Convergence de mesures, processus de Poisson et processus de Lévy

Pierre-Loïc Méliot

1er septembre 2016

Chapitre 1

Convergence de mesures de probabilité

Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilité. On rappelle qu'une variable aléatoire X à valeurs dans un espace mesurable $(\mathfrak{X}, \mathscr{B})$ est une application mesurable $X : \Omega \to \mathfrak{X}$. Le plus souvent, l'espace des variables $\mathfrak{X}$ est un espace topologique, et $\mathscr{B} = \mathscr{B}(\mathfrak{X})$ est l'ensemble des parties boréliennes de $\mathfrak{X}$, c'est-à-dire la plus petite tribu sur $\mathfrak{X}$ qui contient les ouverts et les fermés. Deux exemples importants sont :

— $\mathfrak{X} = \mathbb{Z}$, muni de la topologie discrète et donc de la tribu des parties $\mathscr{B}(\mathbb{Z}) = \mathscr{P}(\mathbb{Z})$ (variables aléatoires entières) ;

— $\mathfrak{X} = \mathbb{R}$, muni de sa topologie usuelle et de la tribu borélienne $\mathscr{B}(\mathbb{R})$ (variables aléatoires réelles).

Dans la théorie des probabilités et des modèles aléatoires, il est utile et parfois nécessaire de travailler avec des espaces $\mathfrak{X}$ plus compliqués que $\mathbb{Z}$ ou $\mathbb{R}$. Par exemple, une marche aléatoire discrète $X = (x_0, x_1, x_2, \ldots, x_n, \ldots)$ peut être considérée comme une variable aléatoire dans $\mathfrak{X} = \mathbb{Z}^{\mathbb{N}}$, et le mouvement brownien $X = (X_t)_{t \geq 0}$ est une variable aléatoire à valeurs dans l'espace $\mathfrak{X} = \mathscr{C}(\mathbb{R}_+, \mathbb{R})$ des fonctions continues sur $\mathbb{R}_+$. L'un des objectifs de ce chapitre est de préciser quels espaces topologiques $\mathfrak{X}$ sont adéquats en théorie des probabilités. Plus précisément, on souhaitera travailler avec des espaces de variables $\mathfrak{X}$ dans lesquels on dispose de résultats généraux sur les divers modes de convergence d'une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$. Dans la section 1.1, on rappelle quels sont les modes de convergence usuels pour une suite de variables aléatoires, et quels sont les liens entre ces modes. Dans les paragraphes 1.2 et 1.3, on étudie plus en détail la convergence en loi, et on montre que si $\mathfrak{X}$ est un espace topologique polonais, alors la topologie associée à ce mode de convergence a de bonnes propriétés. Cette étude revient à comprendre la géométrie de l'espace $\mathscr{M}^1(\mathfrak{X})$ des mesures de probabilité sur $\mathfrak{X}$; c'est l'autre objectif majeur de ce chapitre.

Remarque 1.1. L'étude de la convergence de variables aléatoires dans un espace topologique $\mathfrak{X}$ est importante même lorsque au final on souhaite étudier une seule variable aléatoire $X \in \mathfrak{X}$. En effet, de nombreux modèles aléatoires sont construits par approximation par des modèles discrets, et en prenant des limites d'échelle. Par exemple, le mouvement brownien est une limite d'échelle universelle pour les marches aléatoires à accroissements i.i.d. (voir le chapitre 3), et il est donc utile de comprendre ce qu'est la convergence de variables aléatoires à valeurs dans $\mathscr{C}(\mathbb{R}_+, \mathbb{R})$, même si finalement on souhaite uniquement étudier les propriétés d'un mouvement brownien.

1.1 Modes de convergence des variables aléatoires

On fixe de nouveau un espace de probabilité $(\Omega, \mathscr{F}, \mathbb{P})$, et un espace topologique $\mathfrak{X}$ muni de la tribu des boréliens $\mathscr{B}(\mathfrak{X})$; et on considère une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ à valeurs dans $\mathfrak{X}$. L'objectif de cette section est de rappeler les divers modes de convergence possibles pour cette suite. Ils sont résumés dans le diagramme suivant :

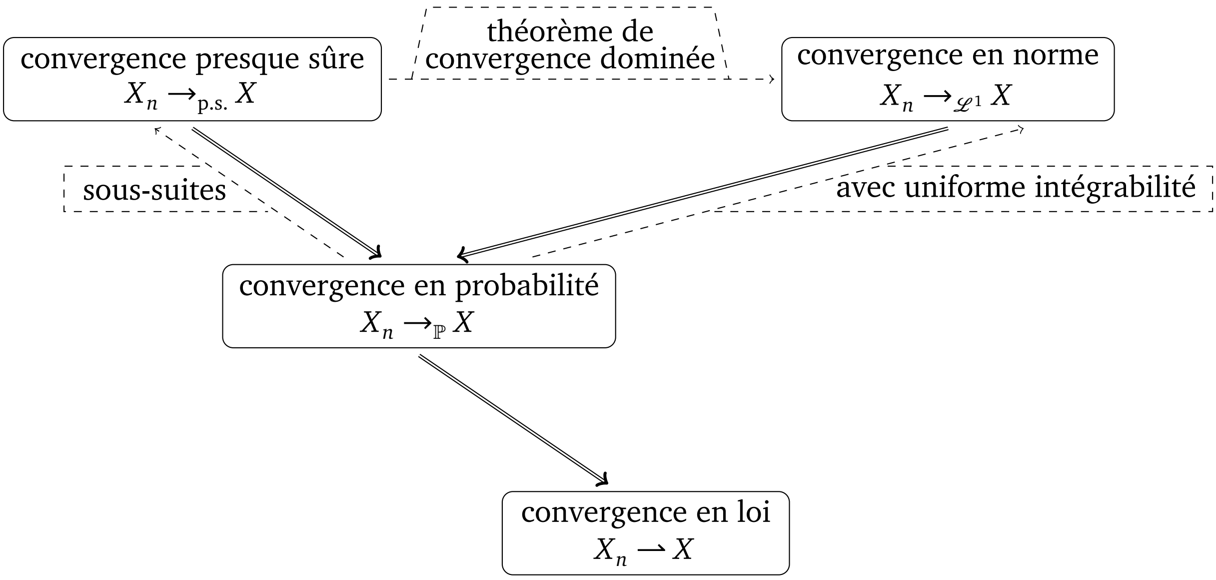


FIGURE 1.1 – Les modes de convergence d'une suite de variables aléatoires.

▷ **Convergence presque sûre et convergence en probabilité.** On rappelle que dans un espace topologique $\mathfrak{X}$, une suite $(x_n)_{n\in\mathbb{N}}$ converge vers x si, pour tout ouvert U contenant x, $x_n \in U$ pour $n \geq n_0$ assez grand (notation : $x_n \to_{n\to\infty} x$, ou plus simplement $x_n \to x$). Si la topologie de $\mathfrak{X}$ est métrisée par une distance d, alors cette condition est équivalente à la formulation plus classique :

$$\forall \varepsilon > 0,\ \exists n_0,\ \forall n \geq n_0,\ d(x_n, x) \leq \varepsilon.$$

La convergence presque sûre est l'adaptation probabiliste la plus simple de cette notion :

Définition 1.2 (Convergence presque sûre)**.** *Une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers une variable X si*

$$\mathbb{P}[\{\omega \in \Omega \,|\, X_n(\omega) \to_{n\to\infty} X(\omega)\}] = 1.$$

On notera dans ce cas $X_n \to_{\text{p.s.}} X$.

Un cas particulier est celui où $\mathfrak{X}$ est un espace métrique muni d'une distance d. Dans ce cas, $X_n \to_{\text{p.s.}} x$ si et seulement si

$$\mathbb{P}[\{\omega \in \Omega \,|\, d(X_n(\omega), X(\omega)) \to_{n\to\infty} 0\}] = 1.$$

Cette condition est encore équivalente à :

$$\forall \varepsilon > 0,\ \lim_{n_0\to\infty} \mathbb{P}[\forall n \geq n_0,\ d(X_n,X) \leq \varepsilon] = 1.$$

Exemple 1.3. Soit $(A_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires intégrables, indépendantes et de même loi dans $\mathfrak{X} = \mathbb{R}^d$. On pose $X_n = \frac{1}{n}(A_1 + A_2 + \cdots + A_n)$. La loi forte des grands nombres assure que

$$X_n \to_{\text{p.s.}} \mathbb{E}[A_1].$$

Une notion plus faible de convergence est celle de convergence en probabilité. Dans ce qui suit, on suppose que $(\mathfrak{X}, d)$ est un espace métrique. Les ouverts de $\mathfrak{X}$ sont donc les parties U telles que, pour tout $x \in U$, il existe $\varepsilon_x > 0$ tel que la boule

$$B_{(x,\varepsilon_x)} = \{y \in \mathfrak{X} \,|\, d(x,y) < \varepsilon_x\}$$

soit incluse dans U.

Définition 1.4 (Convergence en probabilité). *Une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ à valeurs dans $\mathfrak{X}$ converge en probabilité vers une variable X si, pour tout $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}[d(X_n,X) \leq \varepsilon] = 1.$$

On notera dans ce cas $X_n \to_{\mathbb{P}} X$.

Exemple 1.5. Supposons $\mathfrak{X} = \mathbb{R}$, et soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires telle que $\mathbb{E}[(X_n - X)^2] \to 0$. Alors, l'inégalité de Bienaymé–Chebyshev

$$\mathbb{P}[|X_n - X| > \varepsilon] \leq \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2}$$

prouve que $X_n \to_{\mathbb{P}} X$.

Notons $(A_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telles que

$$\mathbb{P}[A_n = n] = \mathbb{P}[A_n = -n] = \frac{1}{2n\,\log(n+1)} \qquad ; \qquad \mathbb{P}[A_n = 0] = 1 - \frac{1}{n\,\log(n+1)}.$$

On a $\mathbb{E}[A_n] = 0$ et $\mathrm{var}(A_n) = \frac{n}{\log(n+1)}$, donc, si $X_n = \frac{A_1+A_2+\cdots+A_n}{n}$, alors

$$\mathbb{E}[X_n] = 0 \qquad ; \qquad \mathrm{var}(X_n) = \frac{1}{n^2}\sum_{k=1}^{n} \frac{k}{\log(k+1)} \simeq_{n\to\infty} \frac{1}{2\,\log n}.$$

Par conséquent, $X_n \to_{\mathbb{P}} 0$ (loi faible des grands nombres). En revanche, on n'a pas $X_n \to_{\text{p.s.}} 0$. En effet, considérons les événements $E_n = \{|A_n| = n\}$. Ces événements sont tous indépendants, et $\sum_{n=1}^{\infty} \mathbb{P}[E_n] = \sum_{n=1}^{\infty} \frac{1}{n\,\log(n+1)} = +\infty$, donc, par le lemme de Borel–Cantelli, $\mathbb{P}[\limsup_n E_n] = 1$. Ainsi, avec probabilité 1, une infinité de variables A_n sont non nulles. Autrement dit, pour presque tout $\omega \in \Omega$, on peut trouver une suite croissante $(n_k(\omega))_{k\geq 1}$ telle que $|A_{n_k(\omega)}| = n_k(\omega)$. Fixons une telle suite aléatoire, et considérons la suite des signes

$$\varepsilon_k = \mathrm{sgn}(A_{n_k})\,\mathrm{sgn}(A_1 + \cdots + A_{n_k - 1}),$$

avec pour convention $\mathrm{sgn}(0) = +1$. On voit facilement que les ε_k sont des variables aléatoires de Bernoulli indépendantes, avec $\mathbb{P}[\varepsilon_k = 1] = \mathbb{P}[\varepsilon_k = -1] = \frac{1}{2}$. On peut donc extraire de $(n_k(\omega))_{k\geq 1}$ une sous-suite aléatoire $(n_{k_l}(\omega))_{l\geq 1}$, telle que

$$\varepsilon_{k_l} = 1 \qquad ; \qquad \left|A_{n_{k_l}}\right| = n_{k_l}.$$

Alors, $|X_{n_{k_l}}| \geq \frac{\left|A_{n_{k_l}}\right|}{n_{k_l}} = 1$, donc on peut toujours extraire de $(X_n)_{n\geq 1}$ une sous-suite aléatoire $(X_{n_{k_l}})_{l\geq 1}$ qui reste en valeur absolue plus grande que 1. Ainsi, la convergence en probabilité n'implique pas la convergence presque sûre.

On voit immédiatement que si une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires converge presque sûrement dans un espace métrique vers une variable X, alors elle converge aussi en probabilité. La proposition suivante précise cette implication, et le lien entre les deux modes de convergence :

Proposition 1.6. *On a $(X_n)_{n\in\mathbb{N}} \to_{\mathbb{P}} X$ si et seulement si, de toute sous-suite $(X_{n_k})_{k\in\mathbb{N}}$, on peut extraire une sous-sous-suite $(X_{n_{k_l}})_{l\in\mathbb{N}}$ qui converge presque sûrement vers X. En particulier, si $X_n \to_{\text{p.s.}} X$, alors $X_n \to_{\mathbb{P}} X$.*

Preuve. Supposons $X_n \to_{\mathbb{P}} X$, et montrons qu'on peut extraire une sous-suite $(X_{n_k})_{k\in\mathbb{N}}$ avec $X_{n_k} \to_{\text{p.s.}} X$. Comme $X_n \to_{\mathbb{P}} X$, on peut trouver une suite croissante $(n_k)_{k\in\mathbb{N}}$ telle que

$$\mathbb{P}\left[d(X_{n_k},X) \leq \frac{1}{2^k}\right] \geq 1 - \frac{1}{2^{k+1}}.$$

Alors, pour tout k_0,

$$\mathbb{P}\left[\forall k \geq k_0,\, d(X_{n_k},X) \leq \frac{1}{2^{k_0}}\right] \geq 1 - \sum_{k=k_0}^{\infty} \frac{1}{2^{k+1}} = 1 - \frac{1}{2^{k_0}}.$$

Comme ceci est vrai quelque soit k_0, $X_{n_k} \to_{\text{p.s.}} X$. Donc, si $X_n \to_{\mathbb{P}} X$, alors pour toute sous-suite $(X_{n_k})_{k\in\mathbb{N}}$, comme on a encore $X_{n_k} \to_{\mathbb{P}} X$, on peut extraire une sous-sous-suite $(X_{n_{k_l}})_{l\in\mathbb{N}}$ telle que $X_{n_{k_l}} \to_{\text{p.s.}} X$.

Réciproquement, supposons que $(X_n)_{n\in\mathbb{N}}$ ne converge pas en probabilité vers X. Alors, il existe une sous-suite $(X_{n_k})_{k\in\mathbb{N}}$ et un $\varepsilon > 0$ tel que

$$\forall k \in \mathbb{N},\ \mathbb{P}\left[d(X_{n_k},X) > \varepsilon\right] \geq \varepsilon.$$

Dans ce cas, aucune sous-sous-suite extraite de $(X_{n_k})_{k\in\mathbb{N}}$ ne peut converger presque sûrement vers X, car

$$\mathbb{P}\left[\forall l \geq l_0,\ d(X_{n_{k_l}},X) \leq \varepsilon\right] \leq \mathbb{P}\left[d(X_{n_{k_{l_0}}},X) \leq \varepsilon\right] \leq 1 - \varepsilon.$$

□

Remarque 1.7. La notion de convergence presque sûre est purement topologique, donc, si d et d' sont deux distances sur $\mathfrak{X}$ qui définissent la même topologie, alors $X_n \to_{\text{p.s.}} X$ dans $(\mathfrak{X}, d)$ si et seulement si $X_n \to_{\text{p.s.}} X$ dans $(\mathfrak{X}, d')$. Comme la convergence en probabilité est liée à la convergence presque sûre par la proposition précédente, on a le même résultat pour la convergence en probabilité : $X_n \to_{\mathbb{P}} X$ dans $(\mathfrak{X}, d)$ si et seulement si $X_n \to_{\mathbb{P}} X$ dans $(\mathfrak{X}, d')$. Ainsi, la notion de convergence en probabilité peut être étendue à tout espace topologique métrisable ; elle ne dépend alors que de la topologie, et pas du choix de la métrique.

▷ Convergence dans $\mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$. Lorsque $\mathfrak{X} = \mathbb{R}^d$, ou plus généralement lorsque $\mathfrak{X}$ est un espace de Banach muni d'une norme $\|\cdot\|_{\mathfrak{X}}$ et de la topologie et de la tribu borélienne associées à cette norme, un autre mode classique de convergence des variables aléatoires est la convergence en norme. Rappelons que l'espace des fonctions intégrables $\mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$ est le complété de l'espace vectoriel des fonctions en escalier

$$\mathrm{Esc}_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P}) = \left\{ f = \sum_{i=1}^{n} 1_{A_i}\, x_i \;\middle|\; \forall i \in [\![1, n]\!],\ x_i \in \mathfrak{X} \text{ et } A_i \in \mathscr{F} \right\}$$

pour la norme

$$\left\| \sum_{i=1}^{n} 1_{A_i}\, x_i \right\|_{\mathscr{L}^1} = \sum_{i=1}^{n} \mathbb{P}[A_i]\, \|x_i\|_{\mathfrak{X}} \quad \text{si les parties } A_i \text{ sont disjointes.}$$

En particulier, par construction, $\mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$ est un espace de Banach, dans lequel l'espace des fonctions en escalier $\mathrm{Esc}_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$ est dense. Si $\mathfrak{X} = \mathbb{R}$, on notera plus simplement $\mathscr{L}^1_{\mathbb{R}}(\Omega, \mathscr{F}, \mathbb{P}) = \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$.

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathfrak{X}$, avec chaque variable $X_n \in \mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$.

Définition 1.8 (Convergence en norme)**.** *On dit que $(X_n)_{n\in\mathbb{N}}$ converge en norme vers $X \in \mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$ si*

$$\|X_n - X\|_{\mathscr{L}^1} = \mathbb{E}[\|X_n - X\|_{\mathfrak{X}}] = \int_{\Omega} \|X_n(\omega) - X(\omega)\|_{\mathfrak{X}}\, \mathbb{P}[d\omega]$$

tend vers 0. Autrement dit, $(X_n)_{n\in\mathbb{N}}$ converge vers X dans l'espace vectoriel normé $\mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$, et on note alors $X_n \to_{\mathscr{L}^1} X$.

