.4.

$\boxed{\text{Marquage}}$ Soit $(X_n, n \ge 1)$ une suite de variables aléatoires , et N une variable aléatoire à valeurs dans $\mathbb{N} \cup \{\infty\}$, telles que $M = \sum_{1 \le i \le N} \delta_{X_i}$ est une mesure de Poisson d'intensité μ. On se donne une autre suite $(Y_n, n \ge 1)$ de variables aléatoires, que l'on suppose i.i.d. de loi μ' sur un espace $(F, \mathcal{F})$, et indépendante de $\sigma(N) \vee \sigma(X_n, n \ge 0)$. Alors la mesure $M' = \sum_{1 \le i \le N} \delta_{(X_i, Y_i)}$ est une mesure de Poisson sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$, d'intensité $\mu \otimes \mu'$.

Preuve. Soit $f : E \times F \to R_+$ mesurable. On calcule

$$\begin{aligned}
\mathbb{E}[\exp(-M'(f)) \,|\, N, X_1, X_2, \ldots] &= \mathbb{E}\left[\exp\left(-\sum_{1 \le i \le N} f(X_n, Y_n)\right) \,|\, N, X_1, X_2, \ldots\right] \\
&= \prod_{i=1}^{N} \int_F \mu'(\mathrm{d}y) \exp(-f(X_n, y)) \\
&= \exp\left(-\sum_{i=1}^{N} -\log \int_F \mu'(\mathrm{d}y) \exp(-f(X_n, y))\right) \\
&= \exp(-M(F)) ,
\end{aligned}$$

où $F(x) = -\log \int_F \mu'(\mathrm{d}y) \exp(-f(x, y))$ est une fonction mesurable positive. On prend l'espérance et on applique la formule de la fonctionnelle de Laplace pour trouver

$$\begin{aligned}
\mathbb{E}[\exp(-M'(f))] &= \exp\left(-\int_E \mathrm{d}\mu(x) \left(1 - \int_F \mu'(\mathrm{d}y) \exp(-f(x, y))\right)\right) \\
&= \exp\left(-\int_E \mathrm{d}\mu(x)\mathrm{d}\mu'(y) \,(1 - \exp(-f(x, y)))\right)
\end{aligned}$$

d'où le résultat. □

$\boxed{\text{Raréfaction}}$ Soit $(X_n, n \ge 1)$ une suite de variables aléatoires , et N une variable aléatoire à valeurs dans $\mathbb{N} \cup \{\infty\}$, telles que $M = \sum_{1 \le i \le N} \delta_{X_i}$ est une mesure de Poisson d'intensité μ, et $(Y_i, i \ge 1)$ une suite de variables aléatoires i.i.d. de loi de Bernoulli de paramètre $p \in [0, 1]$, indépendante de $\sigma(N) \vee \sigma(X_n, n \ge 0)$. Alors les mesures aléatoires

$$M' = \sum_{1 \le i \le N, Y_i = 1} \delta_{X_i} , \qquad M'' = \sum_{1 \le i \le N, Y_i = 0} \delta_{X_i}$$

sont deux mesures de Poisson indépendantes d'intensités $p\mu$ et $(1 - p)\mu$ respectivement.

La preuve est laissée en exercice : on pensera à appliquer les propriétés de marquage et de restriction ci-dessus.

3.4 Processus de Poisson ponctuels

Montrons maintenant comment on peut utiliser les mesures aléatoires de Poisson pour définir certains processus stochastiques. Soit $(E, \mathcal{E})$ un espace mesurable, et μ une mesure σ-finie sur $(E, \mathcal{E})$. Considérons la mesure produit $\mu'(\mathrm{d}t\mathrm{d}x) = \mathrm{d}t \otimes \mu(\mathrm{d}x)$ sur $\mathbb{R}_+ \times E$, où $\mathrm{d}t$ est la mesure de Lebesgue sur $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. Autrement dit, μ' est l'unique mesure telle que $\mu'([0, t] \times A) = t\mu(A)$ for $t \ge 0$ and $A \in \mathcal{E}$.

Définition 3.2 *Une mesure aléatoire de Poisson $M(\mathrm{d}t\mathrm{d}x)$ d'intensité μ de la forme ci-dessus est appelée un* processus de Poisson ponctuel *d'intensité μ.*

Pour justifier cette dénomination, réécrivons M sous la forme $\sum_{i\in I}\delta_{(t_i,x_i)}$, où I est un ensemble d'indice dénombrable (I est toujours infini puisque μ' est une mesure infinie — sauf cas trivial où $\mu = 0$). On pose alors

$$X_t = \begin{cases} x_i & \text{s'il existe } i\in I \text{ tel que } t_i = t \\ \dagger & \text{sinon,} \end{cases}$$

où $\dagger$ est un *point-cimetière*, qui n'appartient pas à E (on voit alors μ comme une mesure sur $E\cup\{\dagger\}$ telle que $\mu(\{\dagger\}) = 0$), ou tout autre point de E qui n'est pas un atome de μ, fixé par convention. Remarquons que cette définition est bien posée, en vertu du lemme suivant, qui implique que p.s., on ne peut pas trouver de temps $t \geq 0$ pour lequel il existerait deux indices $i \neq j$ de I avec $t_i = t_j$.

Lemme 3.2 *Presque sûrement, pour tout $t \geq 0$, on a que $M(\{t\}\times E)\in\{0,1\}$.*

Preuve. Soit $\{E_n, n\geq 1\}$ une partition mesurable de E telle que $\mu(E_n) < \infty$ pour tout $n\geq 1$. Alors les ensembles $\{[k,k+1[\times E_n, k\geq 0, n\geq 1\}$ forment une partition de $\mathbb{R}_+\times E$ en ensembles mesurables de μ'-mesures finies, puisque

$$\mu'([k,k+1[\times E_n) = \mu(E_n)\,.$$

Par la proposition 3.4, les quantités $N_{kn} = M([k,k+1[\times E_n)$ sont toutes finies p.s., et conditionnellement à ces variables, les atomes de $M|_{[k,k+1[\times E_n}$ peuvent être vus comme N_{kn} variables aléatoires i.i.d. de loi $\mu'|_{[k,k+1[\times E_n}$, indépendantes lorsque n,k varie. La première composante de ces variables est donc uniforme sur $[k,k+1[$, et donc l'événement qu'il existe deux de ces variables (pour le même indice n ou pour deux indices différents) ayant la même première composante est de probabilité nulle. Ceci implique le lemme. □

C'est la famille de variables aléatoires $(e_t, t\geq 0)$ qui est le plus souvent appelée *processus de Poisson ponctuel* d'intensité μ. Il convient de prendre garde au fait que ce genre de processus est très différent de ceux que nous avons considérés au chapitre 2. Ils ne sont évidemment pas continus, mais pire, le processus constant égal à $\dagger$ est toujours une modification de $(e_t, t\geq 0)$! En effet, à t fixé, comme $\mu'(\{t\}\times E) = 0$, la probabilité que $e_t \neq \dagger$ est nulle... Ainsi, les temps t tels que $e_t \neq \dagger$ forment un ensemble de temps ≪ exceptionels ≫. Ceci montre aussi que la tribu produit n'est pas la bonne notion pour comprendre le processus $(e_t, t\geq 0)$: le ≪ bon ≫ objet est la mesure aléatoire de Poisson M sous-jacente.

Nous allons maintenant utiliser les mesures ponctuelles de Poisson pour construire de nouveaux processus. Rappelons une définition déjà rencontrée dans l'introduction.

Définition 3.3 *On dit que la famille $(X_t, t\geq 0)$ de variables aléatoires indexées par le paramètre $t\in\mathbb{R}_+$ est un processus de Lévy si les trois propriétés suivantes sont vérifiées :*

- *$X_0 = 0$ p.s.,*
- *presque-sûrement, la fonction $t\mapsto X_t$ est continue à droite et admet des limites à gauche en tout point,*

– *pour tout* $0 = t_0 \leq t_1 \leq \ldots \leq t_k$, *les variables* $(X_{t_i} - X_{t_{i-1}}, 1 \leq i \leq k)$ *sont indépendantes, respectivement de même loi que* $X_{t_i - t_{i-1}}$.

Proposition 3.7 *Soit* $M(\mathrm{d}t\mathrm{d}x)$ *un processus ponctuel de Poisson sur d'intensité* μ *sur un espace mesurable* $(E, \mathcal{E})$. *Si* $f \in L^1(E, \mathcal{E}, \mu)$, *alors le processus*

$$N_t^f = \int_{[0,t]\times E} f(x) M(\mathrm{d}s, \mathrm{d}x)\,, \quad t \geq 0\,,$$

est un processus de Lévy. De plus, le processus

$$M_t^f = \int_{[0,t]\times E} f(x) M(\mathrm{d}s, \mathrm{d}x) - t \int_E f(x)\mu(\mathrm{d}x)\,, \quad t \geq 0,$$

est une martingale par rapport à la filtration $(\mathcal{F}_t)$, *où* $\mathcal{F}_t = \sigma(M|_{[0,t]\times E}), t \geq 0$.

Si de plus $f \in L^2(E, \mathcal{E}, \mu)$, *le processus*

$$(M_t^f)^2 - t \int_E f(x)^2 \mu(\mathrm{d}x),$$

est une martingale par rapport à la filtration $(\mathcal{F}_t)$.

Preuve. Le caractère càdlàg du processus N^f provient directement de sa définition, et du fait que l'application $(t, x) \mapsto f(x)$ est dans $L^1(\mathbb{R}_+ \times E, \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{E}, M|_{[0,T]\times E})$ pour tout $T > 0$ fini fixé, ce qui vient de l'hypothèse $f \in L^1(E, \mathcal{E}, \mu)$ et du lemme 3.1. On conclut par théorème de convergence dominée que N^f est continu à droite et admet les limites à gauche

$$N_{t_-}^f = \int_{[0,t[\times E} f(x) M(\mathrm{d}s\mathrm{d}x)\,.$$

Ensuite, si $0 \leq s \leq t$, alors $N_t^f - N_s^f = \int_{(s,t]\times E} f(x) M(\mathrm{d}u, \mathrm{d}x)$. De plus, par un argument d'invariance par translation évident, la mesure $M(\mathrm{d}u, \mathrm{d}x)\mathbb{1}_{\{u\in]s,t]\}}$ a la même loi que l'image de $M(\mathrm{d}u, \mathrm{d}x)\mathbb{1}_{\{u\in]0,t-s]\}}$ par l'application $(u, x) \mapsto (s+u, x)$ de $\mathbb{R}_+ \times E$ dans lui-même. Cette mesure est de plus indépendante de $M(\mathrm{d}u, \mathrm{d}x)\mathbb{1}_{\{u\in[0,s]\}}$ par la proposition 3.4. De ce fait, N^f a des accroissements indépendants et stationnaires. Le fait que M^f soit une martingale est une conséquence immédiate de la formule du premier moment, et du fait que $N_t^f - N_s^f$ est indépendant de $\mathcal{F}_s$. La dernière partie de l'énoncé vient de ce que

$$(M_t^f)^2 = (M_t^f - M_s^f + M_s^f)^2 = (M_t^f - M_s^f)^2 + 2M_s^f(M_t^f - M_s^f) + (M_s^f)^2\,.$$

Si l'on prend l'espérance conditionnelle sachant $\mathcal{F}_s$, la seconde partie disparaît par le fait que $M_t^f - M_s^f$ est indépendant de $\mathcal{F}_s$ et de moyenne nulle, tandis que la première partie contribue d'un terme $(t-s)\int_E f(x)^2\mu(\mathrm{d}x)$ par la même propriété d'indépendance et par la formule du second moment. $\square$

Si l'on représente $M(\mathrm{d}t\mathrm{d}x)$ sous la forme d'un processus, $(e_t, t \geq 0)$, de la façon décrite plus haut, alors le processus N^f admet la représentation alternative

$$N_t^f = \sum_{0 \leq s \leq t} f(e_s)\,, \qquad t \geq 0\,,$$

où par convention $f(\partial) = 0$, de sorte que la somme est en fait dénombrable. Souvent on note cette somme

$$N_t^f = \sum_{0 < s \leq t} f(e_s)$$

pour insister sur le fait que $N_0^f = 0$, ce qui vient du fait que p.s. $e_0 = \partial$, ou encore, $M(\{0\} \times E) = 0$.

3.4.1 Exemple : le processus de Poisson homogène

Soit $X_1, X_2, \ldots$ une suite de variables aléatoires exponentielles de paramètre $\theta > 0$. On définit $0 = T_0 \leq T_1 \leq \ldots$ par $T_n = X_1 + \ldots + X_n$. Soit

$$N_t^\theta = \sum_{n=1}^{\infty} \mathbb{1}_{\{T_n \leq t\}}\,, \qquad t \geq 0,$$

le processus càdlàg qui compte le nombre de temps T_n qui sont plus petits que t. Le processus $(N_t^\theta, t \geq 0)$ est appelé processus de Poisson homogène de paramètre θ. Ce processus est en effet naturellement décrit en termes de mesures de Poisson.

Proposition 3.8 *Soit $\theta > 0$, et soit M une mesure aléatoire de Poisson d'intensité $\theta\mathrm{d}t$ sur $\mathbb{R}_+$. Alors le processus*

$$M([0,t])\,, \qquad t \geq 0$$

est un processus de Poisson homogène d'intensité θ.

Nous donnons une idée de preuve, laissant les détails en exercice au lecteur. Tout d'abord, avec la définition de N^θ, on montre que ce dernier processus est à accroissements indépendants et stationnaires, et ont des lois de Poisson, ce qui peut se faire directement par un calcul de la loi jointe des accroissements de N^θ. On en déduit que la mesure de Stieltjes $\mathrm{d}N^\theta$ est une mesure de Poisson d'intensité $\theta\mathrm{d}t$ sur $\mathbb{R}_+$, ce qui est une autre façon d'écrire la proposition 3.8.

3.4.2 Exemple : processus de Poisson composés

Une extension naturelle du processus de Poisson homogène est donnée par la construction suivante.

Définition 3.4 *Soit ν une mesure positive finie sur $\mathbb{R}$ ou $\mathbb{R}^d$. Un processus de Poisson d'intensité ν est un processus de la forme*

$$N_t^\nu = \int_{[0,t]\times\mathbb{R}} x M(\mathrm{d}s, \mathrm{d}x)\,, \qquad t \geq 0,$$

où M est une mesure aléatoire de Poisson d'intensité $\mathrm{d}t \otimes \nu(\mathrm{d}x)$.

Si l'on représente la mesure M par un processus $(e_t, t \geq 0)$ comme au début de la section, alors on a

$$N_t^\nu = \sum_{0<s\leq t} e_s \,, \qquad t \geq 0 \,.$$

On remarquera que dans cette situation, p.s., pour tout $T > 0$, le nombre $\#\{s \in [0,T] : e_s \neq \partial\}$ est fini. C'est ce qui fait que cette somme est toujours bien définie, même si la fonction $x \mapsto |x|$ n'est pas dans $L^1(\mathbb{R}^d, \mathcal{B}_{R^d}, \nu)$.

On peut donner la représentation alternative suivante de ces processus. Pour éviter les trivialités, on suppose que ν est non nulle. Soit N^θ un processus de Poisson homogène de paramètre $\theta = \nu(\mathbb{R}^d)$, et dont les temps de saut sont donnés par $0 < T_1 < T_2 < \ldots$ — de sorte que $\sum_{n\geq 1} \delta_{T_n}$ est une mesure aléatoire de Poisson d'intensité $\theta \mathrm{d}t$ sur $\mathbb{R}_+$. Soit également $Y_1, Y_2, \ldots$ une suite de variables aléatoires i.i.d., indépendante de N^θ et de loi commune ν/θ. Alors, le processus

$$\sum_{n\geq 1} Y_n \mathbb{1}_{\{T_n \leq t\}} \,, \qquad t \geq 0 \,,$$

est un processus de Poisson composé d'intensité ν.

Ce résultat provient de la propriété de marquage des mesures aléatoires de Poisson, mentionnée en section 3.3, et qui implique, avec les notations ci-dessus, $M' = \sum_{i\in I} \delta_{(T_i, Y_i)}$ est une mesure de Poisson aléatoire d'intensité $\theta \mathrm{d}t \otimes \nu/\theta = \mathrm{d}t \otimes \nu$.

Nous désignons la loi de N_1^ν par $\mathrm{PC}(\nu)$, et on l'appelle loi de Poisson composée d'intensité ν. On peut l'écrire sous la forme

$$\mathrm{PC}(\nu) = \sum_{n\geq 0} e^{-\nu(\mathbb{R}^d)} \frac{\nu^{*n}}{n!} \,,$$

où ν^{*n} est la convolée n-ième de ν avec elle-même. La fonction caractéristique de $\mathrm{PC}(\nu)$ est donnée par

$$\phi_{\mathrm{PC}(\nu)}(u) = \exp(-\nu(\mathbb{R}^d)(1 - \phi_{\nu/\nu(\mathbb{R}^d)}(u))),$$

où $\phi_{\nu/\nu(\mathbb{R}^d)}$ est la fonction caractéristique de $\nu/\nu(\mathbb{R}^d)$. On peut encore l'écrire

$$\phi_{\mathrm{PC}(\nu)}(u) = \exp\left(\int_{\mathbb{R}^d} \nu(\mathrm{d}x) \left(\exp(\mathrm{i}u \cdot x) - 1\right) \right) \,,$$

ce qui est le premier avatar de la formule de Lévy-Khintchine, qui est le premier objet d'étude du chapitre suivant.

3.5 Exercices pour le chapitre 3

Exercice 1 — Une application de la stabilité des mesures de Poisson
Soit $\theta > 0$ et $\ldots < T_{-2} < T_{-1} < T_0 < T_1 < T_2 < \ldots$ les atomes d'une mesure de Poisson sur $\mathbb{R}$, d'intensité $\theta \mathrm{d}t$. On peut par exemple poser T_0 le premier atome positif de cette mesure, ce qui détermine la numérotation des autres atomes de façon unique.

On se donne une suite $(Y_n, n \in \mathbb{Z})$ i.i.d. de variables aléatoires réelles, indépendante de la mesure de Poisson. Montrer que $\sum_{n\in\mathbb{Z}} \delta_{T_n+Y_n}$ est encore une mesure de Poisson d'intensité $\theta \mathrm{d}t$.

Exercice 2 — Caractérisation du processus de Poisson
Soit $(N_t, t \geq 0)$ un processus à accroissements indépendants et stationnaires, croissant, continu à droite, à valeurs entières, vérifiant $N_0 = 0$ et effectuant uniquement des sauts d'amplitude 1. Montrer que le premier saut de N a une loi exponentielle, et en appliquant la propriété de Markov forte (qu'on justifiera) en déduire que N est un processus de Poisson.

Exercice 3 — Le diable est dans les détails
Compléter la preuve de la proposition 3.8.

Exercice 4 — Calculs sur le processus de Poisson
Vérifier que si N est un processus de Poisson d'intensité θ, alors $(N_t - \theta t, t \geq 0)$ et $((N_t - \theta t)^2 - \theta t, t \geq 0)$ sont des martingales.

Exercice 5 — Amélioration de la stabilité par marquage
Un noyau markovien entre deux espaces mesurés $(E, \mathcal{E})$ et $(F, \mathcal{F})$ est une application K de $E \times \mathcal{F}$ dans $[0, 1]$ telle que

- pour tout $x \in E$, l'application $A \mapsto K(x, A)$ est une mesure de probabilités sur $(F, \mathcal{F})$
- pour tout $A \in \mathcal{F}$, l'application $x \mapsto K(x, A)$ est mesurable de $(E, \mathcal{E})$ dans $[0, 1]$.

Si μ est une mesure σ-finie sur $(E, \mathcal{E})$, et si K est un noyau markovien entre $(E, \mathcal{E})$ et $(F, \mathcal{F})$, montrer qu'il existe une unique mesure σ-finie $\mu \rtimes K$ sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$ telle que

$$\mu \rtimes K(A \times B) = \int_A \mu(\mathrm{d}x) K(x, B)\,, \qquad A \in \mathcal{E}, B \in \mathcal{F}\,.$$

Soit $(X_n, n \geq 0)$ une suite de variables aléatoires à valeurs dans E, et N une variable aléatoire à valeurs dans $\mathbb{Z}_+ \cup \{\infty\}$, telles que $\sum_{n=1}^N \delta_{X_n}$ est une mesure de Poisson d'intensité μ sur E. Soit $(Y_n, n \geq 1)$ une suite de variables aléatoires à valeurs dans F telle que conditionnellement à $(X_n, n \geq 1)$, les variables $(Y_n, n \geq 1)$ sont indépendantes et de lois respectives $K(X_n, \cdot)$. Montrer que

$$\sum_{n=1}^N \delta_{(X_n, Y_n)}$$

est une mesure aléatoire de Poisson d'intensité $\mu \rtimes K$.

Exercice 6 — Temps de saut d'un processus de Poisson ponctuel
Soit μ une mesure σ-finie sur un espace mesurable $(E, \mathcal{E})$, et $(X_t, t \geq 0)$ un processus de Poisson d'intensité μ. On note $\mathcal{T} = \{t \geq 0 : X_t \neq \dagger\}$. Montrer que p.s. tous les points de $\mathcal{T}$ sont isolés si $\mu(E) < \infty$, et que p.s. $\mathcal{T}$ est dénombrable dense si $\mu(E) = \infty$.

Exercice 7 — Mesure de Poisson dans une bande
Soit M une mesure de Poisson dans $[0,1] \times]0, \infty[$, d'intensité

$$\nu(\mathrm{d}t\mathrm{d}x) = \mathrm{d}t \mathbb{1}_{\{0 \leq t \leq 1\}} \frac{\mathrm{d}x}{x^3} \mathbb{1}_{\{x > 0\}}$$

(i) Montrer que p.s. cette mesure est de masse infinie.

(ii) On note $((T_n, X_n), n \geq 1)$ les atomes de M, et on pose par récurrence

$$m_1 = \sup\{X_n, n \geq 1\}, \qquad m_{k+1} = \sup\{X_n : n \geq 1, X_n < m_k\}.$$

Montrer que ceci définit bien p.s. une suite strictement décroissante, et que pour tout $k \geq 1$ il existe un unique N_k tel que $X_{N_k} = m_k$.

(iii) Calculer la loi de m_1.

(iv) Montrer que $M' = \sum_{k \geq 1} \delta_{m_k}$ est une mesure de Poisson, dont on déterminera l'intensité.

(v) Trouver une fonction $f :]0, \infty[\to]0, \infty[$ bijective décroissante telle que $\sum_{k \geq 1} \delta_{f(m_k)}$ est une mesure de Poisson d'intensité $\mathrm{d}t \mathbb{1}_{\{t \geq 0\}}$. On utilisera la stabilité par image.

(vi) En se ramenant par la question précédente au cas du processus de Poisson standard, montrer que conditionnellement à m_1, la mesure $M' - \delta_{m_1}$ est encore une mesure de Poisson, dont on déterminera l'intensité.

Exercice 8 — La percolation Poissonnienne
Soit M une mesure de Poisson dans $\mathbb{R}^d$ d'intensité la mesure de Lebesgue. On note $M = \sum_{n \geq 1} \delta_{X_n}$. Soit une suite $(R_n, n \geq 1)$ de variables aléatoires i.i.d. indépendante de M, à valeurs dans $]0, \infty[$. On pose B_n la boule de centre X_n et de rayon R_n. Ces dernières couvrent le sous-ensemble aléatoire $A = \bigcup_{n \geq 1} B_n$ de $\mathbb{R}^d$, dont on veut étudier quelques propriétés.

(i) Soit $x \in \mathbb{R}^d$. Déterminer la loi du nombre de boules qui recouvrent x, donné par

$$N(x) = \mathrm{Card}\, \{n \geq 1 : x \in B_n\}.$$

On pourra se ramener au cas $x = 0$, puis utiliser la propriété de stabilité des mesures de Poisson par marquage.

(ii) À partir de maintenant, on suppose que $E[(R_1)^d] = \infty$. Montrer que pour tout x, on a que $x \in A$ presque-sûrement. Déduire que A est partout dense dans $\mathbb{R}^d$ presque-sûrement.