Proposition 1.9. *Si $X_n \to_{\mathscr{L}^1} X$, alors $X_n \to_{\mathbb{P}} X$.*

Preuve. Par l'inégalité de Markov,

$$\mathbb{P}[\|X_n - X\|_{\mathfrak{X}} \geq \varepsilon] \leq \frac{\mathbb{E}[\|X_n - X\|_{\mathfrak{X}}]}{\varepsilon} \to_{n\to\infty} 0. \qquad \square$$

Théorème 1.10 (Convergence dominée). *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans l'espace de Banach $(\mathfrak{X}, \|\cdot\|_{\mathfrak{X}})$, et tel qu'existe une variable aléatoire réelle positive R avec $\mathbb{E}[R] < +\infty$ et*

$$\|X_n(\omega)\|_{\mathfrak{X}} \leq R(\omega)$$

pour tout n et pour presque tout ω. Si $X_n \to_{\text{p.s.}} X$, alors $X \in \mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$ et $X_n \to_{\mathscr{L}^1} X$. En particulier, $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

Preuve. Soit $\varepsilon > 0$. Comme R est intégrable, il existe un réel $M > 0$ tel que $\mathbb{E}[R\, 1_{R\geq M}] \leq \varepsilon$ (les fonctions en escalier ont évidemment cette propriété, et il suffit ensuite d'utiliser la densité des fonctions en escalier dans $\mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$). On note A la partie mesurable de Ω constituée des ω tels que $R(\omega) < M$. Comme $X_n \to_{\text{p.s.}} X$, il existe une sous-partie mesurable $A_\varepsilon \subset A$, et un entier n_0 tels que :

$$\mathbb{P}[A \setminus A_\varepsilon] \leq \frac{\varepsilon}{M} \qquad ; \qquad \forall \omega \in A_\varepsilon,\ \forall n \geq n_0,\ \|X_n(\omega) - X(\omega)\|_{\mathfrak{X}} \leq \varepsilon.$$

Alors, pour tous $n_1, n_2 \geq n_0$,

$$\begin{aligned}\mathbb{E}\left[\|X_{n_1} - X_{n_2}\|_{\mathfrak{X}}\right] &\leq \mathbb{E}\left[(\|X_{n-1} - X\|_{\mathfrak{X}} + \|X_{n_2} - X\|_{\mathfrak{X}}) 1_{A_\varepsilon}\right] + \mathbb{E}\left[2R\, 1_{A\setminus A_\varepsilon}\right] + \mathbb{E}\left[2R\, 1_{\Omega\setminus A}\right] \\ &\leq 2\varepsilon + 2M\, \frac{\varepsilon}{M} + 2\varepsilon = 6\varepsilon.\end{aligned}$$

Donc, $(X_n)_{n\in\mathbb{N}}$ est une suite de Cauchy dans $\mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$, et comme cet espace est complet, il existe une variable aléatoire intégrable Y telle que $X_n \to_{\mathscr{L}^1} Y$. En particulier, $X_n \to_{\mathbb{P}} Y$, et comme par ailleurs $X_n \to_{\text{p.s.}} X$, $X_n \to_{\mathbb{P}} X$, donc la suite $(X_n)_{n\in\mathbb{N}}$ converge en probabilité simultanément vers X et Y. Il est facile de voir que ceci implique $X = Y$ presque sûrement, et ainsi, X est intégrable et $X_n \to_{\mathscr{L}^1} X$. Finalement, comme l'espérance est une forme linéaire continue sur $\mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$, $\mathbb{E}[X_n] \to_{n\to\infty} \mathbb{E}[X]$. □

Pour préciser le lien entre convergence dans $\mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$ et convergence en probabilité, on a besoin de la notion d'uniforme intégrabilité. Une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires intégrables est dite uniformément intégrable si, pour tout $\varepsilon > 0$, on peut trouver un réel positif M tel que

$$\forall n \in \mathbb{N},\ \mathbb{E}\left[\|X_n\|_{\mathfrak{X}}\, 1_{\|X_n\|_{\mathfrak{X}} \geq M}\right] \leq \varepsilon.$$

Théorème 1.11 (Uniforme intégrabilité). *Une suite de variables aléatoires intégrables $(X_n)_{n\in\mathbb{N}}$ converge en norme vers $X \in \mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$ si et seulement si $X_n \to_{\mathbb{P}} X$ et $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable.*

Lemme 1.12. *Soit $X \in \mathscr{L}^1_{\mathfrak{X}}(\Omega, \mathscr{F}, \mathbb{P})$. On a*

$$\lim_{t\to 0} \sup_{A \,|\, \mathbb{P}[A] \leq t} \mathbb{E}[\|X\|_{\mathfrak{X}}\, 1_A] = 0.$$

Preuve. Raisonnons par l'absurde et supposons qu'il existe $\varepsilon > 0$ et une suite d'événements $A_n \in \mathscr{F}$ telle que $\mathbb{P}[A_n] \to 0$ et $\mathbb{E}[\|X\|_{\mathfrak{X}}\, 1_{A_n}] \geq \varepsilon$ pour tout $n \in \mathbb{N}$. Quitte à extraire une sous-suite de $(A_n)_{n\in\mathbb{N}}$, on peut supposer que $\sum_{n=0}^{\infty} \mathbb{P}[A_n]$ est finie. Alors, si $B_n = \bigcup_{m\geq n} A_m$, $(B_n)_{n\in\mathbb{N}}$ est une suite d'événements décroissante pour l'inclusion, avec

$$\mathbb{P}[B_n] \leq \sum_{m\geq n} \mathbb{P}[A_m] \to_{n\to\infty} 0,$$

et $\mathbb{E}[\|X\|_{\mathfrak{X}}\, 1_{B_n}] \geq \mathbb{E}[\|X\|_{\mathfrak{X}}\, 1_{A_n}] \geq \varepsilon$ pour tout n. Or, $X_n = X\, 1_{B_n}$ est une suite de variables aléatoires dominées par $R = \|X\|_{\mathfrak{X}}$, et telle que $X_n \to_{\text{p.s.}} 0$, donc $\lim_{n\to\infty} \mathbb{E}[\|X_n\|_{\mathfrak{X}}] = 0$ par le théorème de convergence dominée ; d'où une contradiction. $\square$

Preuve du théorème 1.11. Si $X_n \to_{\mathscr{L}^1} X$, on a déjà vu que $X_n \to_{\mathbb{P}} X$. Pour montrer l'uniforme intégrabilité, fixons $\varepsilon > 0$, et n_0 tel que pour tout $n \geq n_0$, $\|X_n - X\|_{\mathscr{L}^1} \leq \varepsilon$. Comme dans la preuve du théorème de convergence dominée, on peut trouver $M > 0$ tel que la famille finie de variables aléatoires $\{X, X_0, X_1, \ldots, X_{n_0-1}\}$ vérifie

$$\mathbb{E}\left[\|X\|_{\mathfrak{X}}\, 1_{\|X\|_{\mathfrak{X}}\geq M}\right] \leq \varepsilon \qquad ; \qquad \forall k < n_0,\ \mathbb{E}\left[\|X_k\|_{\mathfrak{X}}\, 1_{\|X_k\|_{\mathfrak{X}}\geq M}\right] \leq \varepsilon.$$

Quitte à augmenter la valeur de M, on peut aussi supposer grâce au lemme précédent que

$$\sup_{A\,|\,\mathbb{P}[A]\leq \frac{\varepsilon}{M}} \mathbb{E}[\|X\|_{\mathfrak{X}}\, 1_A] \leq \varepsilon.$$

Alors, $\mathbb{E}[\|X_n\|_{\mathfrak{X}}\, 1_{\|X_n\|_{\mathfrak{X}}\geq 2M}] \leq 3\varepsilon$ pour tout n. En effet, c'est clair si $n < n_0$, et sinon,

$$\begin{aligned}\mathbb{E}[\|X_n\|_{\mathfrak{X}}\, 1_{\|X_n\|_{\mathfrak{X}}\geq 2M}] &\leq \mathbb{E}[\|X_n - X\|_{\mathfrak{X}}] + \mathbb{E}\left[\|X\|_{\mathfrak{X}}\, 1_{\|X_n\|_{\mathfrak{X}}\geq 2M}\right]\\ &\leq \varepsilon + \mathbb{E}\left[\|X\|_{\mathfrak{X}}\, 1_{\|X_n\|_{\mathfrak{X}}\geq 2M,\, \|X\|_{\mathfrak{X}}\leq M}\right] + \mathbb{E}\left[\|X\|_{\mathfrak{X}}\, 1_{\|X\|\geq M}\right]\\ &\leq 2\varepsilon + \mathbb{E}\left[\|X\|_{\mathfrak{X}}\, 1_{\|X_n - X\|_{\mathfrak{X}}\geq M}\right].\end{aligned}$$

Or, par l'inégalité de Markov, l'événement $\{\omega \,|\, \|X_n(\omega) - X(\omega)\|_{\mathfrak{X}} \geq M\}$ a une probabilité inférieure à $\frac{\varepsilon}{M}$, donc $\mathbb{E}[\|X\|_{\mathfrak{X}}\, 1_{\|X_n - X\|_{\mathfrak{X}}\geq M}] \leq \varepsilon$, et finalement $\mathbb{E}[\|X_n\|_{\mathfrak{X}}\, 1_{\|X_n\|_{\mathfrak{X}}\geq 2M}] \leq 3\varepsilon$. Ainsi, $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable.

Supposons réciproquement que $X_n \to_{\mathbb{P}} X$ et que la suite est uniformément intégrable. Il est facile de voir que la suite $(X_n - X)_{n\in\mathbb{N}}$ est encore uniformément intégrable. Soit $\varepsilon > 0$, et M suffisamment grand tel que $\mathbb{E}[\|X_n - X\|_{\mathfrak{X}}\, 1_{\|X_n - X\|_{\mathfrak{X}}\geq M}] \leq \varepsilon$ pour tout n. On a alors

$$\begin{aligned}&\mathbb{E}[\|X_n - X\|_{\mathfrak{X}}]\\ &\leq \mathbb{E}\left[\|X_n - X\|_{\mathfrak{X}} 1_{\|X_n - X\|_{\mathfrak{X}}\leq \varepsilon}\right] + \mathbb{E}\left[\|X_n - X\|_{\mathfrak{X}} 1_{\varepsilon\leq \|X_n - X\|_{\mathfrak{X}}\leq M}\right] + \mathbb{E}\left[\|X_n - X\|_{\mathfrak{X}} 1_{\|X_n - X\|_{\mathfrak{X}}\geq M}\right]\\ &\leq \varepsilon + M\,\mathbb{P}[\|X_n - X\|_{\mathfrak{X}} \geq \varepsilon] + \varepsilon.\end{aligned}$$

Comme $X_n \to_{\mathbb{P}} X$, la probabilité $\mathbb{P}[\|X_n - X\|_{\mathfrak{X}} \geq \varepsilon]$ tend vers 0, donc $\|X_n - X\|_{\mathscr{L}^1} \to 0$. $\square$

▷ **Convergence en loi.** Un dernier mode de convergence classique pour les variables aléatoires est la convergence faible, ou en loi (on parle parfois aussi de convergence étroite). Dans ce qui suit, on suppose seulement que $\mathfrak{X}$ est un espace topologique.

Définition 1.13 (Convergence en loi). *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathfrak{X}$. On dit que $(X_n)_{n\in\mathbb{N}}$ converge en loi, ou faiblement vers une variable X si et seulement si, pour toute fonction continue bornée $f : \mathfrak{X} \to \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

On note dans ce cas $X_n \rightharpoonup X$.

Remarque 1.14. Dans cette définition, il n'est pas nécessaire que les variables X_n soient issues du même espace de probabilité $(\Omega, \mathscr{F}, \mathbb{P})$. En effet, $\mathbb{E}[f(X_n)]$ ne dépend que de la loi $\mu_n = (X_n)_*\mathbb{P}$ de X_n, qui est définie par

$$\mu_n[A \in \mathscr{B}(\mathfrak{X})] = \mathbb{P}[\{\omega \,|\, X_n(\omega) \in A\}].$$

Cette loi μ_n est une probabilité sur $\mathfrak{X}$, et la convergence en loi peut être comprise comme la convergence $\mu_n \rightharpoonup \mu$, où $\mu = X_*\mathbb{P}$ est la loi de X. Ce point de vue sera détaillé dans la suite de ce chapitre ; l'espace de probabilité $(\Omega, \mathscr{F}, \mathbb{P})$ peut être entièrement oublié avec ce point de vue.

Exemple 1.15. Soit $(A_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes, de même loi dans $\mathfrak{X} = \mathbb{R}^d$ et de carré intégrable. Si $m = \mathbb{E}[A_1]$, $\sigma^2 = \mathrm{var}(A_1)$ et

$$G_n = \frac{\sqrt{n}}{\sigma}\left(\frac{A_1 + A_2 + \cdots + A_n}{n} - m\right),$$

alors le théorème central limite assure que $G_n \rightharpoonup G$, où G suit une loi normale de densité $\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,dx$. Ce résultat sera détaillé dans le chapitre 2.

Si $(\mu_n)_{n\in\mathbb{N}}$ et μ sont des mesures de probabilité sur $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}))$, on notera $\mu_n \rightharpoonup \mu$ si $\mu_n(f) = \int_{\mathfrak{X}} f(x)\,\mu_n(dx)$ tend vers $\mu(f)$ pour toute fonction continue bornée f sur $\mathfrak{X}$. D'après la remarque précédente, si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires, alors $X_n \rightharpoonup X$ si et seulement si $\mu_n \rightharpoonup \mu$, où μ_n est la loi de X_n et μ est la loi de X. Le théorème suivant donne des conditions équivalents à $\mu_n \rightharpoonup \mu$ lorsque $\mathfrak{X}$ est un espace métrique (ou métrisable).

Théorème 1.16 (Portmanteau). *Soit $(\mathfrak{X}, d)$ un espace métrique. Les conditions suivantes sont équivalentes :*

1. *La suite de mesures de probabilités $(\mu_n)_{n\in\mathbb{N}}$ vérifie $\mu_n \rightharpoonup \mu$.*
2. *Pour toute fonction f bornée et uniformément continue sur $\mathfrak{X}$, $\mu_n(f) \to \mu(f)$.*
3. *Pour tout fermé F de $\mathfrak{X}$, $\limsup_n \mu_n(F) \leq \mu(F)$.*
4. *Pour tout ouvert U de $\mathfrak{X}$, $\liminf_n \mu_n(U) \geq \mu(U)$.*
5. *Pour toute partie mesurable A de $\mathfrak{X}$ telle que $\mu(\partial A) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.*

Preuve. L'implication 1. ⇒ 2. est évidente, et l'équivalence 3. ⇔ 4. également (par passage au complémentaire). Supposons 2., et fixons un fermé F de $\mathfrak{X}$. La fonction $d(x,F) = \inf_{y\in F} d(x,y)$ est 1-lipschitzienne sur $\mathfrak{X}$, donc pour tout $M > 0$,

$$f_M(x) = \max(0, 1 - M\,d(x,F))$$

est M-lipschitzienne, et par conséquent uniformément continue. On a donc par hypothèse :

$$\lim_{n\to\infty} \mu_n(f_M) = \mu(f_M).$$

Lorsque M tend vers l'infini, f_M tend ponctuellement vers la fonction indicatrice 1_F. Par convergence dominée, $\lim_{M\to\infty} \mu(f_M) = \mu(f)$, donc pour tout $\varepsilon > 0$, on peut trouver un M tel que $\mu(f_M) \leq \mu(F) + \varepsilon$. Alors,

$$\limsup_n \mu_n(F) \leq \limsup_n \mu_n(f_M) = \mu(f_M) \leq \mu(F) + \varepsilon.$$

Comme ceci est vrai pour tout $\varepsilon > 0$, $\limsup_n \mu_n(F) \leq \mu(F)$, et 2. $\Rightarrow$ 3.

Supposons maintenant 3. et 4., et montrons l'hypothèse 5. Soit A une partie telle que la frontière $\partial A = \overline{A} \setminus A^\circ$ ait mesure nulle, où $\overline{A}$ désigne la clôture de A (plus petit fermé contenant A), et A° désigne l'intérieur de A (plus grand ouvert inclus dans A). On a

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

donc tous ces termes sont égaux et $\lim_{n\to\infty} \mu_n(A) = \mu(\overline{A}) = \mu(A^\circ) = \mu(A)$. Finalement, supposons 5., et considérons une fonction f continue bornée sur $\mathfrak{X}$. Pour montrer que $\mu_n(f) \rightharpoonup \mu(f)$, on peut supposer sans perte de généralité que f est positive, et même à valeurs dans $[0,1]$. Alors,

$$\mu_n(f) = \int_{\mathfrak{X}} f(x)\,\mu_n(dx) = \int_{\mathfrak{X}} \left(\int_{y=0}^{1} 1_{y\leq f(x)}\,dy \right) \mu_n(dx)$$

Notons que la mesure $\mu_n(dx) \otimes dy$ est finie sur $\mathfrak{X} \times [0,1]$, et que la fonction $1_{y\leq f(x)}$ est mesurable positive sur cet espace. Par le théorème de Fubini–Tonelli, on peut échanger l'ordre d'intégration, donc

$$\mu_n(f) = \int_{y=0}^{1} \mu_n(\{x \,|\, f(x) \geq y\})\,dy.$$

Notons que l'application $\psi : y \mapsto \mu(\{x \,|\, f(x) \geq y\})$ est décroissante sur $[0,1]$, donc elle admet une quantité dénombrable de discontinuités. Soit y un point de continuité de cette fonction, et $A = \{x \in \mathfrak{X} \,|\, f(x) \geq y\} = f^{-1}([y,1])$. Comme A est l'image réciproque par une fonction continue d'un fermé, A est fermée. Si $x \in \partial A$, alors $x \notin A^\circ$, donc on peut trouver une suite $x_n \to x$ telle que $x_n \notin A$, c'est-à-dire telle que $f(x_n) < y$. Par passage à la limite, $f(x) \leq y$, donc $f(x) = y$. Ainsi, $\partial A \subset f^{-1}(\{y\})$, et l'hypothèse de continuité de ψ en y est équivalente au fait que $\mu(f^{-1}(\{y\})) = 0$. *A fortiori*, $\mu(\partial A) = 0$, donc $\lim_{n\to\infty} \mu_n(A) = \mu(A)$. On vient donc de montrer que la suite de fonctions

$$\psi_n(y) = \mu_n(f^{-1}([y,1]))$$

convergeait en presque tout point y de $[0,1]$ vers $\psi(y)$. D'autre part, toutes ces fonctions sont bornées par 1, donc, par convergence dominée,

$$\mu_n(f) = \int_0^1 \psi_n(y)\,dy \to \int_0^1 \psi(y)\,dy = \mu(f),$$

ce qui est l'hypothèse 1. $\square$

Corollaire 1.17. *Si $X_n \to_{\mathbb{P}} X$ dans un espace métrique (ou métrisable), alors $X_n \rightharpoonup X$.*

Preuve. Soit f une fonction bornée et uniformément continue sur $(\mathfrak{X}, d)$, et $\varepsilon > 0$. On peut trouver $\delta > 0$ tel que, si $d(x,y) \leq \delta$, alors $|f(x) - f(y)| \leq \varepsilon$ (on parlera de module d'ε-continuité pour f). Fixons également une borne M sur f, et soit n_0 assez grand, tel que $\mathbb{P}[d(X_n,X) \geq \delta] \leq \frac{\varepsilon}{M}$ pour tout $n \geq n_0$. On a alors, pour tout $n \geq n_0$,

$$\begin{aligned}\mathbb{E}[|f(X_n) - f(X)|] &\leq \mathbb{E}\left[|f(X_n) - f(X)|\, 1_{d(X_n,X)\leq\delta}\right] + \mathbb{E}\left[|f(X_n) - f(X)|\, 1_{d(X_n,X)\geq\delta}\right] \\ &\leq \mathbb{E}\left[\varepsilon\, 1_{d(X_n,X)\leq\delta}\right] + 2M\, \mathbb{P}[d(X_n,X) \geq \delta] \leq 3\varepsilon. \end{aligned}$$

□

Remarque 1.18. Il n'y a pas de réciproque à ce résultat, sauf dans le cas particulier où la limite X est une constante $x \in \mathfrak{X}$. Dans ce cas, $X_n \to_{\mathbb{P}} x$ si et seulement $X_n \rightharpoonup x$. En effet, si $X_n \rightharpoonup x$, fixons $\varepsilon > 0$, et considérons les deux boules $B_{(x,\varepsilon)} = \{y \in \mathfrak{X} \,|\, d(x,y) < \varepsilon\}$ et $\overline{B}_{(x,\varepsilon)} = \{y \in \mathfrak{X} \,|\, d(x,y) \leq \varepsilon\}$. La première est une partie ouverte, et la seconde est une partie fermée de $\mathfrak{X}$. Par le théorème de Portmanteau,

$$\limsup_n \mathbb{P}[d(X_n, x) \leq \varepsilon] \leq \delta_x(\overline{B}_{(x,\varepsilon)}) = 1 = \delta_x(B_{(x,\varepsilon)}) \leq \liminf_n \mathbb{P}[d(X_n, x) < \varepsilon],$$

donc toutes ces quantités sont égales à 1, et $\lim_{n\to\infty} \mathbb{P}[d(X_n, x) \leq \varepsilon] = 1$. Ainsi, $X_n \to_{\mathbb{P}} x$.

▷ Convergence de variables aléatoires et fonctions continues. Pour conclure notre présentation des divers modes de convergence pour des variables aléatoires, voyons comment ces notions se comportent vis-à-vis des transformations continues. On a :

Proposition 1.19. *Soit $\mathfrak{X}$ et $\mathfrak{Y}$ deux espaces topologiques, et $f : \mathfrak{X} \to \mathfrak{Y}$ une fonction continue. Si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires à valeurs dans $\mathfrak{X}$, alors*

$$\big(X_n \to_{\text{p.s.}} X\big) \Rightarrow \big(f(X_n) \to_{\text{p.s.}} f(X)\big)$$

et

$$\big(X_n \rightharpoonup X\big) \Rightarrow \big(f(X_n) \rightharpoonup f(X)\big).$$

Si $\mathfrak{X}$ et $\mathfrak{Y}$ sont des espaces topologiques métrisables, alors

$$\big(X_n \to_{\mathbb{P}} X\big) \Rightarrow \big(f(X_n) \to_{\mathbb{P}} f(X)\big).$$

Preuve. La première implication est évidente, car si $X_n(\omega) \to X(\omega)$ avec probabilité 1, alors $f(X_n(\omega)) \to f(X(\omega))$ avec la même probabilité, simplement par continuité de f. Pour la seconde implication, si $X_n \rightharpoonup X$ et g est une fonction continue bornée sur $\mathfrak{Y}$, alors $g \circ f$ est une fonction continue bornée sur $\mathfrak{X}$, donc $\mathbb{E}[g(f(X_n))] \to \mathbb{E}[g(f(X))]$; ainsi, $f(X_n) \rightharpoonup f(X)$. Finalement, la troisième implication découle de la première compte tenu du lien entre convergence en loi et convergence en probabilité (sous-suites). □

1.2 Topologie de l'espace des mesures de probabilité

Un résultat topologique important pour construire des objets en analyse est le lemme suivant :

Lemme 1.20. *Soit $\mathfrak{X}$ un espace topologique métrisable, d une distance qui métrise la topologie. On se donne une suite $(x_n)_{n\in\mathbb{N}}$ à valeurs dans $\mathfrak{X}$, qui a les deux propriétés suivantes :*

1. *Toute sous-suite de $(x_n)_{n\in\mathbb{N}}$ a une sous-sous-suite qui converge.*
2. *Si $(x_{n_k})_{k\in\mathbb{N}}$ est une sous-suite convergente de $(x_n)_{n\in\mathbb{N}}$, alors sa seule limite possible est x.*

Dans ce cas, $x_n \to x$.