(iii) En considérant les variables aléatoires $(R_n - 1)^+$, et en utilisant la propriété de raréfaction, déduire que $A = \mathbb{R}^d$ p.s.

Exercice 9 — Autoroutes et mesures de Poisson
On considère une autoroute, assimilée à la droite réelle $\mathbb{R}$. Soit μ une loi de probabilités sur $\mathbb{R}$, telle que $v_m = \int_{\mathbb{R}} |v|\mu(\mathrm{d}v) < \infty$. Soit également $\lambda > 0$.

On considère une mesure de Poisson d'intensité $\lambda \mathrm{d}x \otimes \mu(\mathrm{d}v)$ sur $\mathbb{R} \times \mathbb{R}$, que l'on écrit $M = \sum_{i \in I} \delta_{(X^{(i)}, V^{(i)})}$, et on l'interprète ainsi : chaque atome correspond à une voiture, de position $X^{(i)}$ et de vitesse $V^{(i)}$ à l'instant $t = 0$. Ainsi, à l'instant $t = 0$, les voitures sont réparties comme les instants de saut d'un processus de Poisson standard d'intensité λ, et leurs vitesses sont i.i.d. de loi μ. Les vitesses négatives sont autorisées, correspondant à un sens de circulation double.

On suppose les vitesses des voitures constantes au cours du temps, et on note $X_t^{(i)} = X^{(i)} + tV^{(i)}$ la position au temps $t \in \mathbb{R}$ de la voiture étiquetée $i \in I$.

(1) Montrer que pour tout $t \in \mathbb{R}$, la mesure M_t a même loi que M, où

$$M_t = \sum_{i \in I} \delta_{(X_t^{(i)}, V^{(i)})}.$$

(2) Soit $T^{(i)}$ le temps auquel la voiture i traverse l'origine : $X_{T^{(i)}}^{(i)} = 0$. Montrer que

$$\sum_{i \in I} \delta_{(T^{(i)}, V^{(i)})}$$

est une mesure de Poisson d'intensité $v_m \lambda \mathrm{d}x \otimes \widehat{\mu}(\mathrm{d}v)$, où $\widehat{\mu}(\mathrm{d}v) = |v|\mu(\mathrm{d}v)/v_m$.

(3) Pour $d > 0$ et $t > 0$ fixés, on considère l'événement A suivant

$$A_{d,t} = \Big\{ \forall\, s \in [0,t], M_s([0,d] \times \mathbb{R}) = 0 \Big\}.$$

Il s'agit donc de l'événement selon lequel le segment de route $[0,d]$ est vide pendant tout l'intervalle de temps $[0,t]$.

(i) Montrer que $A_{d,t} = A^0 \cap A^- \cap A^+$, où

$$\begin{aligned} A^0 &= \Big\{ M\big(\{(x,v) : x \in [0,d]\}\big) = 0 \Big\}, \\ A^- &= \Big\{ M\big(\{(x,v) : x < 0, x + tv \geq 0\}\big) = 0 \Big\}, \\ A^+ &= \Big\{ M\big(\{(x,v) : x > d, x + tv \leq d\}\big) = 0 \Big\}. \end{aligned}$$

(ii) Calculer $P(A_{d,t})$, et déterminer la loi de $C_d = \inf\{t \geq 0 : M_t([0,d] \times \mathbb{R}) \neq 0\}$.

Exercice 10 — Isolation d'un atome en 0
On note $\mathcal{M}$ l'espace des mesures finies sur $[-1,1]$, muni de la topologie étroite.

On considère deux mesures de Poisson indépendantes sur $[-1,1]$, d'intensité la mesure de Lebesgue sur $[-1,1]$, notées M, M'. Pour $n \geq 1$, on note M'_n la restriction de M' à $[-1/n, 1/n]$, et M_n la restriction de M à $[-1,1] \setminus [-1/n, 1/n]$.

1. Justifier que M a même loi que $M_n + M'_n$ pour tout $n \geq 1$.

2. Montrer que lorsque $n \to \infty$, presque sûrement, la mesure M_n converge vers M dans $\mathcal{M}$.

3. Montrer que

$$P\left(M'_n\left(\left[-\frac{1}{n}, \frac{1}{n}\right]\right) > 1 \,\middle|\, M'_n\left(\left[-\frac{1}{n}, \frac{1}{n}\right]\right) \geq 1 \right) \underset{n\to\infty}{\longrightarrow} 0$$

et que sachant $\{M'_n([-1/n, 1/n]) \geq 1\}$, la mesure M'_n est avec probabilité tendant vers 1 une mesure ponctuelle ayant un unique atome en un point de $[-1/n, 1/n]$. Déduire que, sous le même conditionnement, M'_n converge en probabilité vers δ_0 dans $\mathcal{M}$ lorsque $n \to \infty$ (on admettra qu'on peut munir $\mathcal{M}$ d'une distance d telle que $d(\delta_0, \delta_x) \leq |x|$ pour tout $x \in \mathbb{R}$).

4. En déduire que la mesure M, considérée sous $P(\cdot \mid M([-1/n, 1/n]) \geq 1)$, converge en loi lorsque $n \to \infty$ vers $M + \delta_0$, pour la topologie de $\mathcal{M}$.

L'exercice qui vient est une sorte de justification générale du résultat de l'exercice précédent. On rappelle d'abord une version fonctionnelle du théorème de la classe monotone.

Théorème 3.1 *Soit X un ensemble, et $\mathcal{A}$ un ensemble de fonctions $X \to \mathbb{R}$ bornées, stable par multiplication.*

Soit également $\mathcal{H}$ un espace vectoriel de fonctions $X \to \mathbb{R}$ bornées, contenant $\mathcal{A}$ et les fonctions constantes, et stable par convergence monotone bornée : si f est bornée et si $(f_n, n \geq 1)$ est une suite croissante de fonctions de $\mathcal{H}$ convergeant simplement vers f, alors $f \in \mathcal{H}$.

Alors, $\mathcal{H}$ contient toutes les fonctions $\sigma(\mathcal{A})$-mesurables bornées.

Exercice 11 — Formule de Palm
(i) On considère l'algèbre $\mathcal{A}$ engendrée par les fonctions de la forme $m \in E^* \mapsto e^{-m(f)}$, avec f une fonction mesurable positive bornée sur E. À l'aide du théorème de Stone-Weierstrass, montrer que pour tout $A \in \mathcal{E}$ et pour toute fonction continue ayant une limite à l'infini $F : \mathbb{R}_+ \to \mathbb{R}_+$, la fonction $m \mapsto F(m(A))$ est limite uniforme d'une suite de fonctions de $\mathcal{A}$. En déduire que π^A est $\sigma(\mathcal{A})$-mesurable, puis que $\mathcal{E}^* = \sigma(\mathcal{A})$.

(ii) Soit M une mesure de Poisson d'intensité μ, σ-finie, sur un espace mesurable E polonais. On considère une fonction mesurable $F : E^* \times E \to \mathbb{R}_+$, où E^* est l'ensemble des mesures σ-finies sur E. Montrer que

$$\mathbb{E}\left[\int_E M(\mathrm{d}x) F(M, x)\right] = \int_E \mu(\mathrm{d}x) \mathbb{E}[F(M + \delta_x, x)] \,.$$

Pour cela, on considérera le cas particulier où $F(M,x) = f(x)\exp(-M(g))$ avec $f,g : E \to \mathbb{R}$ mesurables positives, et on appliquera la formule de la fonctionnelle de Laplace à la fonction $x \mapsto af(x)+g(x)$, pour $a \geq 0$. On conclura à l'aide de la question précédente et du théorème de la classe monotone.

La formule de Palm s'interprète (intuitivement) ainsi : conditionnellement au fait que M a un atome au point x, ce qui advient avec l'intensité $\mu(\mathrm{d}x)$, la mesure M a même loi que $M+\delta_x$, c'est-à-dire la mesure de Poisson à laquelle on a rajouté un atome en x. En considérant le cas où M est une mesure de Poisson d'intensité la mesure de Lebesgue sur $\mathbb{R}$, on pourra réinterpréter le « paradoxe des autobus » à la lumière de cette constatation.

Exercice 12 — Caractérisation de Rényi des mesures de Poisson
On considère une mesure ponctuelle (somme de mesures de Dirac) aléatoire M sur $\mathbb{R}_+$, presque-sûrement finie sur les compacts, et telle que presque sûrement,

$$\text{pour tout } t \in \mathbb{R}_+\,, \qquad M(\{t\}) \in \{0,1\}\,.$$

Enfin, on suppose que $P(M(A)=0) = \exp(-\lambda(A))$ pour tout borélien A de $\mathbb{R}_+$, où λ est la mesure de Lebesgue sur $\mathbb{R}_+$.

1. Soient $I_1, I_2, \ldots, I_m$ des intervalles deux-à-deux disjoints. Montrer que les événements $\{M(I_j)=0, 1 \leq j \leq m\}$ sont indépendants.

2. Pour tout $n,k \geq 0$ on pose $D_{n,k} = [k2^{-n}, (k+1)2^{-n}[$. Soit J un sous-intervalle borné de $\mathbb{R}_+$, d'intérieur non-vide. Pour $n \geq 0$ on note $K_n = \mathrm{Card}\,\{k \geq 0 : D_{n,k} \subseteq J\}$, et

$$M_n(J) = \sum_{k \geq 0 : D_{n,k} \subseteq J} \mathbb{1}_{\{M(D_{n,k}) \neq 0\}}\,.$$

Déterminer la loi de $M_n(J)$ en termes de n et K_n.

3. Montrer que l'on a $M_n(J) \nearrow M(J)$ presque-sûrement lorsque $n \to \infty$ (on considérera les atomes de M dans J et on vérifiera dans un premier temps qu'ils sont tous distincts de $\inf J$ et $\sup J$), et en déduire la loi de $M(J)$.

4. Soient $J_1, \ldots, J_k$ des sous-intervalles bornés deux-à-deux disjoints de $\mathbb{R}_+$. Montrer que les variables aléatoires $M(J_1), \ldots, M(J_k)$ sont indépendantes.

5. On note $N_t = M([0,t]), t \geq 0$ le processus de comptage associé à M. Montrer que N est un processus de Poisson d'intensité 1. En déduire que M est une mesure de Poisson d'intensité $\lambda(\mathrm{d}x)$.

4 Lois indéfiniment divisibles et processus de Lévy

L'objectif de ce dernier chapitre est de donner une caractérisation des processus de Lévy, que nous avons présentés à la définition 3.3.

Si nous voulons décrire la loi d'un processus de Lévy, les premières choses à déterminer sont les marginales de dimension finie. Or si l'on s'intéresse seulement à la loi de X_1, on remarque que pour tout $n \geq 1$,

$$X_1 = \sum_{i=1}^{n} (X_{i/n} - X_{(i-1)/n}) .$$

Par définition d'un processus de Lévy, les éléments de cette somme sont indépendants, et de même loi que $X_{1/n}$. Ainsi, pour tout n donné, X_1 peut s'écrire comme somme de n variables aléatoires indépendantes de même loi. On dit que la variable aléatoire est *indéfiniment divisible*. Les lois indéfiniment divisibles ont été introduites par Lévy dans les années 30. Le but des paragraphes 4.1 et 4.2 est de donner une caractérisation de toutes ces lois à travers une formule analytique explicite pour leur fonction caractéristique, dite *formule de Lévy-Khintchine*.

Une fois que nous aurons déterminé toutes les lois indéfiniment divisibles, nous verrons que pour toute loi indéfiniment divisible μ, il existe une unique loi sur les processus càdlàg (continus à droite, admettant des limites à gauche partout), telle que si $(X_t, t \geq 0)$ suit cette loi, alors X_1 a pour loi μ. Nous terminerons par une construction explicite d'un tel processus, dite construction de Lévy-Itô.

4.1 Lois indéfiniment divisibles : définitions et premières propriétés

Dans tout ce chapitre, on fixe $d \geq 1$.

Définition 4.1 *Une loi de probabilités μ sur $\mathbb{R}^d$ est dite indéfiniment divisible (ID en abrégé) si pour tout $n \geq 1$ il existe une probabilité μ_n telle que $\mu = \mu_n^{*n}$.*

De façon équivalente, une loi μ est ID si une variable aléatoire X de loi μ est égale en loi à une somme de variables i.i.d. $X_{n,1} + X_{n,2} + \ldots + X_{n,n}$ pour tout n. Si l'on note ϕ_X (ou plus simplement ϕ lorsqu'il n'y a pas d'ambiguité sur la variable X) la fonction caractéristique de X

$$\phi_X(u) = \mathbb{E}[\exp(\mathrm{i}u \cdot X)] , \qquad u \in \mathbb{R}^d,$$

ceci revient à dire que pour tout n, il existe une fonction caractéristique ϕ_n qui est une racine n-ième de ϕ, soit $\phi_n^n = \phi$.

Remarque. Dans cette dernière définition, ce n'est pas l'existence d'une racine n-ième qui est problématique, mais bien le fait que cette racine soit une fonction caractéristique.

Exemples. Sont indéfiniment divisibles :
- les masses de Dirac $\delta_a = \delta_{a/n}^{*n}$,
- les lois normales $\mathcal{N}(m, \sigma^2) = \mathcal{N}(m/n, \sigma^2/n)^{*n}$,

• les lois de Poisson $\mathcal{P}(\lambda) = \mathcal{P}(\lambda/n)^{*n}$,
• les lois Gamma(a, θ)=Gamma$(a/n, \theta)^{*n}$, où Gamma(a, θ) est la loi sur $\mathbb{R}_+$ de densité $(\theta^a \Gamma(a))^{-1} x^{a-1} e^{-\theta x} \mathbb{1}_{\{x \geq 0\}} \mathrm{d}x$, $a > 0, \theta > 0$,
• les lois géométriques : voyez-vous pourquoi ?

En revanche, les lois de Bernoulli non constantes, les lois binômiales, les lois uniformes sur un intervalle de longueur finie ne sont pas indéfiniment divisibles. On peut le voir pour les lois de Bernoulli pour des raisons simples de support, plus généralement :

Lemme 4.1 *Une loi ID à support borné est nécessairement une masse de Dirac.*

Preuve. Si X est indéfiniment divisible et vérifie $|X| \leq M$ p.s. ($|\cdot|$ étant une norme quelconque sur $\mathbb{R}^d$), alors en écrivant $X = X_{n,1} + \ldots + X_{n,n}$, avec les $X_{n,i}$ i.i.d., nécessairement on a $|X_{n,i}| \leq M/n$ p.s. En effet si ce n'est pas le cas, le support de la loi de $X_{n,1}$ contient un x tel que $|x| > M/n$, mais alors avec une probabilité strictement positive, $X_{n,i} \in V(x)$ pour tout i, où $V(x)$ est voisinage ouvert de x suffisamment petit pour que $|x_1 + \ldots + x_n| > M$ pour tout $x_1, \ldots, x_n \in V(x)$, ce qui est absurde.

On en déduit que $\mathrm{Var}\,(X_{n,1}) \leq M^2/n^2$, et que $\mathrm{Var}\,(X) \leq M^2/n$ pour tout n, et donc $\mathrm{Var}\,(X) = 0$ et X est constante. □

4.1.1 Premières propriétés des fonctions caractéristiques, exposant caractéristique

L'outil crucial qui va nous permettre de classer les lois ID est leur fonction caractéristique, ou plus précisément leur exposant caractéristique. Pour définir ce dernier, nous commençons par un lemme important.

Lemme 4.2 *La fonction caractéristique d'une loi ID ne s'annule pas.*

Preuve. On commence par se ramener au cas où la f.c. de X ID est réelle, par un argument de symétrisation. Si X est ID, $-X$ aussi et $X - X'$ aussi où X' est une copie de X indépendante. De plus, la f.c. de cette dernière est $\phi_{X-X'} = \phi_X \phi_{-X} = \phi_X \overline{\phi}_X = |\phi|^2$. Comme $X - X'$ est ID, on peut trouver une f.c. ϕ_n telle que $\phi_n^n = |\phi|^2$, et donc nécessairement $\phi_n = |\phi|^{2/n}$ puisque $\phi_n(0) = 1$ et par continuité de ϕ_n. Par conséquent, $\phi_n(u)$ converge vers $\mathbb{1}_{\{\phi(u) \neq 0\}}$ pour tout $u \in \mathbb{R}^d$. Mais ϕ est non nulle sur un voisinage de l'origine, puisqu'une fonction caractéristique est continue, et égale à 1 en 0. Donc ϕ_n converge ponctuellement sur $\mathbb{R}^d$ vers une fonction égale à 1 sur un voisinage de 0. Donc par le théorème de Lévy (théorème 1.2), la fonction limite est une fonction caractéristique, et en particulier, elle est continue. Mais comme elle est à valeurs dans $\{0, 1\}$ (c'est une indicatrice !), on en conclut qu'elle vaut 1 partout. Donc $\phi(u) \neq 0$ pour tout $u \in \mathbb{R}^d$. □

Comme corollaire du dernier lemme, on déduit que la fonction ϕ est une fonction continue à valeurs dans $\mathbb{C}^*$, et cette dernière se relève de façon unique en une fonction continue $\psi : \mathbb{R}^d \to \mathbb{R}$ nulle en 0, telle que $\phi(u) = \exp(\mathrm{i}\psi(u))$ pour tout u. On appellera par la suite ψ l'*exposant caractéristique* de la loi considérée, voir l'exercice 3 du chapitre 1.

Remarque. En particulier, on obtient que si ϕ_n est une fonction caractéristique telle que $\phi_n^n = \phi$, c'est-à-dire que ϕ_n est la fonction caractéristique d'une loi μ_n telle qu'apparaissant dans la définition 4.1, alors ϕ_n elle-même ne s'annule pas, et donc s'écrit de façon unique $\phi_n = \exp(\psi_n)$ avec ψ_n continue nulle en 0. Mais alors $\exp(n\psi_n) = \exp(\psi)$, d'où l'on déduit par unicité de ψ continue nulle en 0 que $\psi_n = \psi/n$. Nous avons donc obtenu que la factorisation d'une loi ID en sa partie n-ième est unique.

4.1.2 Lois ID et lois de Poisson composées

Rappelons maintenant de la section 3.4.2 que la loi de Poisson composée d'intensité ν, mesure positive finie sur $\mathbb{R}^d$, est la loi de la somme $Y_1 + \ldots + Y_N$, où les Y_i sont tous de loi $\nu/\nu(\mathbb{R}^d)$, indépendants entre eux et de N, cette dernière suivant la loi de Poisson de paramètre $\nu(\mathbb{R}^d)$.

Lemme 4.3 *Toute loi de Poisson composée est ID.*

En effet on a par la formule de la fonction caractéristique que $\mathrm{PC}(\nu) = \mathrm{PC}(\nu/n)^{*n}$. Nous avons introduit en section 3.4.2 les processus de Poisson composés d'intensité ν, et nous avons vu qu'ils sont des processus de Lévy dont la loi marginale à l'instant t est $\mathrm{PC}(t\nu)$.

Nous en venons maintenant à un lemme clef de l'étude des lois ID :

Lemme 4.4 (i) *Toute limite étroite de lois ID est ID.*
(ii) *Toute loi ID est limite étroite de lois de Poisson composées.*

Preuve. Nous prouvons (i). Si X_n est ID pour tout n, on peut écrire, pour tout n, p, $X_n = X_{p,1}^{(n)} + \ldots + X_{p,p}^{(n)}$ avec les $X_{p,i}^{(n)}$ i.i.d., de loi disons $\mu_{n,p}$. Mais alors, les mesures $\mu_{n,p}, n \geq 1$ forment une famille tendue de v.a. En effet, on a en notant x_1 la première composante de x,

$$\mu_{n,p}(x_1 > A)^p = P((X_{p,i}^{(n)})_1 > A, 1 \leq i \leq p) \leq P((X_n)_1 > pA),$$

et en raisonnant de même avec les autres coordonnées et leurs opposés, on déduit du fait que les lois de X_n sont tendues que $(\mu_{n,p}, n \geq 1)$ l'est (on aurait également pu utiliser (1)). Ainsi, quitte à extraire, on peut supposer que $X_{p,i}^{(n)}, 1 \leq i \leq p$ convergent en loi vers p variables i.i.d. $X_{p,1}, \ldots, X_{p,p}$. Nécessairement, la loi de leur somme est celle de X.

Pour prouver (ii), on écrit si $\phi_n = \exp(\psi/n)$ est la fonction caractéristique de la fraction n-ième d'une loi ID, $(1 - (1 - \phi_n))^n = \phi$. À u fixé, pour tout n assez grand, on peut appliquer la fonction Log, détermination principale du logarithme, pour obtenir $\exp(\mathrm{Log}(1 - (1 - \phi_n(u)))) = \hat{\mu}_n(u)$, et donc $\exp(n\mathrm{Log}(1 - (1 - \phi_n(u)))) = \phi$. Comme ϕ_n converge ponctuellement vers 1 et que $\mathrm{Log}(1 - z) = -\sum_{k \geq 1} z^k/k$ pour z près de 0, on déduit que $-n(1 - \phi_n) \to \psi$, ou encore $\exp(-n(1 - \phi_n)) \to \phi$. On reconnaît à gauche la fonction caractéristique d'une loi de Poisson composée, de mesure $\nu = n\mu_n$ où μ_n a pour fonction caractéristique ϕ_n. D'où le résultat. $\square$

4.2 La formule de Lévy-Khintchine

La formule de Lévy-Khintchine permet de classifier les exposants caractéristiques ψ des lois ID comme famille dépendant de trois paramètres, comme suit.

On se donne $a \in \mathbb{R}^d$, $\Sigma \in \mathfrak{M}_d(\mathbb{R})$ une matrice symétrique positive, et Π une mesure positive σ-finie sur $\mathbb{R}^d$ ne chargeant pas 0 et vérifiant

$$\int_{\mathbb{R}^d} \Pi(\mathrm{d}x)(|x|^2 \wedge 1) < \infty.$$

Cette dernière condition montre que Π est finie si on la restreint au complémentaire d'un voisinage V de 0, car alors $\mathbb{1}_{V^c}(x) \leq C|x|^2 \wedge 1$, où $C = \sup_{x \in V^c} |x|^{-2}$. Ainsi, Π ne peut éventuellement charger d'une masse infinie que des voisinages de 0.

Soit h une *fonction de troncation*, égale à l'identité sur un voisinage de 0, bornée sur $\mathbb{R}^d$ et nulle sur $\{x : |x| > 1\}$. Par exemple, on choisira souvent $h(x) = x\mathbb{1}_{\{|x| \leq 1\}}$. Une telle fonction étant fixée, on définit pour $u \in \mathbb{R}^d$,

$$\psi_{a,\Sigma,\Pi}(u) = \mathrm{i}u \cdot a - \frac{1}{2}\Sigma u \cdot u + \int_{\mathbb{R}^d} \Pi(\mathrm{d}x)\left(e^{\mathrm{i}u \cdot x} - 1 - \mathrm{i}u \cdot h(x)\right). \qquad (11)$$

Lemme 4.5 *Cette fonction est bien définie sur $\mathbb{R}^d$, elle est continue, nulle en 0, et c'est l'exposant caractéristique d'une loi ID.*

Preuve. Le fait que $\psi_{a,\Sigma,\Pi}$ soit bien définie, nulle en 0 et continue est aisé à partir du fait que, si u évolue dans un compact K,

$$|e^{\mathrm{i}u \cdot x} - 1 - \mathrm{i}u \cdot h(x)| \leq C_K(|x|^2 \wedge 1),$$

pour un certain C, la fonction considérée étant continue en x, u et étant bornée et $O(|u \cdot x|^2)$ au voisinage de 0. On conclut par les théorèmes de Lebesgue.