Preuve. Supposons $x_n \not\to x$. Alors, on peut trouver $\varepsilon > 0$ et une sous-suite $(x_{n_k})_{k\in\mathbb{N}}$ telle que, pour tout k, $d(x_{n_k}, x) \geq \varepsilon$. On peut ensuite trouver une sous-sous-suite de $(x_{n_k})_{k\in\mathbb{N}}$, notée $(x_{n_{k_l}})_{l\in\mathbb{N}}$, qui admet une limite y. Par passage à la limite, $d(x, y) \geq \varepsilon$. Mais ceci contredit le fait que toute sous-suite convergente de $(x_n)_{n\in\mathbb{N}}$ a pour limite x. □

En théorie des probabilités, ce lemme peut être utilisé de la façon suivante. Supposons que l'on veuille construire un objet aléatoire X dans un espace $\mathfrak{X}$, dont la loi μ vérifie certaines propriétés d'une liste P. Lorsque $\mathfrak{X}$ est un espace un tant soit peu compliqué (l'espace des fonctions continues sur $\mathbb{R}_+$; l'espace de tous les arbres munis d'une métrique sur leurs arêtes ; *etc.*), il n'est pas forcément évident qu'une telle loi μ sur $\mathfrak{X}$ existe. Une approche possible pour la construction de μ est alors la suivante. On peut approximer X par des variables X_n plus simples, et telles que :

1. Les lois μ_n des variables X_n restent dans une partie compacte pour la topologie de la convergence en loi de l'ensemble $\mathscr{M}^1(\mathfrak{X})$ des mesures de probabilité sur $\mathfrak{X}$.
2. Les propriétés listées dans P pour la loi μ la caractérisent, et si $(\mu_{n_k})_{k\in\mathbb{N}}$ est une sous-suite de lois qui admet une limite, alors cette limite a les propriétés listées dans P.

Dans ce cas, par le lemme précédent, la mesure μ existe bien, et c'est la limite en loi de $(\mu_n)_{n\in\mathbb{N}}$. On verra dans le chapitre 3 que le mouvement brownien peut être construit de cette façon. Cette approche invite à étudier la topologie de la convergence en loi sur $\mathscr{M}^1(\mathfrak{X})$, et à en identifier les parties relativement compactes ; c'est l'objectif de cette section et de la suivante.

▷ Mesures de probabilité sur un espace polonais. Les hypothèses dont on aura besoin pour l'espace des variables $\mathfrak{X}$ sont réunies dans la définition suivante :

Définition 1.21 (Espace polonais). *Soit $\mathfrak{X}$ un espace topologique. On dit que $\mathfrak{X}$ est un espace polonais si :*

1. *Il existe une distance d qui métrise la topologie de $\mathfrak{X}$: U est un ouvert de $\mathfrak{X}$ si et seulement si, pour tout $x \in U$, il existe $\varepsilon > 0$ tel la boule $B_{(x,\varepsilon)} = \{y \mid d(x, y) < \varepsilon\}$ soit incluse dans U.*
2. *L'espace métrique $(\mathfrak{X}, d)$ est complet : si $(x_n)_{n\in\mathbb{N}}$ est une suite de Cauchy pour d dans $\mathfrak{X}$, alors $(x_n)_{n\in\mathbb{N}}$ a une limite $x \in \mathfrak{X}$.*

3. *L'espace $\mathfrak{X}$ est séparable : il existe une suite $(x_n)_{n\in\mathbb{N}}$ qui est dense dans $\mathfrak{X}$.*

Exemple 1.22. Les ensembles $\mathbb{Z}$ et $\mathbb{R}$ munis de leurs topologies usuelles sont des espaces polonais.

Remarque 1.23. On peut légitimement se demander pourquoi un espace polonais est défini comme un espace séparable *métrisable* par une distance complète, et non directement comme un espace séparable *métrique* complet. L'intérêt de cette définition est le suivant. Supposons que $\mathfrak{X}$ soit une partie d'un espace vectoriel V de dimension infinie. Dans ce cas, la topologie de V est souvent définie par une famille d'observables $\mathscr{O} = \{f : V \to \mathbb{R}\}$: $v_n \to v$ dans V si et seulement si $f(v_n) \to f(v)$ pour tout $f \in \mathscr{O}$. Même lorsque cette topologie (restreinte à $\mathfrak{X}$) est métrisable, il est alors plus pratique de travailler avec les observables $f \in \mathscr{O}$, qu'avec une distance d, qui peut ne pas être très naturelle ou commode à manipuler (elle est rarement canonique). Le cas typique de cette situation est celui de la convergence en loi dans $\mathscr{M}^1(\mathfrak{X})$. On verra dans un instant que si $\mathfrak{X}$ est un espace polonais, alors $\mathscr{M}^1(\mathfrak{X})$ est un espace topologique métrisable (pour la topologie de la convergence en loi). Néanmoins, la manipulation de distances sur cet espace n'est pas très aisée, et la topologie de $\mathscr{M}^1(\mathfrak{X})$ est plus naturellement contrôlée par la famille d'observables $\mathscr{O} = \mathscr{C}_b(\mathfrak{X}, \mathbb{R})$, l'espace des fonctions continues bornées sur $\mathfrak{X}$.

Remarque 1.24. Une autre raison qui justifie qu'on travaille avec des espaces métrisables et non des espaces métriques est la suivante. Si $(\mathfrak{X}, d)$ est un espace métrique non complet, il peut exister une autre distance d' sur $\mathfrak{X}$, qui correspond à la même topologie, mais qui le rend complet. Par exemple, $\mathfrak{X} = [0,1)$ muni de la distance usuelle sur les réels n'est pas complet (ce n'est pas une partie fermée de $\mathbb{R}$), mais comme $[0,1)$ est homéomorphe à $\mathbb{R}_+$ qui est complet, $[0,1)$ est bien un espace topologique polonais. Ainsi, le cadre des espaces polonais offre plus de liberté que le cadre des espaces métriques complets séparables.

Dans toute la suite de ce chapitre, on fixe un espace topologique métrisable et séparable $\mathfrak{X}$, et lorsqu'on en aura besoin, on utilisera une distance d sur $\mathfrak{X}$ qui métrise la topologie. On note $\mathscr{M}^1(\mathfrak{X})$ l'ensemble des mesures de probabilités sur l'espace mesurable $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}))$. Cet espace est muni de l'unique topologie telle qu'une base de voisinages d'une mesure μ est constituée des parties

$$B_{(\mu,\varepsilon,f_1,\ldots,f_n)} = \{\nu \in \mathscr{M}^1(\mathfrak{X}) \mid \forall i \in [\![1,n]\!],\ |\mu(f_i) - \nu(f_i)| < \varepsilon\}$$

avec $\varepsilon > 0$ et $f_1, \ldots, f_n \in \mathscr{C}_b(\mathfrak{X})$ fonctions continues bornées. Alors, la convergence dans $\mathscr{M}^1(\mathfrak{X})$ pour cette topologie est exactement la convergence en loi précédemment introduite. De plus, la proposition 1.19 peut être réinterprétée comme suit : si $\mathfrak{X}$ et $\mathfrak{Y}$ sont deux espaces topologiques et si $f : \mathfrak{X} \to \mathfrak{Y}$ est une application continue, alors elle induit une application

$$\begin{aligned} f_* : \mathscr{M}^1(\mathfrak{X}) &\to \mathscr{M}^1(\mathfrak{Y}) \\ \mu &\mapsto f_*\mu = \mu \circ f^{-1} \end{aligned}$$

qui est continue.

Proposition 1.25. *Soit $\mathfrak{X}$ un espace topologique métrisable et séparable. L'espace topologique $\mathscr{M}^1(\mathfrak{X})$ est métrisable et séparable.*

Preuve. On fixe une distance d sur $\mathfrak{X}$, et on introduit la distance de Lévy–Prohorov sur $\mathscr{M}^1(\mathfrak{X})$:

$$d_{\mathrm{LP}}(\mu,\nu)=\inf\{\varepsilon>0\,|\,\forall A\in\mathscr{B}(\mathfrak{X}),\ \mu(A)\leq \nu(A^\varepsilon)+\varepsilon \text{ et } \nu(A)\leq\mu(A^\varepsilon)+\varepsilon\},$$

où $A^\varepsilon=\{x\in\mathfrak{X}\,|\,\exists y\in A, d(x,y)<\varepsilon\}$. Montrons d'abord que d_{LP} est une distance sur $\mathscr{M}^1(\mathfrak{X})$. Elle est clairement symétrique, et comme $(A^\varepsilon)^\eta\subset A^{\varepsilon+\eta}$ pour toute partie A, on obtient aisément l'inégalité triangulaire. Montrons que si $d_{\mathrm{LP}}(\mu,\nu)=0$, alors $\mu=\nu$. Soit A une partie fermée de $\mathfrak{X}$; alors, $A=\bigcap_{\varepsilon>0}A^\varepsilon$. Comme $\mu(A)\leq\nu(A^\varepsilon)+\varepsilon$ pour tout $\varepsilon>0$, par passage à la limite, $\mu(A)\leq\nu(A)$, et symétriquement, $\nu(A)\leq\mu(A)$, donc $\mu(A)=\nu(A)$ pour toute partie fermée A. Il est bien connu que ceci implique $\mu=\nu$ dans l'espace des mesures de probabilité (voir les exercices en fin de chapitre). Ainsi, d_{LP} est bien une distance sur $\mathscr{M}^1(\mathfrak{X})$. Notons que l'on a en fait :

$$d_{\mathrm{LP}}(\mu,\nu)=\inf\{\varepsilon>0\,|\,\forall A\in\mathscr{B}(\mathfrak{X}),\ \mu(A)\leq\nu(A^\varepsilon)+\varepsilon\}.$$

En effet, soit $\varepsilon>0$ tel que $\mu(A)\leq\nu(A^\varepsilon)+\varepsilon$. Notons que

$$((A^\varepsilon)^{\mathrm{c}})^\varepsilon\subset A^{\mathrm{c}}$$

pour toute partie A : si x appartient au membre de gauche, alors il est à distance strictement inférieure à ε d'éléments y à distance plus grande que ε de A, donc, par l'inégalité triangulaire, x ne peut pas être dans A. Il s'ensuit :

$$\nu(A)=1-\nu(A^{\mathrm{c}})\leq 1-\nu(((A^\varepsilon)^{\mathrm{c}})^\varepsilon)\leq 1-\mu((A^\varepsilon)^{\mathrm{c}})+\varepsilon=\mu(A^\varepsilon)+\varepsilon,$$

c'est-à-dire l'autre inégalité.

Vérifions maintenant que d_{LP} métrise la topologie de $\mathscr{M}^1(\mathfrak{X})$. Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite de mesures de probabilité, telle que $d_{\mathrm{LP}}(\mu_n,\mu)\to 0$. Soit A une partie fermée de $\mathfrak{X}$, et $\varepsilon>0$. Il existe un rang n_0 tel que, pour tout $n\geq n_0$, $\mu_n(A)\leq\mu(A^\varepsilon)+\varepsilon$. Par conséquent,

$$\limsup_n \mu_n(A)\leq\mu(A^\varepsilon)+\varepsilon.$$

En faisant tendre ε vers 0, on conclut que $\limsup_n\mu_n(A)\leq\mu(A)$, car $A=\bigcap_{\varepsilon>0}A^\varepsilon$ en tant que partie fermée. Donc, par le théorème de Portmanteau, $\mu_n\rightharpoonup\mu$. Réciproquement, supposons $\mu_n\rightharpoonup\mu$, et fixons une suite $(x_m)_{m\in\mathbb{N}}$ dense dans $\mathfrak{X}$. Pour tout $\varepsilon>0$, $\mathfrak{X}=\bigcup_{m\in\mathbb{N}}B_{(x_m,\varepsilon)}$. Fixons $\varepsilon>0$, et une partie finie $M\subset\mathbb{N}$ telle que

$$\mu\left(\bigcup_{m\in M}B_{(x_m,\varepsilon)}\right)\geq 1-\varepsilon.$$

L'ensemble $\mathscr{U}$ de tous les ouverts U de la forme

$$U=B_{(x_{m_1},\varepsilon)}\cup B_{(x_{m_2},\varepsilon)}\cup\cdots\cup B_{(x_{m_r},\varepsilon)},\ \ \{m_1,\ldots,m_r\}\subset M$$

est fini, donc, par le théorème de Portmanteau, pour n assez grand, $\mu_n(U)\geq\mu(U)-\varepsilon$ pour tout ouvert $U\in\mathscr{U}$. Fixons maintenant une partie mesurable $A\in\mathscr{B}(\mathfrak{X})$. On approxime A par

$$U(A)=\bigcup_{m\in M\,|\,B_{(x_m,\varepsilon)}\cap A\neq\emptyset}B_{(x_m,\varepsilon)}.$$

Si $x \in A \setminus U(A)$, alors x n'est dans aucune boule $B_{(x_m,\varepsilon)}$, $m \in M$, donc x est dans une partie de μ-mesure inférieure à ε. Alors,

$$\mu(A) \leq \mu(U(A)) + \mu(A \setminus U(A)) \leq \mu_n(U(A)) + 2\varepsilon.$$

Or, $U(A) \subset A^{2\varepsilon}$, donc :

$$\mu(A) \leq \mu_n(A^{2\varepsilon}) + 2\varepsilon.$$

Par la caractérisation non symétrique de la distance de Lévy–Prohorov, $d_{\mathrm{LP}}(\mu_n, \mu) \leq 2\varepsilon$. Ainsi, d_{LP} est bien une distance qui métrise la topologie de la convergence en loi.

Finalement, pour établir la séparabilité de $\mathscr{M}^1(\mathfrak{X})$, considérons de nouveau une suite dense $(x_m)_{m\in\mathbb{N}}$ dans $\mathfrak{X}$, et l'ensemble dénombrable des mesures de probabilité de la forme

$$\nu = \sum_{m\in M} r_m\, \delta_{x_m},$$

où M est une partie finie de $\mathbb{N}$, et les r_m sont des nombres rationnels tels que $\sum_{m\in M} r_m = 1$. On veut approximer à $\varepsilon > 0$ près pour la métrique de Lévy–Prohorov une mesure $\mu \in \mathscr{M}^1(\mathfrak{X})$ par une mesure ν de cette forme. On fixe une partie finie M telle que

$$\mu\left(\bigcup_{m\in M} B_{(x_m,\varepsilon)}\right) \geq 1 - \varepsilon.$$

En remplaçant les boules $B_{(x_m,\varepsilon)}$ par des intersections de celles-ci, on peut trouver des parties mesurables *disjointes* B_m telles que $x_m \in B_m \subset B_{(x_m,\varepsilon)}$ pour tout $m \in M$, et telles que

$$\bigcup_{m\in M} B_{(x_m,\varepsilon)} = \bigsqcup_{m\in M} B_m.$$

On pose $\nu = \sum_{m\in M} r_m \delta_{x_m}$, où les r_m sont des rationnels choisis tels que $\sum_{m\in M} |r_m - \mu(B_m)| \leq \varepsilon$. Soit A une partie mesurable de $\mathfrak{X}$, et

$$V(A) = \bigsqcup_{m\in M \,|\, B_m \cap A \neq \emptyset} B_m.$$

On a alors

$$\mu(A) \leq \mu(V(A)) + \mu(A \setminus V(A)) \leq \mu(V(A)) + \varepsilon,$$

et d'autre part, $V(A) \subset A^{\varepsilon}$ et $|\mu(V(A)) - \nu(V(A))| \leq \sum_{m\in M \,|\, B_m\cap A\neq\emptyset} |\mu(B_m) - r_m| \leq \varepsilon$, donc

$$\mu(A) \leq \nu(V(A)) + 2\varepsilon \leq \nu(A^{2\varepsilon}) + 2\varepsilon \qquad ; \qquad d_{\mathrm{LP}}(\mu, \nu) \leq 2\varepsilon.$$

Ainsi, on a bien trouvé une partie dénombrable dense dans $\mathscr{M}^1(\mathfrak{X})$. □

▷ **Le cas des espaces compacts.** On verra plus loin un résultat encore meilleur : si $\mathfrak{X}$ est un espace topologique polonais, alors $\mathscr{M}^1(\mathfrak{X})$ muni de la topologie de la convergence en loi est aussi un espace topologique polonais. Ainsi, toute suite de Cauchy pour la distance de Lévy–Prohorov dans $\mathscr{M}^1(\mathfrak{X})$ sera convergente. Pour établir ce résultat, on aura besoin de montrer qu'une suite de Cauchy dans $\mathscr{M}^1(\mathfrak{X})$ est relativement compacte ; la section 1.3 déterminera justement quelles sont les parties relativement compactes de $\mathscr{M}^1(\mathfrak{X})$. Un résultat plus simple, et qui est une première étape vers le résultat général, est le suivant :

Théorème 1.26. *Soit $\mathfrak{X}$ un espace topologique métrisable. L'espace des mesures de probabilité $\mathscr{M}^1(\mathfrak{X})$ muni de la topologie de la convergence en loi est compact si et seulement si $\mathfrak{X}$ est compact. Dans ce cas, $\mathscr{M}^1(\mathfrak{X})$ est en fait métrisable compact.*

Rappelons qu'un espace topologique $\mathfrak{X}$ est dit compact s'il est de Hausdorff (deux points $x \neq y$ peuvent être séparés par deux ouverts U et V tels que $x \in U$, $y \in V$ et $U \cap V = \emptyset$) et si, de tout recouvrement $\mathfrak{X} = \bigcup_{m \in N} U_m$ de $\mathfrak{X}$ par des ouverts U_m, on peut extraire un recouvrement fini :

$$\mathfrak{X} = \bigcup_{m \in M} U_m \quad \text{avec } M \subset N \text{ et } M \text{ fini.}$$

Lorsque $\mathfrak{X}$ est métrisable, cette condition est équivalente à la compacité séquentielle : toute suite $(x_n)_{n \in N}$ dans X a une sous-suite $(x_{n_k})_{k \in \mathbb{N}}$ qui est convergente.

Le théorème 1.26 est une conséquence de deux résultats topologiques classiques, que nous énonçons ici sans preuve (voir les références en fin de chapitre). Soit $\mathfrak{X}$ un espace topologique compact ; notons $\mathscr{C}(\mathfrak{X}, \mathbb{R}) = \mathscr{C}(\mathfrak{X})$ l'ensemble des fonctions continues sur $\mathfrak{X}$. C'est un espace de Banach pour la norme $\|f\|_\infty = \max_{x \in \mathfrak{X}} |f(x)|$ (le maximum est atteint par compacité de $\mathfrak{X}$).

Théorème 1.27 (Riesz)**.** *Une forme linéaire continue ϕ sur $\mathscr{C}(\mathfrak{X})$ avec $\mathfrak{X}$ espace topologique compact s'écrit de manière unique sous la forme*

$$\phi(f) = \int_{\mathfrak{X}} f(x)\, \mu(dx),$$

où $\mu = \mu_+ - \mu_-$ est une mesure borélienne sur $\mathfrak{X}$, signée et de masse totale $|\mu| = \mu_+(\mathfrak{X}) + \mu_-(\mathfrak{X})$ finie. Cette masse totale est aussi la norme d'opérateur de ϕ.

On note $\mathscr{M}(\mathfrak{X})$ l'espace vectoriel des mesures signées finies sur $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}))$; il contient l'ensemble $\mathscr{M}^1(\mathfrak{X})$ des mesures de probabilité, et d'après le résultat précédent, c'est le dual topologique de $\mathscr{C}(\mathfrak{X})$. À ce titre, il hérite de la topologie $*$-faible, qui est la topologie la moins fine pour laquelle les applications

$$\phi \mapsto \phi(f)$$

avec $f \in \mathscr{C}(\mathfrak{X})$ sont continues. La restriction à $\mathscr{M}^1(\mathfrak{X})$ de la topologie $*$-faible sur $\mathscr{M}(\mathfrak{X})$ est par définition la topologie de la convergence en loi.

Théorème 1.28 (Banach–Alaoglu). *Soit V un espace vectoriel normé, et V^* son dual topologique (formes linéaires continues sur V), muni de la norme*

$$\|\phi\| = \sup_{v\in V,\ \|v\|\leq 1} |\phi(v)|.$$

La boule unité fermée $\overline{B}^*_{(0,1)}$ de V^* est compacte pour la topologie $*$-faible.