Par ailleurs, comme $\mathrm{i}u{\cdot}a$ et $-\Sigma u{\cdot}u/2$ sont respectivement les exposants caractéristiques des lois δ_a et $\mathcal{N}(0, \Sigma)$, il suffit de montrer que le troisième terme de (11) est un exposant caractéristique. Or ce dernier peut se réécrire, par convergence dominée,

$$\lim_{n\to\infty} \left(-\int_{|x|\geq 1/n} \Pi(\mathrm{d}x)(1 - e^{\mathrm{i}u\cdot x}) - \int_{|x|\geq 1/n} h(x)\Pi(\mathrm{d}x) \right),$$

les quantités considérées étant bien définies puisque Π restreinte à $|x| \geq 1/n$ (appelons-la Π_n) est de masse finie. On reconnaît alors la somme des exposants caractéristiques d'une loi de Poisson composée d'intensité Π_n, et d'une masse de Dirac en $-\int_{|x|\geq 1/n} h(x)\Pi(\mathrm{d}x)$.

On en conclut que $\exp(\psi_{a,\Sigma,\Pi})$ est une limite ponctuelle continue de fonction caractéristique de lois ID, et c'est donc une fonction caractéristique de loi ID par le lemme 4.4. $\square$

Le théorème qui suit est la réciproque du lemme précédent.

Théorème 4.1 (Formule de Lévy-Khintchine) *On fixe la fonction de troncation h. Pour toute loi ID μ, il existe un unique triplet (a, Σ, Π) vérifiant les hypothèses ci-dessus, tel que la fonction caractéristique de μ est $\exp(\psi_{a,\Sigma,\nu})$.*

Remarque. Comme cas particuliers, on obtient bien que les lois des constantes ($\Sigma = 0, \Pi = 0$), les lois normales ($\Pi = 0$), les lois de Poisson composées ($\Sigma = 0$, Π finie et $b - \int_{\mathbb{R}^d} \Pi(\mathrm{d}x)h(x) = 0$) sont ID.

Remarque. Notons que l'unicité est garantie seulement une fois le choix de h effectué. Cependant, si l'on choisit une autre fonction h', l'exposant $\psi'_{a,\Sigma,\Pi}$ associé vérifie facilement

$$(\psi_{a,\Sigma,\Pi} - \psi'_{a,\Sigma,\Pi})(u) = \mathrm{i}u \cdot \int_{\mathbb{R}^d} \Pi(\mathrm{d}x) \left(h(x) - h'(x)\right),$$

cette dernière quantité étant bien définie puisque $h - h'$ est bornée et nulle dans un voisinage de 0. Par conséquent, un changement de choix pour h revient seulement à un changement du paramètre b dans la formule de Lévy-Khintchine, sans affecter Σ et Π.

Nous commençons par montrer l'unicité dans la formule de Lévy-Khintchine.

Lemme 4.6 *La fonction $\psi_{a,\Sigma,\Pi}$ détermine a, Σ, Π de façon unique.*

Preuve. Si ν est suffisamment régulière pour que l'on puisse dériver deux fois sous le signe $\int$ (ce qui équivaut à la sommabilité de $|x|^2$ pour Π), le résultat est assez aisé. Cependant, ceci est faux dans de nombreux cas. L'idée est alors d'effectuer une autre opération que la dérivation sur la fonction.

Pour $v \in \mathbb{R}^d$, nous calculons

$$\begin{aligned}
\frac{1}{2}\int_{-1}^{1} \psi_{a,\Sigma,\Pi}(u + tv)\mathrm{d}t &= \mathrm{i}u \cdot a - \frac{1}{2}\Sigma u \cdot u - \frac{1}{2}\int_{-1}^{1} \mathrm{d}t\, t^2 \Sigma v \cdot v \\
&\quad + \int_{\mathbb{R}^d} \Pi(\mathrm{d}x) \left(e^{\mathrm{i}u\cdot x} \frac{1}{2}\int_{-1}^{1} e^{\mathrm{i}tv\cdot x}\mathrm{d}t - 1 - \mathrm{i}u \cdot h(x) \right) \\
&= \psi_{a,\Sigma,\Pi}(u) + \frac{1}{3}\Sigma v \cdot v - \int_{\mathbb{R}^d} \Pi(\mathrm{d}x) \left(1 - \frac{\sin(v \cdot x)}{v \cdot x}\right) e^{\mathrm{i}u\cdot x}.
\end{aligned}$$

On en déduit que la fonction $\psi_{a,\Sigma,\Pi}(u) - 2^{-1}\int_{-1}^{1} \psi_{a,\Sigma,\Pi}(u + tv)\mathrm{d}t$ est la transformée de Fourier $\widehat{H}_v$ de la mesure

$$H_v(\mathrm{d}x) = \frac{1}{3}\Sigma v \cdot v\, \delta_0(\mathrm{d}x) + \left(1 - \frac{\sin(v \cdot x)}{v \cdot x}\right) \Pi(\mathrm{d}x). \tag{12}$$

On peut donc retrouver ${}^t v \Sigma v$ comme triple de la masse en 0 de cette mesure (le terme de droite attribuant toujours une masse nulle à $\{0\}$), et ces quantités déterminent Σ lorsque v varie dans $\mathbb{R}^d$. Par ailleurs, la fonction $1 - \sin(v \cdot x)/v \cdot x$ s'annule exactement sur l'hyperplan $\{x : v \cdot x = 0\}$, et donc la connaissance de H_v pour v variant dans $\mathbb{R}^d$ permet d'obtenir Π sur le complémentaire de l'intersection de ces hyperplans, qui est $\{0\}$. Comme Π ne charge pas $\{0\}$, ceci détermine Π.

Enfin, on obtient aisément b par soustraction. $\square$

Remarque. Le fait de supposer que Π ne charge pas $\{0\}$ ne sert qu'ici, pour démontrer l'unicité dans la formule de Lévy-Khintchine. Sinon, il est immédiat que changer Π en

$\Pi + c\delta_0$ ne change pas la fonction $\psi_{a,\Sigma,\Pi}$. Il est aussi à remarquer que lorsque $d = 1$, certains ouvrages (par exemple Feller [5]) utilisent une convention différente et écrivent la formule de Lévy-Khintchine sous la forme compacte

$$\psi_{0,\Sigma,F}(u) = \int_{\mathbb{R}} \frac{(e^{\mathrm{i}ux} - 1 - \mathrm{i}u \sin x)}{x^2} F(\mathrm{d}x),$$

où F est une mesure finie est finie. Dans ce cas, la masse de F en 0 sort de l'intégrale sous forme $-qu^2/2$ et donne la partie "normale" de la formule ci-dessus. La mesure $F(\mathrm{d}x)$ correspond à $x^2\Pi(\mathrm{d}x)$ pour nous. Notons aussi que $h = \sin$ ne se confond pas avec l'identité près de 0, mais l'important est que la différence soit un $o(|x|^2)$.

Notons

$$\widetilde{\Sigma} = \Sigma + \int_{\mathbb{R}^d} \Pi(\mathrm{d}x) h(x) h(x)^*.$$

Ici, $h(x)^*$ désigne le vecteur transposé de $h(x)$, et pour tout x, $h(x)h(x)^*$ est donc une matrice de rang 1 positive symétrique. On remarque que la fonction $x \mapsto h(x)h(x)^*$ à valeurs matrices est bien Π-intégrable car son terme générique $x_i x_j \leq x_i^2 + x_j^2 \leq \|x\|_2$ pour x proche de 0, et cette fonction est bornée puisque h l'est.

Proposition 4.1 *On suppose la fonction h continue.*

Soit $(a_n, \Sigma_n, \Pi_n)_{n\geq 1}$ une suite de triplets satisfaisant les hypothèses ci-dessus. On suppose qu'il existe (a, Σ, Π) un tel autre triplet satisfaisant, quand $n \to \infty$

$$\begin{cases} a_n \to a \\ \widetilde{\Sigma}_n \to \widetilde{\Sigma} \\ \Pi_n(f) \to \Pi(f), \end{cases} \tag{13}$$

ce pour toute fonction f continue bornée qui est un $o(|x|^2)$ quand $x \to 0$. Alors on a convergence étroite de la loi ID de triplet caractéristique (a_n, Σ_n, Π_n), vers la loi ID de triplet caractéristique (a, Σ, Π).

Réciproquement, si X_n de loi ID associée à (a_n, Σ_n, Π_n) converge en loi lorsque $n \to \infty$, alors il existe (a, Σ, Π) tels que les trois conditions (13) sont vérifiées.

Preuve. Le sens direct est le plus facile. En effet, en supposant (13), on peut réécrire l'exposant caractéristique de X_n

$$\psi_{a_n,\Sigma_n,\Pi_n}(u) = \mathrm{i}u \cdot a_n - \frac{1}{2}\widetilde{\Sigma}_n u \cdot u + \int_{\mathbb{R}^d} \Pi_n(\mathrm{d}x) \left(e^{\mathrm{i}u\cdot x} - 1 - \mathrm{i}u \cdot h(x) + \frac{1}{2}(u \cdot h(x))^2\right).$$

À u fixé, la fonction dans l'intégrale, que nous appelons $f_u(x)$, est un $o(|x|^2)$ quand $x \to 0$, est bornée et continue, et en particulier elle est bien Π_n-intégrable. Par (13), on obtient donc que $\psi_{a_n,\Sigma_n,\Pi_n}$ converge ponctuellement lorsque $n \to \infty$ vers

$$\mathrm{i}u \cdot a - \frac{1}{2}\widetilde{\Sigma} u \cdot u + \Pi(f_u),$$

ce que l'on peut réécrire $\psi_{a,\Sigma,\Pi}$, et c'est ce qu'il fallait démontrer.

Réciproquement, on suppose donc que les fonction caractéristique $\phi_n = \exp(\psi_{a_n,\Sigma_n,\Pi_n})$ convergent ponctuellement vers une limite ϕ, et on veut montrer que $\phi = \exp(\psi_{a,\Sigma,\Pi})$ pour un (a, Σ, Π) vérifiant (13). Déjà, on sait que la limite est ID, et donc son exposant caractéristique ψ existe et est la limite simple de $\psi_n = \psi_{a_n,\Sigma_n,\Pi_n}$ sur $\mathbb{R}^d$, et même uniforme sur les compacts.

On réintroduit alors la quantité

$$\widehat{H}_v^n(u) = \psi_n(u) - \int_{-1}^{1} \psi_n(u + tv)\mathrm{d}t,$$

et on voit (par convergence uniforme) que le membre de droite converge vers

$$\psi(u) - \int_{-1}^{1} \psi(u + tv)\mathrm{d}t.$$

On en conclut que $\widehat{H}_v^n$ converge ponctuellement vers une limite $\widehat{H}_v$, continue en 0. Or $\widehat{H}_v^n$ est, d'après la preuve du Lemme 4.6, la transformée de Fourier d'une mesure finie H_v^n. On en conclut que cette dernière converge étroitement vers une mesure finie H_v, c'est-à-dire que $H_v^n(1)$ converge vers $H_v(1)$ et que, si $H_v(1) \neq 0$, la probabilité $H_v^n/H_v^n(1)$ converge étroitement vers $H_v/H_v(1)$. Soit alors $e_i, 1 \leq i \leq d$ la base canonique de $\mathbb{R}^d$, $H^n = \sum_{i=1}^d H_{e_i}^n$, $H = \sum_{i=1}^d H_{e_i}$ et

$$r(x) = \sum_{i=1}^{d} \left(1 - \frac{\sin x_i}{x_i}\right),$$

où $x = (x_1, \ldots, x_d)$, de sorte que r ne s'annule qu'en 0.

Définissons alors

$$\Pi(\mathrm{d}x) = \mathbb{1}_{\{x \neq 0\}} \frac{H(\mathrm{d}x)}{r(x)}.$$

Il s'agit d'une mesure σ-finie, et on vérifie que

$$\int_{\mathbb{R}^d} (|x|^2 \wedge 1)\Pi(\mathrm{d}x) = \int_{\mathbb{R}^d \setminus \{0\}} \frac{|x|^2 \wedge 1}{r(x)} H(\mathrm{d}x) < \infty,$$

car H est une mesure finie, $r(x) = \Theta(|x|^2)$ quand $x \to 0$, et $r(x) \to 1$ quand $|x| \to \infty$. De plus, si f est une fonction continue, bornée qui est un $o(|x|^2)$ en 0, on a

$$\Pi_n(f) = H^n\left(\frac{f}{r}\right) \underset{n\to\infty}{\to} H\left(\frac{f}{r}\right) = \Pi(f)$$

où la première égalité est une conséquence de (12), et la convergence provient de la convergence étroite de H^n vers H, car f/r est une fonction continue. Notons qu'il est également crucial que cette fonction soit nulle en 0, c'est ce qui fait que l'on a bien $H^n(f/r) = H^n(\mathbb{1}_{\{x\neq 0\}} f/r) = \Pi_n(f)$, et idem pour $H(f/r) = \Pi(f)$.

Réécrivons, comme pour le sens direct de la preuve, ψ_n sous forme

$$\psi_n(u) = \mathrm{i}u \cdot a_n - \frac{1}{2}\widetilde{\Sigma}_n u \cdot u + \Pi_n(f_u),$$

où $f_u(x) = e^{iu\cdot x} - 1 - iu\cdot x + (u\cdot h(x))^2/2$ est une fonction continue bornée, $o(|x|^2)$ en 0. Par ce qui précède, $\Pi_n(f_u)$ converge pour tout u vers $\Pi(f_u)$. Or par hypothèse, $\psi_n(u)$ converge également pour tout u, ce qui fait que $iu\cdot a_n - \widetilde{\Sigma}_n u\cdot u/2$ converge pour tout u. En tant que polynôme en u de degré au plus, 2, ceci implique que a_n et $\widetilde{\Sigma}_n$ convergent respectivement vers un $a\in\mathbb{R}^d$ et vers un $\widetilde{\Sigma}\in\mathfrak{M}_d(\mathbb{R})$ symétrique positive. La preuve sera complète si l'on montre que $\Sigma := \widetilde{\Sigma} - \int_{\mathbb{R}^d}\Pi(dx)h(x)h(x)^*$ est positive.

Or, pour tout u, on constate que

$$\liminf_{n\to\infty}\int_{\mathbb{R}^d}\Pi_n(dx)(u\cdot h(x))^2 \geq \int_{\mathbb{R}^d}\Pi(dx)(u\cdot h(x))^2,$$

ce qui implique que

$$\Sigma u\cdot u \geq \limsup_{n\to\infty}\Sigma_n u\cdot u \geq 0. \tag{14}$$

Pour montrer l'inégalité annoncée, pour chaque $k\geq 1$ soit f_k une fonction positive continue bornée par 1, égale à 1 sur $\{|x|\geq 2/k\}$ et à 0 sur $\{|x|\leq 1/k\}$. Si g est une fonction positive continue bornée, on a

$$\liminf_{n\to\infty}\Pi_n(g) \geq \lim_{n\to\infty}\Pi_n(g\,f_k) = \Pi(g\,f_k) \underset{k\to\infty}{\uparrow} \Pi(g),$$

où l'égalité provient du fait que $g\,f_k$ est une fonction continue bornée nulle au voisinage de 0. On applique ceci à $g(x) = (u\cdot h(x))^2$. Ceci termine la démonstration. □

Nous sommes maintenant en mesure de donner la preuve de la formule de Lévy-Khintchine.

Démonstration du Théorème 4.1. Soit μ une loi ID. Le Lemme 4.4, (ii) nous dit que μ est limite étroite d'une suite de lois de Poisson composées. Comme ces dernières ont des exposants caractéristiques de la forme $\psi_{a,0,\nu}$ avec ν finie, on en déduit par le réciproque de la Proposition 4.1 que l'exposant caractéristique de μ est nécessairement de la forme $\psi_{a,\Sigma,\Pi}$. L'unicité a été vue au Lemme 4.6. □

La formule de Lévy-Khintchine permet donc de reformuler la Proposition 4.1 sous la forme d'un critère de convergence très utile.

Théorème 4.2 *On suppose h continue. Soit μ_n une suite de lois ID, de triplets caractéristiques (a_n, Σ_n, Π_n). Cette suite converge étroitement si et seulement si (13) est vérifiée.*

Remarque. Il est à noter que en passant à la limite au sein des lois ID, “la partie gaussienne augmente toujours”, comme indiqué par (14). Par exemple, une loi gaussienne $\mathcal{N}(0,1)$ est limite de lois de Poisson composées $\mathrm{PC}(n\mu_n)$ avec par exemple μ_n la loi de $X/\sqrt{n}$ où X suit une loi centrée de variance 1 (c'est une Poissonnisation du théorème central limite). En revanche, aucune loi de Poisson composée (en particulier sans partie gaussienne) ne peut s'obtenir comme limite étroite de lois gausiennes, ou même seulement ayant une partie gaussienne qui ne tend pas vers 0.

Remarque. Dans l'énoncé de la proposition 4.1, on peut remplacer la condition sur les fonctions f par la condition plus faible : f continue bornée nulle dans un voisinage de 0. Le lecteur est invité à montrer cette amélioration.

Le cas des lois positives Dans le cas des lois ID portées par $\mathbb{R}_+$, la formule de Lévy-Khintchine se simplifie comme suit. Rappelons que pour une probabilité μ sur $\mathbb{R}_+$ associée à la v.a. X positive, on peut définir la transformée de Laplace

$$\varphi(\lambda) = \mathbb{E}[\exp(-\lambda X)] \in]0, \infty[,$$

et on notera $\psi = -\log \phi$ l'exposant de Laplace.

Théorème 4.3 *Une loi μ sur $\mathbb{R}_+$ est ID si et seulement s'il existe* $\mathrm{d} \geq 0$ *et une mesure σ-finie Π telle que $\int_0^\infty (|x| \wedge 1)\Pi(\mathrm{d}x) < \infty$, telles que*

$$\psi(\lambda) = \mathrm{d}\lambda + \int_0^\infty (1 - e^{-\lambda x})\Pi(\mathrm{d}x).$$

Preuve. Il est tout d'abord clair qu'une loi μ ID portée par $\mathbb{R}_+$ doit avoir $\Sigma = 0$, car une loi normale est portée par $\mathbb{R}$ tout entier et la convolution de deux lois dont l'une a $\mathbb{R}$ pour support a elle-même $\mathbb{R}$ pour support. Ensuite, pour chaque n, écrivons μ comme convolée de δ_a, $\mathrm{PC}(\Pi \mathbb{1}_{\{|x| \geq 1/n\}})$, δ_{-b_n} et la loi ID de triplet $0, 0, \Pi \mathbb{1}_{\{|x|<1/n\}}$, où $b_n = \int_{|x| \geq 1/n} h(x)\Pi(\mathrm{d}x)$. On en déduit que Π ne charge pas $\mathbb{R}_-$ car sinon, en prenant n assez grand on aurait que le support de $\mathrm{PC}(\Pi \mathbb{1}_{\{|x| \geq 1/n\}})$ contiendrait une partie non bornée de $\mathbb{R}_-$, et ce serait donc aussi le cas pour μ. Mais de ce fait, on déduit facilement que $a - b_n \geq 0$ pour tout n. Ainsi b_n converge en croissant vers une quantité finie b telle que $\mathrm{d} = a - b \geq 0$, et ceci dit que h est Π-intégrable, c'est-à-dire que Π intègre $|x|$ au voisinage de 0. On en déduit que

$$\phi(u) = \exp\left(\mathrm{i}\mathrm{d}u + \int_0^\infty (e^{\mathrm{i}ux} - 1)\Pi(\mathrm{d}x)\right),$$

et on déduit le résultat sur l'exposant de Laplace par prolongement analytique, la fonction $z \mapsto \int_0^\infty (e^{zx} - 1)\Pi(\mathrm{d}x)$ étant holomorphe sur $\{\Re(z) < 0\}$. □

4.3 Construction des processus de Lévy à la Lévy-Itô

On a vu dans un exercice précédent qu'à toute loi ID μ il est possible d'associer de façon naturelle un semigroupe $\mu_t, t \geq 0$ et un processus de Lévy associé à ce semigroupe. Le but de ce paragraphe est de montrer qu'il existe une construction explicite généralisant la version "processus" des lois de Poisson composées. Cette construction, dite de Lévy-Itô, jette en retour une lumière sur la formule de Lévy-Khintchine, car elle permet d'interprèter la partie intégrale en termes de processus de Poisson ponctuels.

Nous commençons par des résultats sur les processus de Poisson composés. Soit ν une mesure finie et $(X_t, t \geq 0)$ un processus de Poisson composé càdlàg d'intensité ν, comme défini au paragraphe 4.1. On notera $\Delta X_t = X_t - X_{t-}$ le saut (éventuellement nul) effectué par X à l'instant t, et pour toute fonction f nulle en 0 et $t \geq 0$ on note

$$N(f)_t = \sum_{0 \leq s \leq t} f(\Delta X_s).$$

Notons que cette somme a bien un sens, car X n'a presque-sûrement qu'un nombre fini de sauts de taille non nulle entre les instants 0 et t. De plus, il est aisé de voir que $N(f)$ est à son tour un processus de Lévy, c'est en fait un processus de Poisson composé d'intensité $f \circ \nu$. On note $(\mathcal{F}_t, t \geq 0)$ une filtration dans laquelle X est un processus de Lévy (par exemple la filtration naturelle de X, mais le cas général servira plus bas).

Lemme 4.7 (i) *Si $f \in \mathbb{L}^1(\nu)$, alors $M(f) = (N(f)_t - t\nu(f), t \geq 0)$ est une $\mathcal{F}_t$-martingale.*
(ii) *Si $f \in \mathbb{L}^2(\nu)$, alors $(M(f)_t^2 - t\nu(f^2), t \geq 0)$ est une $\mathcal{F}_t$-martingale.*

Preuve. (i) Si X est un processus de Lévy d'espérance finie pour un (tout) $t > 0$, $(X_t - t\mathbb{E}[X_1], t \geq 0)$ est une martingale (exercice). Il suffit donc de voir que $\mathbb{E}[N(f)_1] = \nu(f)$. Pour ce faire, on utilise le fait que $N(f)_1$ suit la loi $\mathrm{PC}(f\circ\nu)$, et la formule demandée est un calcul immédiat en dérivant la fonction caractéristique $\exp(-\int_{\mathbb{R}^d}(1-e^{\mathrm{i}uf(x)})\nu(\mathrm{d}x))$ en 0.
(ii) En dérivant deux fois la fonction caractéristique ci-dessus (mais en remplaçant ν par $t\nu$), on obtient de façon similaire que $\mathbb{E}[N(f)_t^2] = t\nu(f^2) + t^2\nu(f)^2$. De ce fait, en notant $M(f)_{t+s} = M(f)_{t+s} - M(f)_t + M(f)_t = \widetilde{M}(f)_s + M(f)_t$ avec $\widetilde{M}_s$ indépendant de $\mathcal{F}_t$, de loi celle de $M(f)_s$,

$$\begin{aligned}\mathbb{E}[M_{t+s}(f)^2|\mathcal{F}_t] &= \mathbb{E}[\widetilde{M}(f)_s^2] + M(f)_t^2 + 2M(f)_t\mathbb{E}[\widetilde{M}(f)_s] \\ &= \mathbb{E}[N(f)_s^2] - 2\nu(f)sE[N(f)_s] + s^2\nu(f)^2 + M(f)_t^2 \\ &= s\nu(f^2) + M(f)_t^2,\end{aligned}$$

et on obtient le résultat en soustrayant $(t+s)\nu(f^2)$. □

On se fixe maintenant une loi μ ID de triplet caractéristique (a, Σ, Π) pour le choix de $h(x) = x\mathbb{1}_{\{|x|\leq 1\}}$. Notons, pour $n \geq 0$

$$E_n = \left\{x \in \mathbb{R}^d : \frac{1}{n+1} \leq |x| < \frac{1}{n}\right\}, \qquad \Pi_n(\mathrm{d}x) = \Pi(\mathrm{d}x)\mathbb{1}_{\{x\in E_n\}},$$

avec la convention $1/0 = +\infty$. Notons que Π_n est une mesure finie pour tout $n \geq 0$ par hypothèse. On voit, en écrivant

$$\begin{aligned}\psi_{a,\Sigma,\Pi}(u) &= \mathrm{i}u\cdot a - \frac{1}{2}\Sigma u\cdot u + \int_{\mathbb{R}^d}\Pi_0(\mathrm{d}x)(e^{\mathrm{i}u\cdot x} - 1) \\ &\quad + \sum_{n\geq 1}\left(\int_{\mathbb{R}^d}\Pi_n(\mathrm{d}x)(e^{\mathrm{i}u\cdot x} - 1) - \mathrm{i}u\cdot\int_{\mathbb{R}^d}x\Pi_n(\mathrm{d}x)\right),\end{aligned}$$

qu'une v.a. X de loi μ peut s'interpréter comme somme de v.a. indépendantes, de lois respectivement

- une constante δ_a,
- une loi normale de matrice de covariance Σ,
- une loi de Poisson composée $\mathrm{PC}(\Pi_0)$,

– des lois de Poisson composées $\mathrm{PC}(\Pi_n)$, décalées d'une constante $\int_{E_n} h(x)\Pi_n(\mathrm{d}x)$, pour $n \geq 1$.