Appliquons ce résultat à $V = \mathscr{C}(\mathfrak{X})$, avec $\mathfrak{X}$ espace topologique compact. Par le théorème de Banach–Alaoglu appliqué à $V^* = \mathscr{M}(X)$, la boule unité $\overline{B}^*$ de $\mathscr{M}(\mathfrak{X})$ est compacte, et $\mathscr{M}^1(\mathfrak{X})$ est une partie de cette boule unité, car si $\|f\|_\infty \leq 1$ et $\mu \in \mathscr{M}^1(\mathfrak{X})$, alors $|\mu(f)| \leq 1$. Pour montrer que $\mathscr{M}^1(\mathfrak{X})$ est compacte, il suffit donc de montrer que c'est une partie fermée dans $\mathscr{M}(\mathfrak{X})$. Or, c'est l'intersection des parties

$$\{\mu \in \mathscr{M}(\mathfrak{X}) \mid \mu(1) = 1\}$$

et

$$\{\mu \in \mathscr{M}(\mathfrak{X}) \mid \mu(f) \geq 0\}$$

pour $f \in \mathscr{C}(\mathfrak{X})$ positive. Ces parties sont fermées pour la topologie $*$-faible, en tant qu'images réciproques par des applications $*$-faiblement continues de fermés de $\mathbb{R}$. Donc, $\mathscr{M}^1(\mathfrak{X})$ est fermée dans $\mathscr{M}(\mathfrak{X})$, et si $\mathfrak{X}$ est un espace topologique compact, alors $\mathscr{M}^1(\mathfrak{X})$ est également compact. Si de plus $\mathfrak{X}$ est métrisable, alors il est automatiquement séparable, et on a vu dans ce cas que $\mathscr{M}^1(\mathfrak{X})$ était aussi métrisable séparable, et donc métrisable compact.

Preuve du théorème 1.26. On vient de voir l'une des implications du théorème. Réciproquement, supposons $\mathscr{M}^1(\mathfrak{X})$ compact, et considérons l'application

$$\begin{aligned} \delta : \mathfrak{X} &\to \mathscr{M}^1(\mathfrak{X}) \\ x &\mapsto \delta_x. \end{aligned}$$

Soit U un ouvert de $\mathscr{M}^1(\mathfrak{X})$, et $V = \delta^{-1}(U)$; montrons que V est ouvert. Si $x \in V$, alors $\delta_x \in U$, donc comme U est ouvert, il existe $\varepsilon > 0$ et des fonctions continues $f_1, \ldots, f_n$ telle que $B_{(\delta_x,\varepsilon,f_1,\ldots,f_n)} \subset U$. Soit y tel que

$$\forall i \in [\![1,n]\!],\ |f_i(x) - f_i(y)| < \varepsilon.$$

Alors, $\delta_y \in B_{(\delta_x,\varepsilon,f_1,\ldots,f_n)} \subset U$, et d'autre part, comme les fonctions f_i sont continues, il existe un voisinage W de x dans $\mathfrak{X}$ tel que si $y \in W$, alors la condition précédente est vérifiée. On conclut que V contient un voisinage W de x, et donc que V est ouvert. Ainsi, l'application δ est continue.

Il est clair que δ est injective, et on va montrer que δ est en fait un homéomorphisme de $\mathfrak{X}$ vers une partie fermée et donc compacte de $\mathscr{M}^1(\mathfrak{X})$. L'image de δ, c'est-à-dire l'ensemble des mesures de Dirac, est caractérisée par la propriété suivante : $\mu \in \delta(\mathfrak{X})$ si et seulement si, pour toute fonction continue f sur $\mathfrak{X}$,

$$\mathrm{var}_\mu(f) = \mu(f^2) - (\mu(f))^2 = 0.$$

Comme $\mu \mapsto \mu(f^2) - (\mu(f))^2$ est continue sur $\mathscr{M}^1(\mathfrak{X})$, $\delta(\mathfrak{X})$ est donc une intersection d'images réciproques par des fonctions continues de fermés de $\mathbb{R}$. C'est donc bien une partie fermée de $\mathscr{M}^1(\mathfrak{X})$, et par conséquent compacte.

Finalement, pour montrer que δ est un homéomorphisme, il reste à montrer que l'image d'un fermé $F \subset \mathfrak{X}$ par δ est un fermé de $\mathscr{M}^1(\mathfrak{X})$. On a besoin ici de l'hypothèse $\mathfrak{X}$ métrisable, et de la caractérisation séquentielle du caractère fermé. Soit $(\delta_{x_n})_{n\in\mathbb{N}}$ une suite de mesures de Dirac qui converge $*$-faiblement dans $\mathscr{M}^1(\mathfrak{X})$ vers δ_x, avec tous les x_n dans F. On doit montrer que $x \in F$, ce qui impliquera que $\delta(F)$ est une partie fermée. Or, la fonction

$$d_F : x \mapsto \min(d(x,F),1)$$

est continue bornée et même 1-lipschitzienne sur $\mathfrak{X}$, donc $\lim_{n\to\infty} \delta_{x_n}(d_F) = \delta_x(d_F) = d_F(x)$. Comme $\delta_{x_n}(d_F) = d_F(x_n) = 0$ pour tout n, on conclut que $d_F(x) = 0$, et donc que $x \in F$. Ainsi, δ est un homéomorphisme de $\mathfrak{X}$ vers un compact, donc $\mathfrak{X}$ est métrisable compact. □

▷ Le dual topologique de $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$. Considérons maintenant un espace topologique polonais $\mathfrak{X}$. On note $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ l'espace des fonctions continues bornées sur $\mathfrak{X}$. Muni de la norme $\|f\|_\infty = \sup_{x\in\mathfrak{X}} |f(x)|$, c'est de nouveau un espace de Banach. Par définition, une suite de mesures de probabilité μ_n converge en loi vers μ si et seulement si, pour tout $f \in \mathscr{C}_{\mathrm{b}}(\mathfrak{X})$, $\mu_n(f) \to \mu(f)$. De plus, pour toute mesure de probabilité μ, l'application $f \mapsto \mu(f)$ est une forme linéaire continue sur $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$. On peut donc réinterpréter la convergence en loi de mesures de probabilité comme suit :

1. L'espace $\mathscr{M}^1(\mathfrak{X})$ est une partie du dual topologique $(\mathscr{C}_{\mathrm{b}}(\mathfrak{X}))^*$, et plus précisément, une partie de sa boule unité.
2. La topologie de la convergence en loi sur $\mathscr{M}^1(\mathfrak{X})$ est la restriction à cet espace de la topologie $*$-faible sur $(\mathscr{C}_{\mathrm{b}}(\mathfrak{X}))^*$.

La différence essentielle avec la discussion du cas compact est la suivante : dans le cas où $\mathfrak{X}$ n'est pas compact, $\mathscr{M}^1(\mathfrak{X})$ n'est pas une partie $*$-faiblement fermée du dual topologique de $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$, et par conséquent, même si le théorème de Banach–Alaoglu s'applique pour $(\mathscr{C}_{\mathrm{b}}(\mathfrak{X}))^*$, il ne dit rien sur la topologie de $\mathscr{M}^1(\mathfrak{X})$.

Le caractère non $*$-faiblement fermé de $\mathscr{M}^1(\mathfrak{X})$ dans $(\mathscr{C}_{\mathrm{b}}(\mathfrak{X}))^*$ est équivalent au résultat suivant : si $\mathfrak{X}$ espace métrisable séparable n'est pas compact, alors le dual topologique de $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ contient strictement plus de formes linéaires que celles induites par $\mathscr{M}(\mathfrak{X})$, l'espace des mesures signées finies. Décrivons plus précisément ce dual topologique. Une mesure finiment additive sur $\mathfrak{X}$ est une fonction $\mu : \mathscr{B}(\mathfrak{X}) \to \mathbb{R}$ telle que :

1. pour toute famille finie $(A_1,\ldots,A_n)$ de parties mesurables disjointes,

$$\mu\left(\bigsqcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i).$$

2. pour toute partie mesurable A,

$$|\mu|(A) = \sup_{A=B_1\sqcup B_2\sqcup\cdots\sqcup B_n} \left(\sum_{i=1}^n |\mu(B_i)|\right) < +\infty.$$

Par rapport à la définition usuelle de mesure signée, on demande l'additivité de μ sur les unions disjointes finies, mais pas forcément sur les unions disjointes dénombrables. L'ensemble $\mathscr{M}_{\mathrm{fa}}(\mathfrak{X})$ des mesures finiment additives est donc en général strictement plus grand que $\mathscr{M}(\mathfrak{X})$. On équipe $\mathscr{M}_{\mathrm{fa}}(\mathfrak{X})$ de la norme $\|\mu\| = |\mu|(\mathfrak{X})$; c'est un espace de Banach pour cette norme. Un sous-espace vectoriel fermé de $\mathscr{M}_{\mathrm{fa}}(\mathfrak{X})$ est l'espace des mesures finiment additives régulières, *i.e.*, telles qu'en plus des deux assertions précédentes, on a :

3. pour toute partie mesurable A, et pour tout $\varepsilon > 0$, il existe un ouvert U contenant A et un fermé F contenu dans A, tel que $|\mu|(U \setminus F) \leq \varepsilon$.

On note $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$ l'ensemble des mesures finiment additives régulières sur $\mathfrak{X}$. Si $\mu \in \mathscr{M}_{\mathrm{far}}(\mathfrak{X})$ et si $f = \sum_{i=1}^{n} 1_{A_i}\, x_i$ est une fonction en escalier sur $\mathfrak{X}$ et à valeurs réelles, alors on peut définir l'intégrale de f par rapport à la mesure finiment additive μ : elle est donnée par la formule

$$\mu(f) = \sum_{i=1}^{n} \mu(A_i)\, x_i \quad \text{si les } A_i \text{ sont disjoints.}$$

La condition d'additivité finie implique que cette définition ne dépend pas de la décomposition de f comme combinaison linéaire de fonctions indicatrices de parties mesurables. On note $\mathrm{Esc}(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), |\mu|)$ l'espace vectoriel des fonctions en escalier sur $\mathfrak{X}$, muni de la norme

$$\|f\|_{\mathscr{L}^1} = \sum_{i=1}^{n} |\mu|(A_i)\, |x_i|$$

si $f = \sum_{i=1}^{n} 1_{A_i}\, x_i$. On définit alors comme dans le cas classique $\mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), |\mu|)$ comme le complété de $\mathrm{Esc}(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), |\mu|)$ pour la norme $\mathscr{L}^1$; c'est un espace de Banach, sur lequel l'application $f \mapsto \mu(f)$ est une forme linéaire continue. Ainsi, on a un analogue de la théorie de l'intégration de Lebesgue par rapport aux mesures finiment additives. On peut montrer que si $\mathfrak{X}$ est polonais et μ est finiment additive régulière, alors $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ est inclus dans $\mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), |\mu|)$.

Théorème 1.29 (Dual de $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$). *Soit $\mathfrak{X}$ un espace topologique polonais. Le dual topologique de $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ est l'ensemble $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$ des mesures finiment additives régulières sur $\mathfrak{X}$. Si $\mu \in \mathscr{M}_{\mathrm{far}}(\mathfrak{X})$, alors sa norme d'opérateur en tant que forme linéaire sur $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ est $|\mu|(\mathfrak{X})$. De plus, $\mathscr{M}(\mathfrak{X})$ est un sous-espace fermé de $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$ pour la topologie forte associée à cette norme d'opérateurs.*

Nous admettrons ce résultat, qui est assez difficile. Un point important est que $\mathscr{M}^1(\mathfrak{X})$ est une partie fortement fermée de $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$, mais n'est pas une partie $*$-faiblement fermée de $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$ lorsque $\mathfrak{X}$ n'est pas compact. Par le théorème de Banach–Alaoglu, $\mathscr{M}^1(\mathfrak{X})$ est bien inclus dans une partie compacte du dual topologique de $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$, mais les limites potentielles de suites de mesures de probabilité pour la topologie $*$-faible peuvent être des mesures seulement finiment additives, et donc en dehors de $\mathscr{M}^1(\mathfrak{X})$.

Remarque 1.30. On peut montrer (voir les exercices) que si $\mathfrak{X}$ est un espace topologique métrisable, alors $\mathfrak{X}$ est compact si et seulement si $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ est séparable. On peut également montrer qu'un espace de Banach est séparable si et seulement si la boule unité de son dual topologique est métrisable pour la topologie $*$-faible. Par conséquent, lorsque $\mathfrak{X}$ est polonais mais non compact, la boule unité de $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$, qui est munie de la topologie $*$-faible qui la rend compacte, n'est en revanche pas métrisable. En particulier, la compacité de cette boule n'est pas caractérisée séquentiellement. Dans la phrase juste avant le début de cette remarque, les "limites potentielles de suites de mesures de probabilité" doivent donc être entendues au sens des suites généralisées à la Moore–Smith (familles indexées par des ensembles ordonnés filtrants).

1.3 Théorème de Prohorov

D'après la discussion de la section précédente, dans le cas général d'un espace topologique polonais $\mathfrak{X}$ non compact, $\mathscr{M}^1(\mathfrak{X})$ n'est pas non plus compacte pour la topologie de la convergence en loi. Pour employer la méthode d'approximation de mesures de probabilité décrite au début de la section 1.2, il faut donc identifier les parties relativement compactes de $\mathscr{M}^1(\mathfrak{X})$, c'est-à-dire les parties dont la clôture est compacte. C'est l'objet de cette dernière section, et on obtiendra comme corollaire une preuve du caractère polonais de $\mathscr{M}^1(\mathfrak{X})$.

▷ **Familles tendues de mesures de probabilité.** Jusqu'à la fin du chapitre, $\mathfrak{X}$ est un espace topologique polonais.

Définition 1.31 (Tension). *Une famille $\mathscr{N} \subset \mathscr{M}^1(\mathfrak{X})$ de mesures de probabilité est dite tendue si, pour tout $\varepsilon > 0$, il existe un compact $K_\varepsilon \subset \mathfrak{X}$ tel que*

$$\forall \mu \in \mathscr{N},\ \ \mu(K_\varepsilon) \geq 1 - \varepsilon.$$

Exemple 1.32. Sur un espace polonais $\mathfrak{X}$ métrisé par une distance complète d, une famille constituée d'une seule mesure μ est toujours tendue. En effet, soit $(x_m)_{m\in\mathbb{N}}$ une suite dense dans $\mathfrak{X}$. Comme μ est de masse finie, pour tout entier k, on peut trouver un indice $\phi(k)$ tel que

$$\mu\left(\bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-1})}\right) \geq 1 - \frac{\varepsilon}{2^{k+1}}.$$

Notons alors $Y = \bigcap_{k=0}^{\infty} \bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-1})}$, et considérons une suite $(y_n)_{n\in\mathbb{N}}$ dans Y. Notons que la mesure de Y est plus grande que $1-\varepsilon$. On raisonne par extraction diagonale, et on construit une première sous-suite $(y_{\psi_1(n)})_{n\in\mathbb{N}}$ telle que tous les termes de cette suite tombent dans la même boule $B_{(x_{m_1},2^{-1-1})}$ avec $m_1 \leq \phi(1)$ (c'est possible par le principe des tiroirs). Puis, on construit une seconde sous-suite $(y_{\psi_1\circ\psi_2(n)})_{n\in\mathbb{N}}$ telle que tous les termes de cette suite tombent dans la même boule $B_{(x_{m_2},2^{-2-1})}$ avec $m_2 \leq \phi(2)$. Par récurrence, on peut construire des extractions successives $\psi_1,\ldots,\psi_l,\ldots$ et trouver des points $x_{m_1},\ldots,x_{m_l}$ tels que, pour tout l et tout n,

$$y_{\psi_1\circ\psi_2\circ\cdots\circ\psi_l(n)} \in B_{(x_{m_l},2^{-l-1})}.$$

On définit alors $y_{\Psi(l)} = y_{\psi_1\circ\psi_2\circ\cdots\circ\psi_l(l)}$. Étant donnés $l_1 \leq l_2$, les termes $y_{\Psi(l_1)}$ et $y_{\Psi(l_2)}$ appartiennent tous deux à la même boule $B_{(x_{m_{l_1}},2^{-l_1-1})}$, donc

$$d(y_{\Psi(l_1)}, y_{\Psi(l_2)}) \leq \frac{1}{2^{l_1}}.$$

On a donc extrait de $(y_n)_{n\in\mathbb{N}}$ une suite de Cauchy, et comme $(\mathfrak{X}, d)$ est complet, cette suite converge. Ainsi, Y est relativement compact, et la clôture $K = \overline{Y}$ est une partie compacte de $\mathfrak{X}$ telle que $\mu(K) \geq \mu(Y) \geq 1-\varepsilon$.

Exemple 1.33. A contrario, si $\mathfrak{X}$ n'est pas compact, alors on peut facilement construire des familles de mesures non tendues. Par exemple, $(\delta_n)_{n\in\mathbb{N}}$ est une famille non tendue de mesures de probabilité sur $\mathbb{R}$, car pour tout compact K, il existe un entier n tel que $n \notin K$ et donc $\delta_n(K) = 0$.

Théorème 1.34 (Prohorov)**.** *Soit $\mathfrak{X}$ un espace topologique polonais. Une partie $\mathscr{N} \subset \mathscr{M}^1(\mathfrak{X})$ est relativement compacte pour la topologie de la convergence en loi si et seulement si elle est tendue.*

Lemme 1.35. *Fixons une suite dense $(x_m)_{m\in\mathbb{N}}$ dans $\mathfrak{X}$. Soit $\mathscr{N}$ une famille relativement compacte dans $\mathscr{M}^1(\mathfrak{X})$. Pour tout $\varepsilon > 0$, et tout entier k, il existe un indice $\phi(k)$ tel que*

$$\mu\left(\bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-1})}\right) \geq 1 - \frac{\varepsilon}{2^{k+1}}$$

pour toute mesure $\mu \in \mathscr{N}$.