La construction de Lévy-Itô du processus de Lévy càdlàg de loi μ à l'instant 1 consiste à remplacer les constantes a par des dérives $at, t \geq 0$, la variable normale par un mouvement brownien, et les lois de Poisson composées par les processus de Poisson composés correspondant.

Lemme 4.8 *Soient $(Y_t^n, t \geq 0), n \geq 1$ des Processus de Poisson composés indépendants d'intensités respectives $\Pi_n, n \geq 1$. Soit $(\mathcal{F}_t, t \geq 0)$ une filtration dans laquelle ces processus restent des processus de Poisson composés indépendants. On note $b_n = \int_{\mathbb{R}^d} x\Pi_n(\mathrm{d}x)$. Alors pour chaque n, $(Y_t^n - b_n t, t \geq 0)$ est une $\mathcal{F}_t$-martingale, et la suite $M^n = (\sum_{1\leq i\leq n}(Y_t^n - b_n t), t \geq 0)$ converge p.s. uniformément sur les compacts vers une martingale càdlàg $\mathbb{L}^2$, appelée $(M_t, t \geq 0)$.*

Preuve. Nous faisons la preuve dans le cas $d = 1$, le cas général nécessitant de travailller coordonnée par coordonnée. Le fait que M^n soit une martingale est une conséquence du Lemme 4.7 appliqué à l'identité, qui est Π_n-intégrable, et même dans $\mathbb{L}^2(\Pi_n)$. On applique ensuite un critère de Cauchy en écrivant pour $n > m$,

$$\mathbb{E}[(M_t^n - M_t^m)^2] = \sum_{i=m+1}^{n} \mathbb{E}[(M_t^i)^2] = \sum_{i=m+1}^{n} t\Pi_n(x^2) \leq \Pi(x^2 \mathbb{1}_{\{|x|\leq 1/m\}}) \underset{m\to\infty}{\to} 0,$$

ce qui montre que M_t^n converge dans $\mathbb{L}^2$ pour chaque t vers une limite M_t. Il est alors facile de voir que le processus $(M_t, t \geq 0)$ est une $\mathcal{F}_t$-matingale en passant à la limite dans $\mathbb{E}[M_{t+s}^n | \mathcal{F}_t] = M_t^n$.

Pour obtenir la convergence uniforme sur les compacts, on utilise l'inégalité $\mathbb{L}^2$ de Doob sur la martingale $M_t - M_t^n$:

$$\mathbb{E}\left[\sup_{0\leq s\leq t}(M_s - M_s^n)^2\right] \leq 4\mathbb{E}[(M_t - M_t^n)^2] = 4\int_{|x|\leq(n+1)^{-1}} x^2\Pi_n(\mathrm{d}x).$$

En particulier, la limite est càdlàg comme limite uniforme de processus càdlàg. □

Il est facile de constater que $((M_t)^2 - t\int_{|x\geq 1|} x^2\Pi(\mathrm{d}x), t \geq 0)$ est également une martingale.

Théorème 4.4 (Théorème de représentation de Lévy-Itô) *Soit $(X_t, t \geq 0)$ un processus de Lévy càdlàg de semigroupe associé à une mesure ID μ d'exposant caractéristique $\psi_{a,\Sigma,\Pi}$ comme ci-dessus. On définit pour $n \geq 0$*

$$Y_t^n = \sum_{0\leq s\leq t} \Delta X_s \mathbb{1}_{\{\Delta X_s \in E_n\}}, \qquad t \geq 0.$$

Alors les Y^n sont des processus de Poisson composés indépendants d'intensités respectives Π_n, et il existe un mouvement brownien $(B_t, t \geq 0)$ de matrice de covariance Σ, indépendant des Y^n, tel que

$$X_t = at + B_t + Y_t^0 + \sum_{n\geq 1}\left(Y_t^n - t\int_{\mathbb{R}^d} x\Pi_n(\mathrm{d}x)\right), \qquad t \geq 0$$

où la somme définit une martingale de carré intégrable avec convergence uniforme sur les compacts p.s.

Un corollaire important de ce théorème est que la mesure ponctuelle $\sum_{s\geq 0}\delta_{(s,\Delta X_s)}$ est une mesure de Poisson d'intensité $\mathrm{d}t\otimes\Pi(\mathrm{d}x)$, les processus de saut des processus de Poisson composés apparaissant dane le théorème précédent étant des processus ponctuels de Poisson indépendants d'intensités respectives $\Pi_n, n\geq 0$.

Preuve. Il est aisé d'après le Lemme 4.8 et la discussion qui le précède de montrer que si $Y^n, n\geq 0$ et B sont comme dans l'énoncé du théorème, alors le processus

$$at+B_t+Y_t^0+\sum_{n\geq 1}\left(Y_t^n-t\int_{\mathbb{R}^d}x\Pi_n(\mathrm{d}x)\right)\,,\qquad t\geq 0$$

définit un processus de Lévy càdlàg, et où la convergence de la série est uniforme sur les compacts. Or, le processus $(\Delta X_s\mathbb{1}_{\{\Delta X_s\in E_n\}}, s\geq 0)$ des saut de X tombant dans E_n est un processus mesurable d'un processus càdlàg X, et on en conclut que ce dernier est un processus de Poisson composé d'intensité Π_n, puis que la série de martingales compensées converge p.s., à nouveau d'après le Lemme 4.8. Une fois Y^0 et M construits, on en déduit que $X_t-M_t-Y_t^0-at$ est un mouvement brownien de covariance Σ, indépendant. $\square$

Remarque. On peut se demander si la compensation des sauts de E_n par la dérive de pente $\Pi_n(\mathrm{Id})$ est nécessaire. Notons que si l'on peut s'en passer, alors pour tout s,t, la quantité

$$\sum_{s<u\leq t}\Delta X_u$$

doit être finie. Comme on ne peut pas ordonner les termes non nuls de cette série dès qu'il y en a une infinité, le seul sens raisonnable à donner à cette convergence est celui de la convergence absolue. On arrive donc au critère que $\sum_{0\leq s\leq t}|\Delta X_s|<\infty$ pour tout t, ce qui advient si et seulement si Π intègre $|x|$ au voisinage de 0 (exercice). Dans ce cas, la partie martingale de saut du processus de Lévy est à variation finie.

Si cette dernière condition n'est pas vérifiée, la compensation est nécessaire, et le processus M obtenu est à variation infinie (il ne peut pas s'obtenir comme différence de deux processus croissants), et en particulier il n'est pas somme de ses sauts (même si par exemple la mesure Π est symétrique).

Corollaire 4.1 *Un processus de Lévy continu est nécessairement un mouvement brownien avec dérive.*

4.4 Lois et processus stables

Dans ce paragraphe, nous donnons quelques résultats sur une classe importante de lois ID, appellées lois stables. La plupart des résultats est donnée sans démonstration, pour lesquelles nous renvoyons à [5, 6, 7, 2]. Dans toute cette partie on se place en dimension $d=1$.

Définition 4.2 *Une loi μ (ou toute v.a. de loi μ) est dite stable si pour tout $n \geq 2$, il existe a_n, b_n tels que si $X_1, \ldots, X_n$ sont des copies indépendantes de X, alors $X_1 + \ldots + X_n$ a même loi que $a_n X + b_n$.*

En particulier, il est immédiat que toute loi stable est ID. Un exemple fondamental de loi stable est donné par les lois normales, puisqu'on a

$$\mathcal{N}(a_1, \sigma_1^2) * \mathcal{N}(a_2, \sigma_2^2) = a_1 + a_2 + \sqrt{\sigma_1^2 + \sigma_2^2}\mathcal{N}(0, 1).$$

La propriété fondamentale des lois stables est

Proposition 4.2 *Soit $(X_n, n \geq 1)$ une suite i.i.d. On suppose qu'il existe $\alpha_n, \beta_n, n \geq 1$ tels que*

$$\frac{X_1 + \ldots + X_n}{\alpha_n} - \beta_n \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} Y,$$

pour une certaine v.a. Y. Alors Y est stable.

On dit alors que X_1 est dans le domaine d'attraction de la loi stable Y. La réciproque, i.e. que toute loi stable peut s'obtenir comme une telle limite, est évidente par définition. Autrement dit, les lois stables apparaissent naturellement comme généralisation de la loi normale pour un théorème central limite. On voit en particulier que les lois normales sont les seules lois stables (non triviales) de variance finie, puisque de telles variables satisfont le TCL standard.

La classification des lois stables, c'est-à-dire l'écriture systématique de la formule de Lévy-Khintchine pour ces lois, commence par le lemme suivant. On dit que la fonction $f : \mathbb{R}_+ \to \mathbb{R}$ varie régulièrement à l'infini avec exposant $\alpha \geq 0$, si pour tout $\lambda > 0$ on a

$$\frac{f(\lambda x)}{f(x)} \underset{x\to\infty}{\to} \lambda^\alpha.$$

Lemme 4.9 *Dans la Définition 4.2, il existe un $\alpha \in]0, 2]$ tel que $a_n = n^{1/\alpha}$ pour tout n. On dit alors que la loi μ est stable d'indice α.*

Dans la Proposition 4.2, si la limite est stable d'indice α, alors la suite α_n varie régulièrement à l'infini, avec exposant α.

En particulier, les lois normales sont stables d'indice 2. Nous avons également croisé des lois stables d'indice 1, avec les lois de Cauchy symétriques de densité $a/(\pi(a^2 + x^2))$ sur $\mathbb{R}$.

Théorème 4.5 (i) *Les lois stables d'indice $\alpha = 2$ sont les lois normales. Si μ est stable d'indice $\alpha \in]0, 2[$, alors son exposant caractéristique est $\psi_{a,0,\Pi}$ pour un $a \in \mathbb{R}$, et il existe $c_+, c_- \in \mathbb{R}_+$ non toutes deux nulles telles que*

$$\Pi(\mathrm{d}x) = \frac{c_+}{x^{1+\alpha}} \mathbb{1}_{\{x>0\}} \mathrm{d}x + \frac{c_-}{(-x)^{1+\alpha}} \mathbb{1}_{\{x<0\}} \mathrm{d}x.$$

(ii) *À recentrage et multiplication par une constante près, l'exposant de Laplace d'une loi μ stable d'indice α est*

$$\begin{cases} -u^2 & \textit{si } \alpha = 2 \\ -|u|^\alpha \left(1 - \mathrm{i}\beta \operatorname{sgn}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right) & \textit{si } \alpha \in]0,2[\setminus\{1\}, \ \textit{pour un } \beta \in [-1,1], \\ -|u| \left(1 - \mathrm{i}\beta \operatorname{sgn}(t) \frac{2}{\pi} \log|t|\right) & \textit{si } \alpha = 1, \ \textit{pour un } \beta \in [-1,1]. \end{cases}$$

4.5 Exercices pour le chapitre 4

Exercice 1 — Lois ID dans $\mathbb{Z}$
La loi μ, portée par $\mathbb{Z}$, est dite $\mathbb{Z}$-indéfiniment divisible ($\mathbb{Z}$-ID) si pour tout $n \geq 1$, il existe μ_n une loi *portée par* $\mathbb{Z}$ telle que $\mu_n^{*n} = \mu$.

(i) Montrer qu'une loi ID portée par $\mathbb{Z}$ n'est pas nécessairement $\mathbb{Z}$-ID.