Preuve. Dans le cas contraire, on peut trouver k et $\delta = \frac{\varepsilon}{2^{k+1}}$, et une suite de mesures $(\mu_n)_{n\in\mathbb{N}}$ dans $\mathscr{N}$, tel que pour tout n,

$$\mu_n\left(\bigcup_{m\leq n} B_{(x_m,2^{-k-1})}\right) \leq 1-\delta.$$

Soit $(\mu_{\psi(n)})_{n\in\mathbb{N}}$ une sous-suite convergente, et μ la limite de cette sous-suite. On a pour tout entier n_0

$$\begin{aligned}\mu\left(\bigcup_{m\leq n_0} B_{(x_m,2^{-k-1})}\right) &\leq \liminf_n \mu_{\psi(n)}\left(\bigcup_{m\leq n_0} B_{(x_m,2^{-k-1})}\right)\\ &\leq \liminf_n \mu_{\psi(n)}\left(\bigcup_{m\leq \psi(n)} B_{(x_m,2^{-k-1})}\right) \leq 1-\delta.\end{aligned}$$

Par passage à la limite $n_0 \to \infty$, $\mu(\mathfrak{X}) = \mu(\bigcup_{m\in\mathbb{N}} B_{(x_m,2^{-k-1})}) \leq 1-\delta$, ce qui est absurde pour une mesure de probabilité. □

Le lemme implique que si $\mathscr{N}$ est une partie relativement compacte de $\mathscr{M}^1(\mathfrak{X})$, alors $\mathscr{N}$ est tendue ; en effet, on peut maintenant utiliser exactement le même raisonnement que lorsqu'on a montré qu'une seule mesure μ était tendue. L'implication réciproque va découler des résultats connus dans le cas compact, et du lemme suivant :

Lemme 1.36. *Soit $\mathfrak{X}$ un espace topologique polonais. Il existe un espace topologique métrisable compact $\mathfrak{W}$, et une application continue $\Phi : \mathfrak{X} \to \mathfrak{W}$ qui est un homéomorphisme entre $\mathfrak{X}$ et $\Phi(\mathfrak{X}) \subset \mathfrak{W}$.*

Preuve. On fixe comme précédemment une suite $(x_m)_{m\in\mathbb{N}}$ dense dans $\mathfrak{X}$, et on considère l'espace des suites $\mathfrak{W} = ([0,1])^{\mathbb{N}}$, muni de la topologie produit. Par le théorème de Tychonoff, comme $[0,1]$ est compact, $\mathfrak{W}$ est également compact, et il est métrisé par

$$d((a_n)_{n\in\mathbb{N}},(b_n)_{n\in\mathbb{N}}) = \sum_{n=0}^{\infty} \frac{|a_n - b_n|}{2^{n+1}}.$$

On considère l'application $\Phi : \mathfrak{X} \to \mathfrak{W}$, définie par

$$\Phi(x) = (\min(1, d(x_n, x)))_{n\in\mathbb{N}}.$$

Cette application est 1-lipschitzienne :

$$\begin{aligned} d(\Phi(x),\Phi(y)) &= \sum_{n=0}^{\infty} \frac{|\min(1,d(x_n,x)) - \min(1,d(x_n,y))|}{2^{n+1}} \\ &\leq \sum_{n=0}^{\infty} \frac{|d(x_n,x)-d(x_n,y)|}{2^{n+1}} \leq \sum_{n=0}^{\infty} \frac{d(x,y)}{2^{n+1}} = d(x,y). \end{aligned}$$

En particulier, c'est une application continue. Elle est injective, car si $\Phi(x) = \Phi(y)$, alors $\min(1,d(x_m,x)) = \min(1,d(x_m,y))$ pour tout y. Choisissant une sous-suite $(x_{m_l})_{l\in\mathbb{N}}$ qui converge vers x, on en déduit :

$$0 = \lim_{l\to\infty} \min(1,d(x_{m_l},x)) = \lim_{l\to\infty} \min(1,d(x_{m_l},y)) = \min(1,d(x,y)),$$

donc $d(x,y) = 0$ et $x = y$. Finalement, montrons que Φ est un homéomorphisme sur son image. Il faut montrer que si $y_n \not\to y$ dans $\mathfrak{X}$, alors $\Phi(y_n) \not\to \Phi(y)$ dans $\mathfrak{W}$. Comme $y_n \not\to y$, on peut trouver une sous-suite $(y_{n_k})_{k\in\mathbb{N}}$ et un $\varepsilon \in (0,1)$ tel que $d(y_{n_k}, y) \geq \varepsilon$ pour tout $k \in \mathbb{N}$. On peut également trouver un x_m proche de y à $\frac{\varepsilon}{3}$ près, et donc, par l'inégalité triangulaire, $d(y_{n_k}, x_m) \geq \frac{2\varepsilon}{3}$ pour tout k. Alors,

$$d(\Phi(y_{n_k}),\Phi(y)) \geq \frac{|\min(1,d(y_{n_k},x_m)) - \min(1,d(y,x_m))|}{2^{m+1}} \geq \frac{\varepsilon}{3\,2^{m+1}},$$

donc $\Phi(y_n) \not\to \Phi(y)$. □

Preuve du théorème 1.34. On a déjà vu l'une des implications, et il reste à montrer que si $\mathscr{N}$ est tendue, alors elle est relativement compacte. Comme $\mathscr{M}^1(\mathfrak{X})$ est métrisable, cette relative compacité est caractérisée séquentiellement : il suffit de montrer que tout suite de mesures $(\mu_n)_{n\in\mathbb{N}}$ dans $\mathscr{N}$ admet une sous-suite convergente en loi dans $\mathscr{M}^1(\mathfrak{X})$. On fixe comme dans le lemme précédent une injection continue $\Phi : \mathfrak{X} \to \mathfrak{W}$ vers un espace métrisable compact, et on note $\nu_n = \Phi_* \mu_n$ les lois images dans $\mathscr{M}^1(\mathfrak{W})$. Comme $\mathfrak{W}$ est métrisable compact, $\mathscr{M}^1(\mathfrak{W})$ l'est également, et il existe une sous-suite $(\nu_{n_k})_{k\in\mathbb{N}}$ qui converge en loi vers une mesure $\nu \in \mathscr{M}^1(\mathfrak{W})$. Montrons que cette mesure ν est supportée par $\Phi(\mathfrak{X})$. Soit $\varepsilon > 0$, et K_ε un compact de $\mathfrak{X}$ tel que $\mu_{n_k}(K_\varepsilon) \geq 1 - \varepsilon$ pour tout k. Comme Φ est continue, $\Phi(K_\varepsilon)$ est une partie compacte de $\mathfrak{W}$, donc fermée dans $\mathfrak{W}$. Par le théorème de Portmanteau,

$$\nu(\Phi(K_\varepsilon)) \geq \limsup_k \nu_{n_k}(\Phi(K_\varepsilon)) = \limsup_k \mu_{n_k}(K_\varepsilon) \geq 1 - \varepsilon.$$

A fortiori, $\nu(\Phi(\mathfrak{X})) \geq 1-\varepsilon$, et comme ceci est vrai pour tout $\varepsilon > 0$, $\nu(\Phi(\mathfrak{X})) = 1$. Ainsi, ν peut être considéré comme un élément de $\mathscr{M}^1(\Phi(\mathfrak{X}))$, et on peut alors définir une mesure de probabilité μ sur $\mathfrak{X}$ par la formule

$$\mu(A) = \nu(\Phi(A)) = (\Phi^{-1})_* \nu.$$

Comme $\Phi^{-1} : \Phi(\mathfrak{X}) \to \mathfrak{X}$ est continue, et $\nu_{n_k} \rightharpoonup \nu$, $(\Phi^{-1})_* \nu_{n_k} \rightharpoonup \mu$. Or,

$$(\Phi^{-1})_* \nu_{n_k} = \nu_{n_k} \circ \Phi = \mu_{n_k},$$

donc on a trouvé une sous-suite convergente en loi $\mu_{n_k} \rightharpoonup \mu$. □

▷ Complétude de l'espace des mesures de probabilité. Avant d'étudier les conséquences théoriques du théorème de Prohorov, revenons un instant sur la méthode d'approximation de mesures de probabilité décrite au début de la section 1.2. Pour construire une mesure de probabilité μ sur un espace $\mathfrak{X}$ qui vérifie certaines propriétés d'une liste P, il suffit maintenant :

1. de vérifier que les propriétés de la liste P, si elles sont satisfaites par une mesure μ, la caractérisent entièrement ;
2. de construire des approximations $(\mu_n)_{n\in\mathbb{N}}$ (lois de variables aléatoires $X_n \in \mathfrak{X}$), qui forment une famille tendue de mesures de probabilité, et qui satisfont asymptotiquement les propriétés de la liste P (autrement dit, toute limite potentielle de la suite μ_n doit satisfaire les propriétés dans P).

Alors, l'existence et l'unicité de la mesure μ sera assurée, et ce sera la limite en loi des mesures μ_n. C'est une méthode très efficace de construction de variables aléatoires, à deux conditions :

1. Il faut disposer de critères de tension simples pour une famille de mesures de probabilité dans $\mathscr{M}^1(\mathfrak{X})$. Ceci demande en premier lieu d'identifier les parties compactes de $\mathfrak{X}$. Dans le chapitre 2, on verra comment faire ceci dans un espace de dimension finie (variables aléatoires vectorielles), et dans le chapitre 3 on traitera le cas de l'espace des fonctions continues sur $\mathbb{R}_+$ (processus aléatoires continus).
2. Il faut également être capable de caractériser les lois sur $\mathfrak{X}$, et de vérifier qu'une liste de conditions P caractérise une loi. Dans un espace de dimension finie, la transformée de Fourier fournira un tel outil, et dans certains cas, on pourra aussi utiliser les moments des lois. Dans l'espace des fonctions continues sur $\mathbb{R}_+$, on verra que les lois fini-dimensionnelles caractérisent la loi d'un processus.

Voyons maintenant une conséquence théorique importante du théorème 1.34 :

Théorème 1.37. *Soit $\mathfrak{X}$ un espace topologique polonais. L'espace des mesures de probabilité $\mathscr{M}^1(\mathfrak{X})$ est également polonais.*

En analyse, une règle d'or est de toujours travailler dans des espaces complets (sinon, ceci veut dire que l'on a oublié des points limites possibles). Cette règle est bien satisfaite par $\mathscr{M}^1(\mathfrak{X})$, et la topologie de la convergence en loi est donc une bonne topologie sur cet espace.

Preuve. On sait déjà que $\mathscr{M}^1(\mathfrak{X})$ est métrisable séparable, et on va montrer que la distance de Lévy–Prohorov d_{LP} rend cet espace complet. Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite de Cauchy dans $\mathscr{M}^1(\mathfrak{X})$ pour la distance de Lévy–Prohorov. On va montrer que cette famille de mesures est tendue. Alors, $(\mu_n)_{n\in\mathbb{N}}$ aura des sous-suites convergentes, et d'autre part, les limites de ces sous-suites sont toutes les mêmes lorsqu'on a une suite de Cauchy, donc, par le critère habituel de convergence, $(\mu_n)_{n\in\mathbb{N}}$ aura bien une limite dans $\mathscr{M}^1(\mathfrak{X})$ (c'est le même raisonnement que lorsqu'on montre que $\mathbb{R}$ est complet en utilisant le théorème de Bolzano–Weierstrass).

On utilise une nouvelle fois le critère de tension suivant : une suite dense $(x_m)_{m\in\mathbb{N}}$ étant fixée, pour tout $\varepsilon \in (0,1)$ et tout entier k, il existe un indice $\phi(k)$ tel que

$$\mu_n\left(\bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-1})}\right) \geq 1-\frac{\varepsilon}{2^{k+1}}$$

pour tout n. Comme la suite $(\mu_n)_{n\in\mathbb{N}}$ est de Cauchy, on peut trouver un entier n_0 tel que $d_{\mathrm{LP}}(\mu_{n_0},\mu_n) \leq \frac{\varepsilon}{2^{k+2}}$ pour tout $n \geq n_0$. D'autre part, comme $\{\mu_0,\ldots,\mu_{n_0}\}$ est une famille finie de mesures de probabilité, et comme $(x_m)_{m\in\mathbb{N}}$ est dense, on peut trouver un indice $\phi(k)$ tel que

$$\mu_n\left(\bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-2})}\right) \geq 1-\frac{\varepsilon}{2^{k+2}}$$

pour tout $n \in [\![0,n_0]\!]$. Alors, l'inégalité

$$\mu_n\left(\bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-1})}\right) \geq 1-\frac{\varepsilon}{2^{k+1}}$$

est *a fortiori* vérifiée si $n \leq n_0$, et pour $n \geq n_0$, on a

$$\begin{aligned}\mu_n\left(\bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-1})}\right) &\geq \mu_n\left(\left(\bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-2})}\right)^{\frac{\varepsilon}{2^{k+2}}}\right)\\ &\geq \mu_{n_0}\left(\bigcup_{m\leq\phi(k)} B_{(x_m,2^{-k-2})}\right)-\frac{\varepsilon}{2^{k+2}} \geq 1-\frac{\varepsilon}{2^{k+1}}.\end{aligned}$$

Ceci achève la preuve du théorème. □

▷ **Représentation de Skorohod.** Nous concluons ce chapitre par un dernier résultat théorique au sujet de la convergence en loi.

Théorème 1.38 (Skorohod). *Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite de mesures de probabilité sur un espace polonais $\mathfrak{X}$, telle que $\mu_n \rightharpoonup \mu$. Il existe un espace de probabilité $(\Omega,\mathscr{F},\mathbb{P})$, et des variables aléatoires X_n de lois μ_n et X de loi μ qui sont définies sur cet espace, et telles que*

$$X_n \to_{\mathrm{p.s.}} X.$$

Attention, le théorème ne dit pas que si $X_n \rightharpoonup X$, alors $X_n \to_{\text{p.s.}} X$: si $X_n \rightharpoonup X$, alors quitte à étendre l'espace de probabilité, on peut trouver d'autres variables aléatoires X_n' et X' qui ont mêmes lois que X_n et X, et telles que $X_n' \to_{\text{p.s.}} X'$. La preuve de ce théorème repose sur la théorie de la tension de Prohorov, mais elle est assez difficile, et nous l'omettons ici. En pratique, c'est un résultat peu utile, mais il a des conséquences théoriques intéressantes.

Exemple 1.39. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires uniformément intégrables, qui converge en loi vers une variable intégrable X. Alors, $\mathbb{E}[X_n] \to \mathbb{E}[X]$. En effet, on peut remplacer X_n et X par des variables X_n' et X' de mêmes lois et telles que $X_n' \to_{\text{p.s.}} X'$. Comme $(X_n')_{n\in\mathbb{N}}$ est uniformément intégrable, et $X_n' \to_{\mathbb{P}} X'$, ceci implique

$$\mathbb{E}[X_n] = \mathbb{E}[X_n'] \to \mathbb{E}[X'] = \mathbb{E}[X].$$

Références

L'essentiel du contenu de ce chapitre est issu de [P. Billingsley, *Convergence of Probability Measures*, 2nd edition, Wiley, 1999, Chapter 1]. Pour la construction de l'intégrale de Lebesgue et de l'espace $\mathscr{L}^1_{\mathfrak{X}}$ par complétion de l'espace des fonctions en escalier, on renvoie à [S. Lang, *Real and Functional Analysis*, Graduate Texts in Mathematics vol. 142, Springer-Verlag, 1993, Chapter VI]. Cette construction est nettement plus simple que celle de [W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, 1987, Chapter 1], qui part de l'intégrale de fonctions positives, ce qui n'est ni très pertinent (par exemple, avec cette construction plus classique, on ne peut pas intégrer de fonctions à valeurs Banach), ni très efficace (on pourra comparer la longueur des preuves du théorème de convergence dominée avec ces deux approches).

Le théorème difficile de dualité entre $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ et $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$ pour $\mathfrak{X}$ polonais est démontré dans [N. Dunford and J. T. Schwartz, *Linear Operators*, Wiley, 1988, Theorem IV.6.2]. Pour le résultat plus simple de dualité entre $\mathscr{C}(\mathfrak{X})$ et $\mathscr{M}(\mathfrak{X})$ lorsque $\mathfrak{X}$ est compact, on renvoie à [S. Lang, *Real and Functional Analysis*, Graduate Texts in Mathematics vol. 142, Springer-Verlag, 1993, Chapter IX] ; le théorème de Banach–Alaoglu y est également démontré [*loc. cit.*, Chapter IV]. Finalement, pour le théorème de représentation de Skorohod, voir de nouveau [P. Billingsley, *Convergence of Probability Measures*, 2nd edition, Wiley, 1999, Chapter 1].

Exercices

1. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi, à valeurs dans $\mathbb{R}_+$. On suppose que $\mathbb{P}[X_n \geq t] > 0$ pour tout $t \in \mathbb{R}_+$.
   (a) Montrer que pour toute suite $(c_n)_{n\in\mathbb{N}}$ telle que $c_n \to 0$, $c_n X_n \to_{\mathbb{P}} 0$.
   (b) Construire une suite $(c_n)_{n\in\mathbb{N}}$ telle que $c_n \to 0$, mais $c_n X_n \not\to_{\text{p.s.}} 0$.

2. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables réelles indépendantes et de carré intégrable : $X_n \in \mathscr{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ pour tout n. On pose $S_n = \sum_{k=0}^n X_k$.
   (a) Comparer les espaces de Banach $(\mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P}), \|\cdot\|_{\mathscr{L}^1})$ et $(\mathscr{L}^2(\Omega, \mathscr{F}, \mathbb{P}), \|\cdot\|_{\mathscr{L}^2})$.
   (b) Montrer que si $(\mathbb{E}[S_n])_{n\in\mathbb{N}}$ et $(\mathrm{var}(S_n))_{n\in\mathbb{N}}$ convergent, alors S_n converge dans $\mathscr{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ vers une variable de carré intégrable S.
   (c) Montrer la réciproque de cette implication.

3. Dans cet exercice, $(\Omega, \mathscr{F}, \mathbb{P})$ est l'espace mesuré $([0,1], \mathscr{B}([0,1]), dx)$. On note

$$x = \frac{a_1(x)}{2} + \frac{a_2(x)}{4} + \frac{a_3(x)}{8} + \cdots = \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n}$$

   le développement dyadique propre d'un nombre réel, bien défini pour $x \in [0,1)$ (chaque $a_n(x)$ vaut 0 ou 1). On définit ensuite

$$n_k(x) = 2^k + \sum_{n=1}^{k} a_n(x)\, 2^{k-n} \in [\![2^k, 2^{k+1} - 1]\!].$$

   (a) Calculer le développement dyadique de $x = \frac{2}{3}$, puis la suite $(n_k(x))_{k\geq 1}$ pour $x = \frac{2}{3}$.
   (b) On pose $X_n = 1_{n=n_k(x)}$, où x est tiré au hasard dans $[0,1]$ suivant la mesure de Lebesgue. Que vaut $\mathbb{E}[X_n]$? Montrer que $X_n \to_{\mathscr{L}^1} 0$, mais que $X_n \not\to_{\text{p.s.}} 0$.
   (c) Soit $Y_n = nX_n$. Montrer que $Y_n \to_{\mathbb{P}} 0$, mais que $Y_n \not\to_{\mathscr{L}^1} 0$.

4. Soit $(\mathfrak{X}, \mathscr{B}, \mu)$ un espace mesuré, avec μ mesure positive mais pas forcément finie. On considère une suite de fonctions $(f_n)_{n\in\mathbb{N}}$ dans $\mathscr{L}^1(\mathfrak{X}, \mathscr{B}, \mu)$ telles que :

$$f_n \geq 0 \qquad ; \qquad f_n \to f \ \mu\text{-presque partout.}$$

   (a) Montrer le lemme de Scheffé : si $\int_{\mathfrak{X}} f_n(x)\,\mu(dx) \to \int_{\mathfrak{X}} f(x)\,\mu(dx)$, alors $f_n \to_{\mathscr{L}^1} f$. On pourra introduire la fonction $g_n = f - f_n$, et utiliser l'identité

$$|g_n| = 2\,\max(g_n, 0) - g_n$$

   et le théorème de convergence dominée pour $\mu(\max(g_n, 0))$.

(b) Soit $(X_n)_{n\in\mathbb{N}}$ une famille de variables aléatoires réelles. On suppose que les lois des variables X_n sont toutes absolument continues par rapport à une même mesure μ sur $\mathbb{R}$, et de densités $f_n(x)\,\mu(dx)$. Montrer que si f_n converge μ-presque partout vers une densité de probabilité f, alors $X_n \rightharpoonup X$, où X est une variable aléatoire de densité f par rapport à μ.

5. Soit $\{\mathscr{N}_{\mathbb{R}}(m_i,(\sigma_i)^2)\}_{i\in I}$ une famille de lois normales. Donner un critère nécessaire et suffisant sur $\{(m_i,(\sigma_i)^2)\}_{i\in I}$ pour que cette famille soit tendue.

6. Soit $\mathfrak{X}$ un espace polonais, et $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes de même loi $\mu \in \mathscr{M}^1(\mathfrak{X})$. Montrer que la suite des mesures empiriques

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

qui est une suite aléatoire dans $\mathscr{M}^1(\mathfrak{X})$, converge presque sûrement pour la topologie de la convergence en loi vers μ (indication : utiliser une base dénombrable d'ouverts de $\mathfrak{X}$).

7. Soit $\mathscr{X} \subset \mathscr{L}^1(\Omega,\mathscr{F},\mathbb{P})$ un ensemble de variables aléatoires réelles intégrables. Cette famille est dite uniformément intégrable si, pour tout $\varepsilon > 0$, il existe un réel positif M tel que

$$\forall X \in \mathscr{X},\ \mathbb{E}[|X|\,1_{|X|\geq M}] \leq \varepsilon.$$

(a) Montrer que dans ce cas, $\lim_{t\to 0}\sup_{X\in\mathscr{X}}\sup_{A\,|\,\mathbb{P}[A]\leq t}\mathbb{E}[|X|\,1_A] = 0$. On pourra s'inspirer du lemme du cours qui traite le cas où $\mathscr{X}$ est réduit à un seul élément.

(b) On suppose $\mathbb{E}[|X|] \leq K$ pour tout $X \in \mathscr{X}$ et pour une certaine constante $K \geq 0$. Montrer que réciproquement, si $\lim_{t\to 0}\sup_{X\in\mathscr{X}}\sup_{A\,|\,\mathbb{P}[A]\leq t}\mathbb{E}[|X|\,1_A] = 0$, alors $\mathscr{X}$ est uniformément intégrable.

La propriété d'uniforme intégrabilité pour $\mathscr{L}^1(\Omega,\mathscr{F},\mathbb{P})$ joue un rôle analogue à la tension pour $\mathscr{M}^1(\mathfrak{X})$. Plus précisément, on dispose du théorème de Dunford–Pettis. On rappelle que le dual topologique de $\mathscr{L}^1(\Omega,\mathscr{F},\mathbb{P})$ est l'espace $\mathscr{L}^\infty(\Omega,\mathscr{F},\mathbb{P})$ des fonctions presque sûrement bornées $(\Omega,\mathscr{F},\mathbb{P}) \to \mathbb{R}$. Ce dual définit la topologie faible de $\mathscr{L}^1(\Omega,\mathscr{F},\mathbb{P})$, et on peut montrer qu'une partie $\mathscr{X}$ de $\mathscr{L}^1(\Omega,\mathscr{F},\mathbb{P})$ est relativement compacte pour la topologie faible si et seulement si elle est uniformément intégrable.

8. Dans cet exercice, on s'intéresse à des conditions suffisantes pour que deux lois μ et ν sur un espace topologique $\mathfrak{X}$ soient égales.

(a) Soit μ et ν deux mesures de probabilité sur $\mathfrak{X}$ (muni de la tribu des boréliens), telles que $\mu(F) = \nu(F)$ pour tout fermé F de $\mathfrak{X}$. Montrer que $\mu = \nu$. On pourra introduire l'ensemble $\mathscr{A}$ des parties $A \subset \mathfrak{X}$ tel que $\mu(A) = \nu(A)$, et montrer que c'est une tribu.

(b) Soit μ et ν deux mesures finies sur $\mathfrak{X}$, telles que $\mu(f) = \nu(f)$ pour toute fonction continue bornée f. Montrer que $\mu = \nu$. En déduire qu'on a toujours un plongement de $\mathscr{M}(\mathfrak{X})$ dans $(\mathscr{C}_{\mathrm{b}}(\mathfrak{X}))^*$.

9. Si μ et ν sont deux mesures de probabilité sur un espace topologique $\mathfrak{X}$, leur distance en variation totale est définie par

$$d_{\mathrm{VT}}(\mu, \nu) = \sup_{A \in \mathscr{B}(\mathfrak{X})} |\mu(A) - \nu(A)| \in [0, 1].$$

   (a) Montrer que si $d_{\mathrm{VT}}(\mu_n, \mu)$ tend vers 0, alors $\mu_n \rightharpoonup \mu$. La réciproque est-elle vraie ?

   (b) On suppose $\mathfrak{X}$ polonais, et on voit $\mathscr{M}^1(\mathfrak{X})$ comme une partie fortement fermée de $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$, le dual topologique de $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$. Quel est le lien entre d_{VT} et la topologie forte de $\mathscr{M}_{\mathrm{far}}(\mathfrak{X})$?

   (c) Supposons $\mathfrak{X} = \mathbb{Z}$. Montrer que d_{VT} métrise la topologie de la convergence en loi sur $\mathscr{M}^1(\mathbb{Z})$. Montrer aussi que

$$d_{\mathrm{VT}}(\mu, \nu) = \frac{1}{2} \sum_{k \in \mathbb{Z}} |\mu(k) - \nu(k)|.$$

   (d) Plus généralement, soit μ et ν deux mesures de probabilité qui sont toutes les deux absolument continues par rapport à une mesure commune λ sur $\mathfrak{X}$. On note $\mu(dx) = m(x)\,\lambda(dx)$ et $\nu(dx) = n(x)\,\lambda(dx)$. Montrer que

$$d_{\mathrm{VT}}(\mu, \nu) = \frac{1}{2} \int_{\mathfrak{X}} |m(x) - n(x)|\,\lambda(dx).$$

10. Dans cet exercice, on donne des interprétations probabilistes des distances de Lévy–Prohorov et en variation totale, en termes de couplages.

   (a) Soient μ et ν deux mesures de probabilité sur un espace topologique métrisable $\mathfrak{X}$. Un couplage de μ et ν est une mesure de probabilité M sur $\mathfrak{X} \times \mathfrak{X}$, telle que la première marginale $(\pi_1)_* M$ de M est μ, et la seconde marginale $(\pi_2)_* M$ est ν (π_1 et π_2 sont les deux projections coordonnées $\mathfrak{X} \times \mathfrak{X} \to \mathfrak{X}$). Montrer que

$$\begin{aligned} &d_{\mathrm{VT}}(\mu, \nu) \\ &= \inf\{\varepsilon > 0, \text{ il existe un couplage } M \text{ tel que } M(\mathfrak{X}^2 \setminus D(\mathfrak{X})) \leq \varepsilon\}, \end{aligned}$$

   où $D(\mathfrak{X}) = \{(x, x) \,|\, x \in \mathfrak{X}\}$ est la diagonale dans $\mathfrak{X}^2$ (question préliminaire : pourquoi $D(\mathfrak{X})$ est-elle une partie mesurable ?).

   (b) Soit d une distance qui métrise la topologie de $\mathfrak{X}$, et d_{LP} la distance de Lévy–Prohorov qui lui est associée. Montrer que

$$\begin{aligned} &d_{\mathrm{LP}}(\mu, \nu) \\ &= \inf\{\varepsilon > 0, \text{ il existe un couplage } M \text{ tel que } M(\{(x, y) \,|\, d(x, y) \geq \varepsilon\}) \leq \varepsilon\}. \end{aligned}$$

   (c) Montrer qu'on a toujours $d_{\mathrm{LP}}(\mu, \nu) \leq d_{\mathrm{VT}}(\mu, \nu)$.

11. Détailler la preuve de l'inégalité triangulaire pour la distance de Lévy–Prohorov.

12. Soit $(V, \|\cdot\|)$ un espace vectoriel normé, et V^* son dual topologique, c'est-à-dire l'ensemble des formes linéaires $\phi : V \to \mathbb{R}$ qui sont continues. On rappelle que ceci est équivalent à demander que
$$\|\phi\| = \sup_{\|v\| \leq 1} |\phi(v)|$$
soit une quantité finie. On note $\overline{B}^* = \overline{B}^*_{(0,1)} = \{\phi \in V^* \,|\, \|\phi\| \leq 1\}$ la boule unité fermée de V^* pour cette norme sur les formes linéaires continues, et l'objectif de cet exercice est de montrer que $\overline{B}^*$ est compacte pour la topologie $*$-faible, c'est-à-dire la topologie la plus grossière sur V^* qui rend les applications
$$\begin{aligned} V^* &\to \mathbb{R} \\ \phi &\mapsto \phi(v) \end{aligned}$$
continues. Soit $\overline{B}$ la boule unité fermée de V, et Ψ l'application
$$\begin{aligned} \Psi : \overline{B}^* &\to ([-1,1])^{\overline{B}} \\ \phi &\mapsto (\phi(v))_{v \in \overline{B}}. \end{aligned}$$
On équipe $([-1,1])^{\overline{B}}$ de la topologie produit, et on rappelle que par le théorème de Tychonoff, c'est un espace topologique compact.
   (a) Montrer que Ψ est une application continue, et même un homéomorphisme entre $\overline{B}^*$ muni de la topologie $*$-faible, et $\Psi(\overline{B}^*)$ (indication : comparer les topologies sur $\Psi(\overline{B}^*)$ obtenues par transport de la topologie $*$-faible de $\overline{B}^*$ par Ψ, et par restriction de la topologie de $([-1,1])^{\overline{B}}$).
   (b) Montrer que $\Psi(\overline{B}^*)$ est une partie fermée de $([-1,1])^{\overline{B}}$. Conclure que $\overline{B}^*$ est compacte pour la topologie $*$-faible.
   (c) On suppose V séparable, et on se donne une suite $(v_n)_{n\in\mathbb{N}}$ dense dans V. Montrer que
$$d(\phi_1, \phi_2) = \sum_{n=0}^{\infty} \frac{\min(1, |\phi_1(v_n) - \phi_2(v_n)|)}{2^{n+1}}$$
et une distance sur V^* qui métrise la topologie $*$-faible.

13. Soit $\mathfrak{X}$ un espace polonais, et d une distance qui métrise la topologie. Une partie A de $\mathfrak{X}$ est dite précompacte si, pour tout $\varepsilon > 0$, il existe une famille finie $\{a_1, \ldots, a_r\}$ de points de A telle que $A \subset \bigcup_{i=1}^r B_{(a_i,\varepsilon)}$. Montrer qu'une partie A est précompacte si et seulement si elle est relativement compacte.

14. Soit $(\mathfrak{X}, d)$ un espace métrique.
   (a) On suppose $\mathfrak{X}$ compact. Montrer que l'algèbre des fonctions continues $\mathscr{C}_b(\mathfrak{X}) = \mathscr{C}(\mathfrak{X})$ est séparable. On pourra utiliser le théorème de Stone-Weierstrass et considérer la $\mathbb{Q}$-sous-algèbre engendrée par les fonctions constantes et les fonctions $d(\cdot, x_n)$, où $(x_n)_{n\in\mathbb{N}}$ est une suite dense dans $\mathfrak{X}$.

(b) On suppose maintenant $\mathfrak{X}$ non compact. Montrer que $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ n'est pas séparable. On pourra d'abord montrer qu'il existe une famille infinie de boules ouvertes $B_{(x_n,\varepsilon_n)}$ disjointes dans $\mathfrak{X}$, puis utiliser les fonctions $d(\cdot,x_n)$ pour construire une famille non dénombrable de fonctions dans $\mathscr{C}_{\mathrm{b}}(\mathfrak{X})$ qui sont toutes distantes de 1 pour la norme uniforme.

15. Dans cet exercice, on donne un critère simple sur un espace topologique pour qu'il soit métrisable (lemme d'Urysohn). On dit que $\mathfrak{X}$ espace topologique est normal si ses points $x \in \mathfrak{X}$ sont des fermés, et si pour tous fermés disjoints F_1 et F_2, il existe deux ouverts disjoints U_1 et U_2 tels que $F_1 \subset U_1$ et $F_2 \subset U_2$; autrement dit, on peut séparer les fermés disjoints par des ouverts. Dans tout ce qui suit, on fixe un espace topologique normal $\mathfrak{X}$.

   (a) Soit A une partie fermée de $\mathfrak{X}$, et U_1 un ouvert contenant A. Montrer qu'il existe un ouvert $U_{1/2}$ tel que
   $$A \subset U_{1/2} \subset \overline{U_{1/2}} \subset U_1.$$

   (b) Soit A et B deux fermés disjoints de $\mathfrak{X}$. Construire une famille d'ouverts $(U_r)_r$ indexée par les dyadiques $\frac{n}{2^k} \in (0,1]$, et telle que $A \subset U_r$ pour tout r, $U_1 = \mathfrak{X} \setminus B$, et $\overline{U_r} \subset U_s$ si $r \leq s$.

   (c) Avec les mêmes hypothèses, on définit une fonction $f : \mathfrak{X} \to [0,1]$ par la formule
   $$f(x) = \begin{cases} 1 & \text{si } x \in B \\ \inf\{r \mid x \in U_r\} & \text{si } x \notin B. \end{cases}$$
   Montrer que f est une fonction continue sur $\mathfrak{X}$, telle que $f_{|A} = 0$ et $f_{|B} = 1$.

   (d) On suppose maintenant $\mathfrak{X}$ normal et séparable, au sens suivant : il existe une famille dénombrable d'ouverts $(U_n)_{n\in\mathbb{N}}$ telle que tout ouvert U de $\mathfrak{X}$ soit une union de certains de ces ouverts de base (la définition diffère un peu du cas où $\mathfrak{X}$ est métrisable, mais lui est équivalente dans ce cas). Soit $(U_{n_i}, U_{m_i})_{i\in\mathbb{N}}$ la liste de toutes les paires d'ouverts de base telle que $\overline{U_{n_i}} \subset U_{m_i}$, et $(f_i)_{i\in\mathbb{N}}$ une liste de fonctions continues à valeurs dans $[0,1]$, telles que
   $$(f_i)_{|\overline{U_{n_i}}} = 0 \qquad ; \qquad (f_i)_{\mathfrak{X}\setminus U_{m_i}} = 1.$$
   On pose $d(x,y) = \sum_{i=0}^{\infty} \frac{1}{2^i}|f_i(x) - f_i(y)|$. Montrer que d est une distance sur $\mathfrak{X}$, et qu'elle métrise la topologie de $\mathfrak{X}$. Ainsi, tout espace topologique normal et séparable est métrisable.

16. Soit $\mathfrak{X}$ un espace polonais, et $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires indépendantes, avec $X_n \rightharpoonup X$ et $Y_n \rightharpoonup Y$. Montrer que si Z est la variable aléatoire dans $\mathfrak{X}^2$ dont la loi est $\mathbb{P}_Z = \mathbb{P}_X \otimes \mathbb{P}_Y$, alors $(X_n, Y_n) \rightharpoonup Z$. Si l'on retire l'hypothèse d'indépendance entre les X_n et les Y_n, montrer que le couple (X_n, Y_n) peut ne pas avoir de limite en loi.

17. Soit $\mathfrak{X}$ un espace polonais, et $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires, telles que $X_n \rightharpoonup X$ et $Y_n \to_{\mathbb{P}} y$, où y est une constante. Montrer qu'on a la convergence en loi $(X_n, Y_n) \rightharpoonup (X, y)$ (c'est le lemme de Slutsky).

18. Soit $\mathfrak{X}$ un espace métrisable. Une mesure borélienne sur $\mathfrak{X}$ est dite régulière si, pour toute partie A mesurable,

$$\mu(A) = \inf\{\mu(U) \,|\, U \text{ ouvert contenant } A\} = \sup\{\mu(F) \,|\, F \text{ fermé contenu dans } A\}.$$

Montrer que toute mesure de probabilité sur $\mathfrak{X}$ est régulière.

Chapitre 2

Convergence en loi et transformée de Fourier

Dans le chapitre précédent, nous avons présenté la notion de convergence en loi de variables aléatoires, dans le cadre général de variables à valeurs dans un espace métrique ou métrisable $\mathfrak{X}$. Dans ce chapitre et le suivant, on spécialise cette théorie au cas où $\mathfrak{X}$ est un espace numérique ou un espace de fonctions. On traite dans ce chapitre le cas où $\mathfrak{X} = \mathbb{R}$, la théorie se généralisant *mutatis mutandis* au cas d'espaces vectoriels réels $\mathbb{R}^d$ de dimension finie.

Dans la première section 2.1, on revisite la notion de convergence en loi dans le cas de variables réelles, et on la relie à la convergence des fonctions de répartition. On présente ensuite dans le paragraphe 2.2 le théorème de Lévy, qui est un critère pratique pour vérifier la convergence en loi de variables réelles : une suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si et seulement si les transformées de Fourier $\mathbb{E}[e^{i\xi X_n}]$ convergent ponctuellement vers la transformée de Fourier $\mathbb{E}[e^{i\xi X}]$. Si l'on peut contrôler localement uniformément la vitesse de convergence de ces transformées de Fourier, alors on peut en déduire une borne sur la distance de Kolmogorov $d_{\mathrm{Kol}}(X_n, X)$: c'est ce que l'on explique dans la section 2.3, avec comme application les inégalités de Berry–Esseen, qui sont vérifiées dans le cadre du théorème central limite avec une hypothèse de troisième moment.

2.1 Fonction de répartition et distance de Kolmogorov

▷ Fonctions de répartition et convergence en loi. Dans toute cette section, $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires réelles, et si X est une variable réelle, on note $F_X(s) = \mathbb{P}[X \leq s]$ la fonction de répartition de cette variable ; c'est une fonction croissante, continue à droite, avec $\lim_{s\to-\infty} F_X(s) = 0$ et $\lim_{s\to+\infty} F_X(s) = 1$. Si F_X est discontinue en s, alors

$$F_X(s) - F_X(s_-) = \lim_{r\nearrow s} \left(\mathbb{P}[X \in [r,s]]\right) = \mathbb{P}[X = s],$$

et s est un atome de la loi de X, dont le poids est la discontinuité de la fonction de répartition. Comme $\mathbb{P}_X$ est une mesure de probabilité, elle ne peut pas avoir plus de $\frac{1}{\varepsilon}$ atomes de poids supérieur à ε ; par conséquent, l'ensemble des discontinuités de F_X est inclus dans une union dénombrable d'ensembles finis

$$\bigcup_{k=1}^{\infty} \left\{ s \;\middle|\; \mathbb{P}[X=s] \geq \frac{1}{k} \right\},$$

et est donc dénombrable. Réciproquement, toute fonction croissante continue à droite sur $\mathbb{R}$ et avec limites 0 et 1 à l'infini est une fonction de répartition d'une mesure de probabilité.

La proposition suivante décrit la convergence en loi $X_n \rightharpoonup X$ en termes des fonctions de répartition :

Proposition 2.1. *On a convergence en loi de $(X_n)_{n\in\mathbb{N}}$ vers X si et seulement si, pour tout $s \in \mathbb{R}$ qui est un point de continuité de F_X, $\lim_{n\to\infty} F_{X_n}(s) = F_X(s)$.*

Preuve. Supposons $X_n \rightharpoonup X$. Si s est un point de continuité de F_X, alors $A = (-\infty, s]$ est une partie mesurable et $\delta A = \{s\}$ a mesure nulle pour $\mathbb{P}_X$. Donc, par le théorème de Portmanteau, $\lim_{n\to\infty} \mathbb{P}[X_n \in A] = \mathbb{P}[X \in A]$, c'est-à-dire que $\lim_{n\to\infty} F_{X_n}(s) = F_X(s)$.

Supposons réciproquement la convergence simple des fonctions de répartition $F_{X_n} \to F_X$ en tout point de continuité de F_X. Si $\varepsilon > 0$ est fixé, alors on peut trouver un compact $K = [-C, C]$ tel que $F_X(-C) \leq \varepsilon$ et $F_X(C) \geq 1-\varepsilon$, et comme F_X a un ensemble dénombrable de discontinuités, on peut supposer F_X continue en $-C$ et en C. Alors, pour n assez grand, $F_{X_n}(-C) \leq 2\varepsilon$ et $F_{X_n}(C) \geq 1-2\varepsilon$, car $F_{X_n}(-C) \to F_X(-C)$ et $F_{X_n}(C) \to F_X(C)$. Quitte à augmenter la valeur de C, on peut donc trouver $C > 0$ tel que, pour tout $n \in \mathbb{N}$,

$$F_{X_n}(-C) \leq 2\varepsilon \qquad ; \qquad F_{X_n}(C) \geq 1-2\varepsilon.$$

Par conséquent,

$$\mathbb{P}_{X_n}([-C, C]) = F_{X_n}(C) - F_{X_n}((-C)_-) \geq 1-4\varepsilon$$

pour tout $n \in \mathbb{N}$, ce qui montre que les lois $(\mathbb{P}_{X_n})_{n\in\mathbb{N}}$ forment une famille tendue. Il reste à voir que la seule limite possible d'une sous-suite convergente de $(\mathbb{P}_{X_n})_{n\in\mathbb{N}}$ est $\mathbb{P}_X$. D'après l'autre implication de la proposition, si $X_{n_k} \rightharpoonup Y$, alors $F_{X_{n_k}}(s) \to F_Y(s)$ sauf éventuellement sur un ensemble dénombrable, donc $F_X(s) = F_Y(s)$ pour presque tout $s \in \mathbb{R}$. Or, deux fonctions continues à droite sur $\mathbb{R}$ qui sont égales en dehors d'une partie dénombrable $D \subset \mathbb{R}$ sont en fait égales partout : en effet, si $d \in D$, alors on peut trouver une suite $x_n \searrow d$ avec $x_n \notin D$ pour tout n, et donc,

$$F_X(d) = \lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} F_Y(x_n) = F_Y(d).$$

Donc, si $(X_{n_k})_{k\in\mathbb{N}}$ converge en loi, alors sa limite est forcément X, ce qui implique $X_n \rightharpoonup X$ par le lemme 1.20. □

Corollaire 2.2. *Soit X une variable aléatoire réelle dont la loi $\mathbb{P}_X$ admet une densité f_X par rapport à la mesure de Lebesgue. Une suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si et seulement si F_{X_n} converge ponctuellement vers F_X.*

Preuve. Si X admet une densité f_X, alors sa fonction de répartition $F_X(s) = \int_{u=-\infty}^{s} f_X(u)\,du$ est continue sur $\mathbb{R}$. □

▷ **Distance de Kolmogorov.** En fait, on peut démontrer un résultat beaucoup plus fort, en utilisant un argument proche du théorème de Dini.