(ii) Montrer qu'une loi de Poisson composée d'intensité ν, où ν est une mesure finie sur $\mathbb{Z}$, est $\mathbb{Z}$-ID.

(iii) Réciproquement, soit μ une loi $\mathbb{Z}$-ID. À l'aide de la formule de Lévy-Khintchine, montrer que μ est une loi de Poisson composée.

(iv) Montrer que les lois géométriques sont $\mathbb{Z}$-ID, et déterminer leur mesure d'intensité.

Exercice 2 — Les lois ID sont les limites de tableaux triangulaires
(i) Montrer que la limite étroite d'une suite $(\mu^{(n)}, n \geq 1)$ de lois ID est elle-même ID. Pour cela, on écrira $\mu^{(n)}$ sous la forme $(\mu_p^{(n)})^{*p}$ pour tout $p \geq 1$, et on montrera que les suites $(\mu_p^{(n)}, n \geq 1)$ sont uniformément tendues.

(ii) Montrer qu'une loi de probabilités μ sur $\mathbb{R}$ est ID si et seulement si pour tout n, il existe une loi de probabilités μ_n, de sorte que la suite $(\mu_n^{*n}, n \geq 1)$ converge étroitement vers μ lorsque $n \to \infty$. Pour le sens réciproque, on pourra constater que $\mu_{np}^{*np} = (\mu_{np}^{*n})^{*p}$ pour tout $n, p \geq 1$ entiers, et s'inspirer de la preuve du fait que toute limite étroite de lois ID est ID.

Exercice 3 — Limites de tableaux triangulaires
Montrer par des méthodes similaires à celles de la preuve du lemme 4.4 que toute limite de suite X_n s'écrivant sous la forme $X_{n,1} + \ldots + X_{n,n}$ avec les $X_{n,i}$ i.i.d. est nécessairement ID.

Même exercice mais en remplaçant i.i.d. par "indépendantes" et en ajoutant l'hypothèse

$$\max_{1 \leq i \leq n} P(|X_{n,i}| > \varepsilon) \underset{n \to \infty}{\to} 0,$$

ce pour tout $\varepsilon > 0$. Donner des contre-exemples sans cette hypothèse.

Exercice 4 — Formule de Lévy-Khintchine pour les lois ID sur $\mathbb{R}_+$
Soit μ une mesure de probabilités ID sur $\mathbb{R}_+$. La formule de Lévy-Khintchine donne son exposant caractéristique

$$\psi(u) = \mathrm{i}au - \frac{\sigma^2}{2}u^2 + \int_{\mathbb{R}} (e^{\mathrm{i}ux} - 1 - \mathrm{i}uh(x))\pi(\mathrm{d}x)\,,$$

où h est une fonction mesurable bornée, nulle au voisinage de l'infini et égale à l'identité au voisinage de 0.

(i) Montrer que $\sigma = 0$.

(ii) Soit μ_n la loi de Poisson composée d'intensité $\pi(\mathrm{d}x)\mathbb{1}_{\{|x|\geq 1/n\}}$ et

$$b_n = \int_{\{|x|\geq 1/n\}} h(x)\pi(\mathrm{d}x)\,.$$

Montrer que μ s'écrit $\delta_a * \delta_{-b_n} * \mu_n * \mu'_n$ avec μ'_n qui converge étroitement vers δ_0, et en déduire que $\pi(\mathbb{R}_-) = 0$.

(iii) En déduire que $a - b_n \geq 0$ pour tout n, puis que b_n converge vers une limite b telle que $\mathrm{d} = a - b \geq 0$ quand $n \to \infty$. En conclure que $\int_{\mathbb{R}_+} (x \wedge 1)\pi(\mathrm{d}x) < \infty$, et enfin que pour tout $\lambda \geq 0$, l'exposant de Laplace de μ est donné par

$$-\log\left(\int_{\mathbb{R}_+} \mu(\mathrm{d}x)e^{-\lambda x}\right) = \mathrm{d}\lambda + \int_{\mathbb{R}_+} (1 - e^{-\lambda x})\pi(\mathrm{d}x)\,.$$

Exercice 5 — Quelques représentations de Lévy-Khintchine
(i) Donner la représentation de Lévy-Khintchine de l'exposant de Laplace d'une loi gamma de paramètres (a, θ), donnée par $\mu(\mathrm{d}x) = \Gamma(a)^{-1}\theta^a x^{a-1}e^{-\theta x}\mathrm{d}x\mathbb{1}_{\{x\geq 0\}}$. On pourra calculer la limite étroite de $n\mu_n$ où $\mu_n^{*n} = \mu$.

(ii) Montrer que la loi de Cauchy de paramètre $a > 0$, de densité $a\mathrm{d}x/\pi(a^2 + x^2)$, est indéfiniment divisible, et donner la représentation de Lévy-Khintchine de son exposant caractéristique. On rappelle que la transformée de Fourier de la loi de Cauchy de paramètre a est $\hat{\mu}(u) = \exp(-a|u|)$.

(iii) Soit $(B_t, t \geq 0)$ un mouvement Brownien standard. On note $T_a = \inf\{t \geq 0 : B_t = a\}$, pour $a \geq 0$. Montrer que la loi de T_a est ID. On rappelle une conséquence du principe de réflexion, selon laquelle T_a a même loi que $(a/B_1)^2$. Déterminer la représentation de Lévy-Khintchine de son exposant de Laplace (indice : cet exposant est $\psi(\lambda) = a\sqrt{2\lambda}$).

(iv) Soit $\alpha \in]0, 2]$. Montrer que $u \mapsto -|u|^\alpha$ est l'exposant caractéristique d'une loi ID, en déterminant sa représentation de Lévy-Khintchine. Cette loi est appelée loi stable symétrique d'exposant α. On notera que le cas $\alpha = 2$ est particulier.

Exercice 6 — Moments et mesure d'intensité
Soit μ une mesure de probabilités ID, et π la mesure d'intensité apparaissant dans la représentation de Lévy-Khintchine. Montrer que

$$\int_{\mathbb{R}} |x|^p \mu(\mathrm{d}x) < \infty \qquad \Longleftrightarrow \qquad \int_{\{|x|\geq 1\}} |x|^p \pi(\mathrm{d}x) < \infty \,.$$

Pour cela, on pourra considérer d'abord le cas où π est de support compact. Puis on pourra écrire $\mu = \mu' * \mu''$ où μ' est la loi de Poisson composée d'intensité $\pi(\mathrm{d}x)\mathbb{1}_{\{|x|\geq 1\}}$, et où μ'' admet des moments de tous ordres.

Exercice 7 — Processus de Poisson composés dans $\mathbb{N}$
Soit ν une mesure finie portée par $\mathbb{N}^* = \{1, 2, 3, \ldots\}$. Montrer que le processus de Poisson composé de mesure intensité ν a même loi qu'une somme pondérée de processus de Poisson (standard) indépendants, d'intensités respectives $\nu(n), n \in \mathbb{N}^*$.

Exercice 8 — Mouvement brownien plan et processus de Cauchy
Soit $B_t = (B_t^{(1)}, B_t^{(2)}), t \geq 0$ un mouvement brownien standard de dimension 2. On note $T_a = \inf\{t \geq 0 : B_t^{(2)} = a\}$ et $X_a = B_{T_a}^{(1)}, a \geq 0$. Montrer que $(X_a, a \geq 0)$ est un processus de Lévy et que pour tout $a \geq 0$, la loi de X_a est la loi de Cauchy de paramètre a. On pourra utiliser certains points rappelés dans un exercice précédent.

Exercice 9 — Indépendance de processus de Poisson composés
Soit $(X_t^{(1)}, X_t^{(2)})_{t\geq 0}$ un processus de Poisson composé à valeurs dans $\mathbb{R}^2$, dont la mesure d'intensité est ν, mesure finie sur $\mathbb{R}^2$, ne chargeant pas $(0, 0)$.

(i) Montrer que les processus $(X_t^{(1)}, t \geq 0)$ et $(X_t^{(2)}, t \geq 0)$ sont deux processus de Poisson composés dont on précisera les mesures d'intensité.

(ii) Montrer que les processus $X^{(1)}$ et $X^{(2)}$ sont indépendants si et seulement si ν est une mesure de la forme $\nu^{(1)} \otimes \delta_0 + \delta_0 \otimes \nu^{(2)}$, c'est-à-dire que ν ne charge que les axes $\{x = 0\}, \{y = 0\}$ de $\mathbb{R}^2$. On pourra penser à utiliser les fonctions caractéristiques. Montrer que cela équivaut encore au fait que $X^{(1)}$ et $X^{(2)}$ ne sautent p.s. jamais au même instant.

Exercice 10 — Moments des lois ID et mesure de Lévy
Montrer qu'une loi ID associée à a, Σ, Π a un moment d'ordre $p > 0$ si et seulement si $\int_{|x|\geq 1} |x|^p \Pi(\mathrm{d}x) < \infty$.

Exercice 11 — Quelques triplets caractéristiques explicites
Déterminer le triplet caractéristique des lois ID suivantes.
- La loi gamma(a, θ), $a, \theta > 0$.
- La loi géométrique de paramètre $a \in (0, 1)$ (la loi de probabilité intensité de la loi de Poisson composée associée est appelée loi logarithmique).

• La loi de Cauchy de paramètre a, dont la densité est $a/(\pi(a^2+x^2))$ sur $\mathbb{R}$.
• La loi du premier temps d'atteinte de 1 par un mouvement brownien (voir aussi le paragraphe 4.4).

Exercice 12 —
Soit B un mouvement Brownien standard et τ un temps de loi exponentielle indépendant de B. Montrer que (τ, B_τ) est ID, et donner son triplet caractéristique.

Exercice 13 — Changement d'échelle et processus stables
Soit $(X_t, t \geq 0)$ un processus de Lévy dont la mesure de Lévy est de la forme

$$\frac{c_+ \mathrm{d}x}{x^{1+\alpha}} \mathbb{1}_{\{x>0\}} + \frac{c_- \mathrm{d}x}{(-x)^{1+\alpha}} \mathbb{1}_{\{x<0\}} ,$$

avec $c_+ + c_- > 0$, et dont la partie gaussienne est nulle. A quelle condition sur la dérive (le coefficient réel a de la formule de Lévy-Khintchine) le processus X vérifie-t-il la propriété d'invariance par changement d'échelle suivante ?

$$\forall \lambda > 0 , \qquad (X_{\lambda t}, t \geq 0) \overset{\text{(loi)}}{=} (\lambda^{1/\alpha} X_t, t \geq 0) .$$

On dit alors que $(X_t, t \geq 0)$ est un processus stable d'indice α.

Exercice 14 — Avec quelques connaissances de processus de Markov
Si μ est ID de fonction caractéristique $\phi = \exp(\psi)$ où ψ est continue nulle en 0, alors pour tout $t \geq 0$, $\phi_t = \exp(t\psi)$ est la fonction caractéristique d'une loi μ_t. Montrer que $(\mu_t, t \geq 0)$ est un semigroupe de convolution ayant la propriété de Feller. En déduire par des méthodes de processus de Markov quà toute loi ID on peut associer un processus de Lévy à trajectoires càdlàg admettant $(\mu_t, t \geq 0)$ pour semigroupe.

Réciproquement, montrer que si $(X_t, t \geq 0)$ est un processus de Lévy sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ tel que $(\omega, t) \mapsto X_t(\omega)$ est mesurable de $\Omega \times \mathbb{R}_+$ dans $\mathbb{R}^d$, alors X est le processus de Lévy associé à la loi ID μ comme précédemment, où μ est la loi de X_1.

Références

[1] P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999.

[2] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1989.

[3] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[4] S. N. Ethier and T. G. Kurtz. *Markov processes.* Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[5] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[6] B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables.* Translated from the Russian, annotated, and revised by K. L. Chung. With appendices by J. L. Doob and P. L. Hsu. Revised edition. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont., 1968.

[7] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, second edition, 2003.

[8] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability.* The Clarendon Press Oxford University Press, New York, 1993. Oxford Science Publications.

[9] K. R. Parthasarathy. *Probability measures on metric spaces.* Probability and Mathematical Statistics, No. 3. Academic Press Inc., New York, 1967.

[10] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, third edition, 1987.