Définition 2.3 (Distance de Kolmogorov). *Si X et Y sont deux variables aléatoires réelles, leur distance de Kolmogorov est définie par la formule*

$$d_{\mathrm{Kol}}(X,Y) = \sup_{s\in\mathbb{R}} |F_X(s) - F_Y(s)|.$$

Si $X_n \rightharpoonup X$, alors il n'est pas vrai en général que $d_{\mathrm{Kol}}(X_n,X) \to 0$, car $F_{X_n}(s)$ peut ne pas converger vers $F_X(s)$ si s n'est pas un point de continuité de F_X. C'est en fait la seule obstruction :

Théorème 2.4. *Soit X une variable aléatoire réelle dont la fonction de répartition F_X est continue sur tout $\mathbb{R}$. Une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si et seulement si $d_{\mathrm{Kol}}(X_n,X) \to 0$.*

Preuve. Si $d_{\mathrm{Kol}}(X_n,X) \to 0$, alors en particulier F_{X_n} converge ponctuellement vers F_X en tout point $s \in \mathbb{R}$, donc $X_n \rightharpoonup X$ par la proposition précédente. Réciproquement, supposons $X_n \rightharpoonup X$; on sait alors que $F_{X_n}(s) \to F_X(s)$ pour tout $s \in \mathbb{R}$, et il s'agit de montrer que la convergence est en fait uniforme. Fixons $k \geq 1$, et des points $s_1 < s_2 < \cdots < s_{k-1}$ tels que $F_X(s_i) = \frac{i}{k}$ pour tout $i \in [\![1,k]\!]$; c'est possible car F_X est une fonction continue croissante de $\mathbb{R}$ vers $[0,1]$, avec $\lim_{s\to-\infty} F_X(s) = 0$ et $\lim_{s\to+\infty} F_X(s) = 1$. Pour n assez grand,

$$\frac{i-1}{k} \leq F_{X_n}(s_i) \leq \frac{i+1}{k}$$

pour tout $i \in [\![1,k-1]\!]$, compte tenu de la convergence simple des fonctions de répartition en tous les points s_i. Alors, si $s \in \mathbb{R}$, choisissant s_i tel que $s_i \leq s < s_{i+1}$ avec pour convention $s_0 = -\infty$ et $s_k = +\infty$, on a pour n assez grand

$$F_{X_n}(s) \leq F_{X_n}(s_{i+1}) \leq \frac{i+2}{k} = F_X(s_i) + \frac{2}{k} \leq F_X(s) + \frac{2}{k},$$

et de même,

$$F_{X_n}(s) \geq F_{X_n}(s_i) \geq \frac{i-1}{k} = F_X(s_{i+1}) - \frac{2}{k} \geq F_X(s) - \frac{2}{k}.$$

La distance de Kolmogorov entre X_n et X est donc plus petite que $\frac{2}{k}$ pour n assez grand, ce que l'on voulait démontrer. □

Ainsi, bien que la distance de Kolmogorov métrise une topologie plus fine que la topologie de la convergence en loi, elle contrôle bien cette convergence lorsque la limite a une fonction de répartition continue. En particulier, pour mesurer la vitesse d'une convergence en loi $X_n \rightharpoonup X$, on cherchera à estimer $d_{\mathrm{Kol}}(X_n,X)$ dès que X est par exemple une variable aléatoire à densité.

2.2 Transformée de Fourier et théorème de Lévy

▷ **Transformée de Fourier de mesures de probabilité.** Si X est une variable aléatoire réelle, on rappelle que :

Définition 2.5 (Transformée de Fourier)**.** *La transformée de Fourier de X est la fonction d'une variable $\xi \in \mathbb{R}$ définie par*

$$\phi_X(\xi) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X}].$$

On parle également de fonction caractéristique de X. Si μ est une mesure de probabilité sur $\mathbb{R}$, on définit de même sa transformée de Fourier par la formule $\widehat{\mu}(\xi) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi x}\,\mu(dx)$. Ainsi, si X a loi $\mathbb{P}_X$, alors $\phi_X(\xi) = \widehat{\mathbb{P}_X}(\xi)$.

Proposition 2.6. *La transformée de Fourier est une application injective de $\mathscr{M}^1(\mathbb{R})$ vers l'espace $\mathscr{C}_{\mathrm{b}}(\mathbb{R},\mathbb{C})$ des fonctions continues bornées de $\mathbb{R}$ dans $\mathbb{C}$.*

Preuve. Si μ est une mesure de probabilité, alors la fonction $(\xi, x) \mapsto \mathrm{e}^{\mathrm{i}\xi x}$ est continue en les deux variables et uniformément dominée par la fonction $x \mapsto 1$ qui est intégrable par rapport à μ, donc, par le théorème de continuité sous le signe intégral, $\xi \mapsto \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi x}\,\mu(dx)$ est continue et bornée par $\int_{\mathbb{R}} 1\,\mu(dx) = 1$. Il reste à montrer que $\mu \mapsto \widehat{\mu}$ est une application injective. On rappelle que si $\nu_{\sigma^2} = \mathscr{N}_{\mathbb{R}}(0,\sigma^2)$ est une loi gaussienne de densité $\nu_{\sigma^2}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{x^2}{2\sigma^2}}\,dx$, alors sa transformée de Fourier est $\widehat{\nu_{\sigma^2}}(\xi) = \mathrm{e}^{-\frac{\sigma^2\xi^2}{2}}$, voir les exercices en fin de chapitre. Si μ est une mesure de probabilité sur $\mathbb{R}$, on pose $\mu_{\sigma^2} = \mu * \nu_{\sigma^2}$, où $*$ désigne le produit de convolution de mesures. Ainsi,

$$\mu_{\sigma^2}(A) = \iint_{\mathbb{R}^2} 1_{(x+y\in A)}\,\mu(dx)\,\nu_{\sigma^2}(dy) = \int_A dt\left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{(x-t)^2}{2\sigma^2}}\,\mu(dx)\right)$$

pour toute partie mesurable A. L'injectivité de la transformée de Fourier est une conséquence des deux faits suivants :

1. Si l'on connaît la transformée de Fourier $\widehat{\mu}$, alors on peut calculer μ_{σ^2} pour tout $\sigma^2 > 0$.
2. Si l'on connaît la famille de mesures de probabilité $(\mu_{\sigma^2})_{\sigma>0}$, alors on peut retrouver la mesure de probabilité μ.

Pour le premier point, notons que $\mathrm{e}^{-\frac{(x-t)^2}{2\sigma^2}} = \widehat{\nu_{\frac{1}{\sigma^2}}}(x-t)$, et donc que

$$\begin{aligned}\mu_{\sigma^2}(A) &= \int_A dt\left(\int_{(x,y)\in\mathbb{R}^2} \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{\mathrm{i}y(x-t)}\,\nu_{\frac{1}{\sigma^2}}(dy)\,\mu(dx)\right)\\ &= \int_A dt\left(\int_{y\in\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\mathrm{i}yt}\,\widehat{\mu}(y)\,\nu_{\frac{1}{\sigma^2}}(dy)\right)\end{aligned}$$

ne dépend effectivement que de $\widehat{\mu}$. Pour la seconde partie, montrons que $\lim_{\sigma\to 0}\mu_{\sigma^2} = \mu$ au sens de la convergence en loi. Soit f une fonction uniformément continue et bornée sur $\mathbb{R}$;

on a $\mu_{\sigma^2}(f) = \mathbb{E}[f(X+Y_{\sigma^2})]$ avec $X \sim \mu$, $Y_{\sigma^2} \sim \nu_{\sigma^2}$, et X et Y_{σ^2} indépendants. Or,

$$\mathbb{P}[|Y_{\sigma^2}| \geq \eta] = \frac{2}{\sqrt{2\pi\sigma^2}} \int_\eta^\infty \mathrm{e}^{-\frac{y^2}{2\sigma^2}}\,dy = \frac{2}{\sqrt{2\pi}} \int_{\frac{\eta}{\sigma}}^\infty \mathrm{e}^{-\frac{y^2}{2}}\,dx = g\left(\frac{\eta}{\sigma}\right)$$

est une fonction décroissante de $\frac{\eta}{\sigma}$, avec $\lim_{\sigma\to 0} g(\frac{\eta}{\sigma}) = 0$. Soit $\varepsilon > 0$, B une borne sur $|f|$, et η un module d'ε-continuité pour f : si $|y_1 - y_2| \leq \eta$, alors $|f(y_1) - f(y_2)| \leq \varepsilon$. On peut écrire :

$$\begin{aligned} |\mu_{\sigma^2}(f) - \mu(f)| &= |\mathbb{E}[f(X+Y_{\sigma^2}) - f(X)]| \\ &\leq \mathbb{E}[|f(X+Y_{\sigma^2}) - f(X)|\,1_{|Y_{\sigma^2}|\leq\eta}] + 2B\,\mathbb{P}[|Y_{\sigma^2}| \geq \eta] \\ &\leq \varepsilon + 2B\,g\left(\frac{\eta}{\sigma}\right) \leq 2\varepsilon \text{ pour } \sigma \text{ suffisamment petit.} \end{aligned}$$

Donc, $\lim_{\sigma\to 0} \mu_{\sigma^2}(f) = \mu(f)$ pour toute fonction uniformément continue bornée, et $\mu_{\sigma^2} \rightharpoonup \mu$. Ceci achève la preuve de la proposition. □

Remarque 2.7. Une autre propriété importante de la transformée de Fourier de mesures de probabilité, et qui sous-tend le calcul reliant μ_{σ^2} à $\widehat{\mu}$, est la compatibilité avec le produit de convolution. Ainsi, pour toutes mesures de probabilité μ et ν sur $\mathbb{R}$, si $\mu * \nu$ est la loi de la somme de deux variables indépendantes de lois respectives μ et ν, alors :

$$\begin{aligned} \widehat{(\mu * \nu)}(\xi) &= \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi t}\,(\mu*\nu)(dt) = \int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}\xi(x+y)}\,\mu(dx)\,\nu(dy) \\ &= \left(\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi x}\,\mu(dx)\right)\left(\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi y}\,\nu(dy)\right) = \widehat{\mu}(\xi)\,\widehat{\nu}(\xi). \end{aligned}$$

▷ **Théorème de Lévy.** Comme $x \mapsto \mathrm{e}^{\mathrm{i}\xi x}$ est une fonction continue bornée, si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires réelles avec $X_n \rightharpoonup X$, alors $\phi_{X_n}(\xi) \to \phi_X(\xi)$ pour tout $\xi \in \mathbb{R}$. Le théorème de Lévy assure que la réciproque est vraie :

Théorème 2.8 (Lévy)**.** *Une suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si et seulement si $(\phi_{X_n})_{n\in\mathbb{N}}$ converge ponctuellement vers ϕ_X.*

Preuve. On vient de voir l'une des implications, et il s'agit maintenant de montrer que si $\phi_{X_n}(\xi) \to \phi_X(\xi)$ pour tout $\xi \in \mathbb{R}$, alors $X_n \rightharpoonup X$. Montrons d'abord que la famille des lois $(\mathbb{P}_{X_n})_{n\in\mathbb{N}}$ est tendue. On utilisera l'inégalité suivante : pour tout $x \in \mathbb{R}$ et tout $\delta > 0$,

$$1_{|\delta x|>2} \leq 2\left(1 - \frac{\sin \delta x}{\delta x}\right) = \frac{1}{\delta}\int_{-\delta}^{\delta} (1 - \cos(\xi x))\,d\xi.$$

En intégrant par rapport à $\mathbb{P}_{X_n}$, on obtient

$$\mathbb{P}\left[|X_n| > \frac{2}{\delta}\right] \leq \frac{1}{\delta}\int_{-\delta}^{\delta} \left(1 - \frac{\phi_{X_n}(\xi) + \phi_{X_n}(-\xi)}{2}\right) d\xi = \frac{1}{\delta}\int_{-\delta}^{\delta} \left(1 - \phi_{X_n}(\xi)\right) d\xi.$$

Fixons $\varepsilon > 0$. Comme ϕ_X est continue en 0, il existe $\delta > 0$ tel que $|1-\phi_X(\xi)| \leq \varepsilon$ pour tout $\xi \in [-\delta, \delta]$, et donc tel que

$$\frac{1}{\delta}\int_{-\delta}^{\delta} (1-\phi_X(\xi))\, d\xi \leq 2\varepsilon.$$

Comme ϕ_{X_n} converge partout vers ϕ_X et est bornée par 1, par convergence dominée, pour n assez grand, on a donc également

$$\mathbb{P}\left[|X_n| > \frac{2}{\delta}\right] \leq \frac{1}{\delta}\int_{-\delta}^{\delta} \left(1-\phi_{X_n}(\xi)\right) d\xi \leq 3\varepsilon.$$

Ceci prouve la tension des lois des variables $(X_n)_{n\in\mathbb{N}}$. Il reste alors à montrer que si $(X_{n_k})_{k\in\mathbb{N}}$ est une sous-suite convergente en loi, alors sa seule limite possible est (la loi de) X. Or, si $X_{n_k} \rightharpoonup Y$, alors $\phi_{X_{n_k}} \to \phi_Y$ par le sens direct du théorème, donc $\phi_X = \phi_Y$, et par la proposition 2.6, ceci implique $\mathbb{P}_X = \mathbb{P}_Y$. □

Exemple 2.9. Le théorème de Lévy est classiquement utilisé dans la preuve du théorème central limite : si $X_1, X_2, \ldots, X_n, \ldots$ sont des variables aléatoires indépendantes et de même loi avec $\mathbb{E}[X_1] = 0$ et $\mathbb{E}[(X_1)^2] = \sigma^2 < +\infty$, alors

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{n}}$$

converge en loi vers une gaussienne standard $\mathcal{N}_{\mathbb{R}}(0,1)$. En effet,

$$\phi_{S_n}(\xi) = \mathbb{E}\left[e^{i\xi\frac{X_1+\cdots+X_n}{\sigma\sqrt{n}}}\right] = \prod_{i=1}^{n}\mathbb{E}\left[e^{\frac{i\xi X_i}{\sigma\sqrt{n}}}\right] = \prod_{i=1}^{n}\phi_{X_i}\left(\frac{\xi}{\sigma\sqrt{n}}\right) = \left(\phi_{X_1}\left(\frac{\xi}{\sigma\sqrt{n}}\right)\right)^n.$$

Or, si Y est une variable aléatoire réelle avec $\mathbb{E}[|Y|^k] < +\infty$, alors le théorème de dérivation sous le signe intégral montre que ϕ_Y est dérivable continuement k fois sur $\mathbb{R}$, avec $(\phi_Y)^{(k)}(\xi) = \mathbb{E}[(iY)^k\, e^{i\xi Y}]$. Sous les hypothèses du théorème central limite, ϕ_{X_1} est dérivable deux fois, et on a le développement de Taylor $\phi_{X_1}\left(\frac{\xi}{\sigma\sqrt{n}}\right) = 1 - \frac{\xi^2}{2n}(1 + o_{n\to\infty}(1))$. Par conséquent,

$$\lim_{n\to\infty} \phi_{S_n}(\xi) = \lim_{n\to\infty}\left(1 - \frac{\xi^2}{2n}(1 + o_{n\to\infty}(1))\right)^n = e^{-\frac{\xi^2}{2}},$$

qui est la transformée de Fourier d'une gaussienne standard.

2.3 Inégalités de Berry–Esseen

Si l'on suit attentivement la preuve du théorème de Lévy 2.8, on peut s'attendre à ce qu'un contrôle au voisinage de 0 de la différence $\phi_{X_n} - \phi_X$ implique un contrôle sur la distance de Kolmogorov $d_{\mathrm{Kol}}(X_n, X)$. L'objectif de cette section est d'établir un résultat de ce type, lorsque X est une variable à densité avec une densité $f_X(x) = \frac{d\mathbb{P}_X(x)}{dx}$ uniformément bornée sur $\mathbb{R}$ par une constante $M > 0$. On va démontrer l'inégalité suivante :

Théorème 2.10 (Feller). *Soit X et Y deux variables aléatoires réelles, avec Y telle que f_Y existe et soit bornée par $M > 0$. Pour tout $T > 0$,*

$$d_{\mathrm{Kol}}(X,Y) \leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\phi_X(\xi)-\phi_Y(\xi)}{\xi}\right| d\xi + \frac{24\,M}{\pi T}.$$

Pour démontrer ce résultat, supposons dans un premier temps que X et Y ont des densités f_X et f_Y qui sont extrêmement régulières, par exemple dans l'espace de Schwartz. Alors, $F_X - F_Y$ est également dans l'espace de Schwartz, et en utilisant la formule d'inversion de Fourier, on obtient :

$$\begin{aligned} d_{\mathrm{Kol}}(X,Y) = \sup_{s\in\mathbb{R}} |F_X(s)-F_Y(s)| &= \frac{1}{2\pi}\sup_{s\in\mathbb{R}}\left|\int_{\mathbb{R}} \widehat{(F_X-F_Y)}(\xi)\,\mathrm{e}^{-\mathrm{i}\xi s}\,d\xi\right| \\ &\leq \frac{1}{2\pi}\int_{\mathbb{R}}\left|\widehat{(F_X-F_Y)}(\xi)\right| d\xi. \end{aligned}$$

Or, par intégration par parties,

$$\begin{aligned} \widehat{(F_X-F_Y)}(\xi) &= \int_{\mathbb{R}} (F_X(x)-F_Y(x))\,\mathrm{e}^{\mathrm{i}\xi x}\,dx \\ &= -\int_{\mathbb{R}} (f_X(x)-f_Y(x))\,\frac{\mathrm{e}^{\mathrm{i}\xi x}}{\mathrm{i}\xi}\,dx = \mathrm{i}\left(\frac{\phi_X(\xi)-\phi_Y(\xi)}{\xi}\right), \end{aligned}$$

donc on a dans ce cas l'inégalité

$$d_{\mathrm{Kol}}(X,Y) \leq \frac{1}{2\pi}\int_{\mathbb{R}}\left|\frac{\phi_X(\xi)-\phi_Y(\xi)}{\xi}\right| d\xi.$$

Pour traiter le cas général, on va essentiellement se ramener au cas précédent, en utilisant des noyaux de convolution qui vont nous permettre de travailler avec des densités dont les transformées de Fourier sont à support compact dans $[-T,T]$. Plus précisément, posons

$$k(x) = \frac{1}{\pi}\,\frac{1-\cos x}{x^2},$$

la fonction étant prolongée par continuité en $x = 0$.

Lemme 2.11. *On a $\int_{\mathbb{R}} k(x)\,dx = 1$, et plus généralement, $\int_{\mathbb{R}} k(x)\,\mathrm{e}^{\mathrm{i}\xi x}\,dx = \max(1-|\xi|,0)$.*

Preuve. On utilisera sans démonstration le fait suivant : si $f \in \mathscr{L}^2(\mathbb{R})$, alors sa transformée de Fourier $\widehat{f}(\xi) = \int_{\mathbb{R}} f(x)\,\mathrm{e}^{\mathrm{i}\xi x}\,dx$ appartient aussi à $\mathscr{L}^2(\mathbb{R})$ (la transformée de Fourier est définie d'abord sur $\mathscr{L}^1(\mathbb{R})\cap\mathscr{L}^2(\mathbb{R})$, puis prolongée par densité). De plus, si f et $\widehat{f}$ sont dans $\mathscr{L}^1(\mathbb{R})\cap\mathscr{L}^2(\mathbb{R})$, alors on a la formule d'inversion de Fourier

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\xi)\,\mathrm{e}^{-\mathrm{i}\xi x}\,d\xi$$

ainsi que la formule de Parseval $(\|f\|_{\mathscr{L}^2})^2 = \frac{1}{2\pi}(\|\widehat{f}\|_{\mathscr{L}^2})^2$ (voir les références données à la fin du chapitre). Comme le noyau k est dans $\mathscr{L}^1(\mathbb{R})\cap\mathscr{L}^2(\mathbb{R})$, la formule $\widehat{k}(\xi) = \max(1-|\xi|, 0)$ est donc équivalente à $k(x) = \frac{1}{2\pi}\int_{\mathbb{R}} h(\xi)\,\mathrm{e}^{-\mathrm{i}\xi x}\,d\xi$ avec $h(\xi) = \max(1-|\xi|, 0)$. Or,

$$\begin{aligned}\frac{1}{2\pi}\int_{\mathbb{R}} h(\xi)\,\mathrm{e}^{-\mathrm{i}\xi x}\,d\xi &= \frac{1}{2\pi}\int_{-1}^{1}(1-|\xi|)\,\mathrm{e}^{-\mathrm{i}\xi x}\,d\xi = \frac{1}{\pi}\int_0^1 (1-\xi)\,\cos(\xi x)\,d\xi \\ &= \frac{1}{\pi}\int_0^1 \frac{\sin(\xi x)}{x}\,d\xi = \frac{1}{\pi}\,\frac{1-\cos x}{x^2}.\end{aligned}$$

□

Lemme 2.12. *Soit X et Y deux variables aléatoires choisies comme dans l'énoncé du théorème 2.10. On pose $\Delta(s) = F_X(s) - F_Y(s)$, $\delta = \sup_{s\in\mathbb{R}} |\Delta(s)| = d_{\mathrm{Kol}}(X, Y)$,*

$$\Delta_T(s) = \int_{\mathbb{R}} \Delta(t)\; k(T(s-t))\,T\,dt,$$

et finalement $\delta_T = \sup_{s\in\mathbb{R}} |\Delta_T(s)|$. On a

$$\delta_T \geq \frac{\delta}{2} - \frac{12\,M}{\pi T}.$$

Preuve. Choisissons un point s où $|\Delta(s)|$ est proche de son maximum : $|F_X(s)-F_Y(s)| \geq \delta-\varepsilon$. On traitera le cas où $F_X(s) - F_Y(s) \geq 0$, l'autre cas s'y ramenant en remplaçant X par $-X$ et Y par $-Y$. Comme F_X est croissante et F_Y est M-Lipschitzienne, on a

$$\Delta(s+\eta) = F_X(s+\eta) - F_Y(s+\eta) \geq \delta - \varepsilon - M\eta$$

pour tout $\eta > 0$. Soit $s' = s + \frac{\delta-\varepsilon}{2M}$, et $\eta_0 = \frac{\delta-\varepsilon}{2M}$; pour tout $t \in [s'-\eta_0 = s, s'+\eta_0]$, on a

$$\Delta(t) \geq \frac{\delta-\varepsilon}{2} - M(t-s').$$

D'autre part, $\Delta(t) \geq -\delta$ pour les autres valeurs de t. Maintenant,

$$\begin{aligned}\delta_T \geq \Delta_T(s') &= \int_{\mathbb{R}} \Delta(t)\; k(T(s'-t))\,T\,dt \\ &\geq \int_{u=-\eta_0}^{\eta_0} \left(\frac{\delta-\varepsilon}{2} - Mu\right) k(Tu)\,T\,du - 2\delta\int_{\eta_0}^{+\infty} k(Tu)\,T\,du \\ &\geq \frac{\delta-\varepsilon}{2} - (3\delta-\varepsilon)\int_{\eta_0}^{+\infty} k(Tu)\,T\,du = \frac{\delta-\varepsilon}{2} - (3\delta-\varepsilon)\int_{\frac{T(\delta-\varepsilon)}{2M}}^{+\infty} k(x)\,dx.\end{aligned}$$

Comme ceci est vrai pour tout $\varepsilon > 0$, on conclut que

$$\delta_T \geq \frac{\delta}{2} - 3\delta\int_{\frac{T\delta}{2M}}^{+\infty} k(x)\,dx.$$

Or, la queue de distribution $\int_{\frac{T\delta}{2M}}^{+\infty} k(x)\,dx$ est plus petite que $\frac{4M}{\pi T\delta}$, car $k(x) \leq \frac{2}{\pi x^2}$; ceci permet de conclure. □

Preuve du théorème 2.10. Posons $k_T(x) = T\,k(Tx)$. La fonction $\Delta_T(x)$ introduite dans le lemme précédent a pour transformée de Fourier

$$\widehat{\Delta_T}(\xi) = \widehat{\Delta * k_T}(\xi) = \widehat{\Delta}(\xi)\,\widehat{k_T}(\xi) = \max\left(1 - \frac{|\xi|}{T}, 0\right)\widehat{\Delta}(\xi).$$

On a donc, avec le même raisonnement que dans le cas où les variables X et Y ont des densités très régulières :

$$\begin{aligned}\delta_T = \sup_{s\in\mathbb{R}} |\Delta_T(s)| &\leq \frac{1}{2\pi}\int_{\mathbb{R}} \left|\widehat{\Delta_T}(\xi)\right| d\xi \\ &\leq \frac{1}{2\pi}\int_{-T}^{T} \left|\widehat{\Delta}(\xi)\right| d\xi = \frac{1}{2\pi}\int_{-T}^{T} \left|\frac{\phi_X(\xi)-\phi_Y(\xi)}{\xi}\right| d\xi.\end{aligned}$$

Le théorème découle alors immédiatement du lemme précédent. □

Exemple 2.13. Appliquons l'inégalité précédente dans le cadre du théorème central limite, avec des variables aléatoires $X_1, \ldots, X_n$ indépendantes et de même loi, et

$$\mathbb{E}[X_1] = 0 \qquad ; \qquad \mathbb{E}[(X_1)^2] = \sigma^2 \qquad ; \qquad \mathbb{E}[|X_1|^3] = \rho < +\infty.$$

Posons $S_n = \frac{X_1+\cdots+X_n}{\sigma\sqrt{n}}$, et notons Y une variable gaussienne standard. On pose $T = \frac{\sigma^3\sqrt{n}}{\rho}$; comme $\sigma^3 \leq \rho$, $T \leq \sqrt{n}$. L'inégalité du théorème donne :

$$d_{\mathrm{Kol}}(S_n, Y) \leq \frac{1}{\pi}\int_{-T}^{T} \left|\frac{\left(\phi_{X_1}\left(\frac{\xi}{\sigma\sqrt{n}}\right)\right)^n - \mathrm{e}^{-\frac{\xi^2}{2}}}{\xi}\right| d\xi + \frac{24}{\pi\sqrt{2\pi}\,T}.$$

D'autre part, on a pour tout $t \in \mathbb{R}$

$$\left|\mathrm{e}^{\mathrm{i}t} - 1 - \mathrm{i}t + \frac{t^2}{2}\right| \leq \frac{|t|^3}{6}$$

par la formule de Taylor avec reste intégral. Par conséquent, en intégrant par rapport à la loi de X_1, on obtient

$$\left|\phi_{X_1}\left(\frac{\xi}{\sigma\sqrt{n}}\right) - 1 + \frac{\xi^2}{2n}\right| \leq \frac{\rho|\xi|^3}{6\sigma^3 n^{3/2}} = \frac{|\xi|^3}{6Tn} \leq \frac{\xi^2}{6n}.$$

si $|\xi| \leq T$. Donc, sur l'intervalle $[-T, T]$,

$$\left|\phi_{X_1}\left(\frac{\xi}{\sigma\sqrt{n}}\right)\right| \leq 1 - \frac{\xi^2}{2n} + \frac{\xi^2}{6n} = 1 - \frac{\xi^2}{3n} \leq \mathrm{e}^{-\frac{\xi^2}{3n}}.$$

Si a, b sont des réels de valeur absolue plus petite que c, alors $|a^n - b^n| \leq n|a-b|\,c^{n-1}$. Par conséquent, avec $a = \phi_{X_1}(\frac{\xi}{\sigma\sqrt{n}})$, $b = \mathrm{e}^{-\frac{\xi^2}{2n}}$ et $c = \mathrm{e}^{-\frac{\xi^2}{3n}}$, on obtient

$$\left|\left(\phi_{X_1}\left(\frac{\xi}{\sigma\sqrt{n}}\right)\right)^n - \mathrm{e}^{-\frac{\xi^2}{2}}\right| \leq n\,|a-b|\,c^{n-1}$$

avec

$$\begin{aligned}|a-b| &\leq \left|\phi_{X_1}\left(\frac{\xi}{\sigma\sqrt{n}}\right)-1+\frac{\xi^2}{2n}\right|+\left|\mathrm{e}^{-\frac{\xi^2}{2n}}-1+\frac{\xi^2}{2n}\right| \leq \frac{\xi^3}{6\,Tn}+\frac{\xi^4}{8n^2} \leq \frac{7|\xi|^3}{24\,Tn};\\ c^{n-1} &\leq \mathrm{e}^{-\frac{(n-1)\xi^2}{3n}} \leq \mathrm{e}^{-\frac{\xi^2}{4}}\end{aligned}$$

si $|\xi| \leq T$ et $n \geq 4$. Ainsi,

$$\left|\frac{\left(\phi_{X_1}\left(\frac{\xi}{\sigma\sqrt{n}}\right)\right)^n - \mathrm{e}^{-\frac{\xi^2}{2}}}{\xi}\right| \leq \frac{7\,\xi^2}{24\,T}\,\mathrm{e}^{-\frac{\xi^2}{4}}$$

sur l'intervalle $[-T, T]$, d'où

$$d_{\mathrm{Kol}}(S_n, Y) \leq \left(\frac{7\sqrt{\pi}}{6}+\frac{24}{\sqrt{2\pi}}\right)\frac{1}{\pi T} \leq \frac{4}{T} = \frac{4\rho}{\sigma^3\sqrt{n}}.$$

Comme le terme de droite est plus grand que $\frac{4}{\sqrt{n}}$, l'inégalité est également trivialement vraie si $n < 4$, donc on conclut :

Théorème 2.14 (Berry–Esseen)**.** *Soit $X_1, \ldots, X_n$ des variables indépendantes et de même loi, centrées et admettant des moments d'ordre* 3*. Si $S_n = \frac{X_1+X_2+\cdots+X_n}{\sqrt{n\,\mathbb{E}[(X_1)^2]}}$, alors pour tout $n \geq 1$,*

$$\sup_{s\in\mathbb{R}}\left|\mathbb{P}[S_n \leq s] - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{s}\mathrm{e}^{-\frac{x^2}{2}}\,dx\right| \leq \frac{4\,\mathbb{E}[|X_1|^3]}{(\mathbb{E}[(X_1)^2])^{3/2}}\,\frac{1}{\sqrt{n}}.$$

Remarque 2.15. On peut encore utiliser le théorème 2.10 pour évaluer des distances de Kolmogorov concernant des suites de variables aléatoires plus compliquées que des sommes renormalisées de variables indépendantes. De façon générale, un bon contrôle sur les transformées de Fourier donne toujours des estimées de type Berry–Esseen, mais un tel contrôle peut être très difficile à démontrer.

Références

L'essentiel de la discussion autour du théorème de Lévy peut être trouvé dans [O. Kallenberg, *Foundations of Modern Probability*, 2nd edition, Probability and Its Applications, Springer-Verlag, 2001, Chapter 5]. Pour les arguments de transformée de Fourier, on renvoie à [W. Rudin, *Functional Analysis*, 2nd edition, International Series in Pure and Applied Mathematics, McGraw-Hill, 1991, Chapter 7]. Le théorème de Berry–Esseen et l'inégalité générale qui relie la distance de Kolmogorov aux transformées de Fourier proviennent de [W. Feller, *An Introduction to Probability Theory and Its Applications, Volume II*, 2nd edition, Wiley Series in Probability and Mathematical Statistics, Wiley, 1971, Chapter XVI] (dans cet ouvrage, la preuve du théorème 2.14 est légèrement plus précise qu'ici, et elle donne une constante 3 au lieu de 4 dans l'inégalité).

Exercices

1. Soit X une variable aléatoire réelle.
   (a) Montrer que la fonction de répartition $F_X(s) = \mathbb{P}[X \leq s]$ est continue à droite.
   (b) On suppose que la loi $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue. Montrer que la fonction F_X est continue sur $\mathbb{R}$.
   (c) Si F_X est continue sur $\mathbb{R}$, la loi $\mathbb{P}_X$ est-elle absolument continue par rapport à la mesure de Lebesgue ?

2. Soit $X \sim \mathcal{N}_{\mathbb{R}}(m, \sigma^2)$ une variable aléatoire gaussienne de densité $\frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$. Montrer que sa transformée de Fourier est
$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X}] = \mathrm{e}^{\mathrm{i}\xi m - \frac{\sigma^2\xi^2}{2}}.$$
   On pourra d'abord se ramener au cas où $m = 0$ et $\sigma^2 = 1$, puis écrire une équation différentielle satisfaite par $\phi_X(\xi)$ et la résoudre.

3. Si $\lambda > 0$, on rappelle que la loi de Poisson de paramètre λ est la mesure de probabilité p_λ sur $\mathbb{N}$ telle que
$$p_\lambda(\{k\}) = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}.$$
   Soit $X_1, \ldots, X_n$ des variables indépendantes avec $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = \frac{\lambda}{n}$. On pose $S_n = X_1 + \cdots + X_n$. Calculer les transformées de Fourier de S_n et de la loi de Poisson p_λ, et en déduire que $S_n \rightharpoonup p_\lambda$ lorsque n tend vers l'infini (λ restant fixé).

4. Si μ et ν sont deux lois discrètes sur $\mathbb{Z}$, pour évaluer la distance entre μ et ν, on peut utiliser la distance en variation totale $d_{\mathrm{VT}}(\mu, \nu) = \frac{1}{2}\sum_{k\in\mathbb{Z}} |\mu(k) - \nu(k)|$ à la place de la distance de Kolmogorov.
   (a) Montrer que pour toutes lois discrètes $\mu_1, \ldots, \mu_n, \nu_1, \ldots, \nu_n$,
$$d_{\mathrm{VT}}(\mu_1 * \mu_2 * \cdots * \mu_n,\, \nu_1 * \nu_2 * \cdots * \nu_n) \leq d_{\mathrm{VT}}(\mu_1, \nu_1) + d_{\mathrm{VT}}(\mu_2, \nu_2) + \cdots + d_{\mathrm{VT}}(\mu_n, \nu_n).$$
   (b) En déduire qu'étant données des variables de Bernoulli indépendantes $X_1, \ldots, X_n$ avec $X_i \sim \mathcal{B}(p_i)$, si P suit une loi de Poisson de paramètre $\lambda = p_1 + \cdots + p_n$, alors
$$d_{\mathrm{VT}}(X_1 + \cdots + X_n, P) \leq \sum_{i=1}^{n} (p_i)^2.$$
   (c) Retrouver le résultat de l'exercice précédent.

5. Dans cet exercice, on considère une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ qui admettent des moments de tout ordre, et telles que $\mathbb{E}[(X_n)^k] \to \mathbb{E}[X^k]$ pour tout $k \geq 1$ (convergence des moments). Ceci n'implique pas en général la convergence en loi $X_n \rightharpoonup X$, car les moments d'une variable aléatoire ne déterminent pas en général sa loi.

(a) Soit μ une loi sur $\mathbb{R}$, telle que l'intégrale $m_k = \int_{\mathbb{R}} x^k \,\mu(dx)$ soit absolument convergente pour tout $k \geq 1$. On note $M_k = \int_{\mathbb{R}} |x|^k \,\mu(dx)$, et on suppose que la série entière
$$\mathbb{E}[\mathrm{e}^{zX}] = \sum_{k=0}^{\infty} \frac{m_k\, z^k}{k!}$$
a un rayon de convergence r non nul. Montrer que $\sum_{k=0}^{\infty} \frac{M_k\, z^k}{k!}$ a aussi un rayon de convergence R non nul (indication : utiliser l'inégalité $|x|^{2k-1} \leq 1 + x^{2k}$).

(b) Soit $\phi(\xi) = \widehat{\mu}(\xi)$. En utilisant l'inégalité
$$\left| \mathrm{e}^{itx} \left(\mathrm{e}^{ihx} - \sum_{k=0}^{K} \frac{(ihx)^k}{k!} \right) \right| \leq \frac{(|hx|)^{K+1}}{(K+1)!},$$
montrer que pour tout $|h| < R$ et tout $t \in \mathbb{R}$, on a le développement en série entière
$$\phi(t+h) = \sum_{k=0}^{\infty} \frac{\phi^{(k)}(t)}{k!}\, h^k.$$

(c) En déduire que si la série $\mathbb{E}[\mathrm{e}^{zX}]$ a un rayon de convergence r non nul, alors μ est l'unique mesure de probabilité dont les moments sont m_k (indication : montrer par récurrence sur $n \geq 1$ que les moments m_k déterminent $\widehat{\mu}$ sur l'intervalle $(-nR, nR)$).

(d) On suppose que X a une loi μ entièrement déterminée par ses moments comme dans les questions précédentes. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires admettant des moments de tout ordre, avec $\lim_{n\to\infty} \mathbb{E}[(X_n)^k] = \mathbb{E}[X^k]$. Montrer que la suite des lois $(\mathbb{P}_{X_n})_{n\in\mathbb{N}}$ est tendue, puis que $X_n \rightharpoonup X$ (indication : si Y est une limite d'une sous-suite convergente en loi $(X_{n_l})_{l\in\mathbb{N}}$, alors $(X_{n_l})^k \rightharpoonup Y^k$ et $((X_{n_l})^k)_{l\geq 1}$ est uniformément intégrable, donc $\mathbb{E}[X^k] = \lim_{l\to\infty} \mathbb{E}[(X_{n_l})^k] = \mathbb{E}[Y^k]$).

6. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi μ, et
$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$
les mesures empiriques de ces variables. On note F_n la fonction de répartition de μ_n, et F celle de μ.

(a) On suppose d'abord F continue. En s'inspirant de la preuve du théorème 2.4, montrer que $\lim_{n\to\infty} \sup_{s\in\mathbb{R}} |F_n(s) - F(s)| = 0$ presque sûrement.

(b) On ne suppose plus F continue. Soit $Y_1, \ldots, Y_n$ des variables i.i.d. de loi uniforme sur $[0,1]$, et $G(x \in [0,1]) = \inf\{s \in \mathbb{R} \,|\, F(s) > x\}$ l'inverse continu à droite de F. Si $X_i = G(Y_i)$, montrer que $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires indépendantes et de même loi μ, et que $F_n = A_n \circ F$, où A_n est la fonction de répartition empirique des Y_i. En déduire qu'on a encore presque sûrement $\lim_{n\to\infty} \sup_{s\in\mathbb{R}} |F_n(s) - F(s)| = 0$. C'est le théorème de Glivenko–Cantelli.

Chapitre 3

Convergence en loi de processus

Dans ce chapitre, on étudie la convergence en loi de processus aléatoires, c'est-à-dire des variables aléatoires à valeurs dans des espaces de fonctions. Par exemple, on pourra considérer une variable aléatoire X dans l'espace $\mathfrak{X} = \mathscr{C}(\mathbb{R}_+, \mathbb{R})$ de toutes les fonctions continues de $\mathbb{R}_+$ dans $\mathbb{R}$; c'est la bonne façon d'étudier une famille $(X_t)_{t\geq 0}$ de variables aléatoires qui varient continuement en t. Dans les deux premières sections 3.1 et 3.2, on introduit les espaces de fonctions $\mathfrak{X}$ dans lesquels on souhaitera étudier des processus aléatoires :

- l'espace des fonctions continues sur $[0,1]$, ou sur $\mathbb{R}_+$, ou plus généralement sur un espace métrique compact ou σ-localement compact ;
- et l'espace des fonctions continues à droite et avec limites à gauche (càdlàg) sur $[0,1]$ ou sur $\mathbb{R}_+$.

On montre dans chaque cas que l'espace de fonctions $\mathfrak{X}$ est polonais, et on donne un critère simple pour caractériser une loi sur $\mathfrak{X}$. On décrit aussi les parties relativement compactes de ces espaces ; pour les fonctions continues, une condition nécessaire et suffisante est donnée par le théorème d'Arzelà–Ascoli.

Le paragraphe 3.3 donne des critères de tension pour des familles de mesures de probabilité sur les espaces de fonctions $\mathfrak{X}$ précédemment listés ; dans le cas des fonctions continues sur $\mathbb{R}_+$, un critère simple reposant sur les moments des accroissements du processus est dû à Kolmogorov, et on en donne une preuve complète. Finalement, on applique ces résultats dans la section 3.4, en étudiant les limites en loi de marches aléatoires renormalisées. Plus précisément, soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles indépendantes et de même loi, avec $\mathbb{E}[X_1 = 0]$ et $\mathbb{E}[(X_1)^2] = 1$. On peut associer à $(X_n)_{n\in\mathbb{N}}$ les deux marches aléatoires suivantes :

$$C^{(n)}(t \geq 0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i + \frac{nt - \lfloor nt \rfloor}{\sqrt{n}} X_{\lfloor nt \rfloor + 1};$$

$$D^{(n)}(t \geq 0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i.$$

On verra dans la section 3.4 que les deux suites de processus $(C^{(n)})_{n\in\mathbb{N}}$ et $(D^{(n)})_{n\in\mathbb{N}}$ admettent la même limite en loi, à savoir le mouvement brownien, qui est un processus gaussien continu.

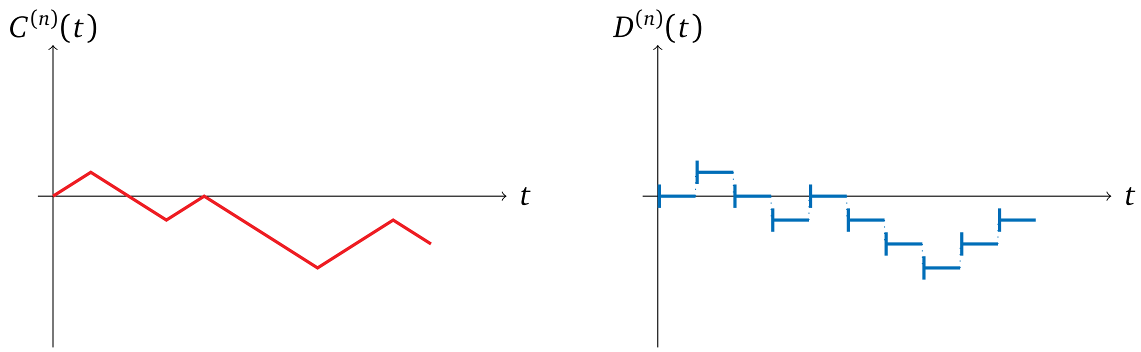


FIGURE 3.1 – Marches aléatoires associées à une suite de variables i.i.d.

3.1 L'espace des fonctions continues

Dans toute cette section, (T,d) est un espace métrique, et $\mathfrak{X} = \mathscr{C}(T) = \mathscr{C}(T,\mathbb{R})$ désigne l'espace des fonctions continues sur T et à valeurs dans $\mathbb{R}$. Les deux exemples importants qu'il faut garder en tête sont $\mathscr{C}([0,1])$ (espace des trajectoires continues à horizon fini) et $\mathscr{C}(\mathbb{R}_+)$ (espace des trajectoires continues à horizon infini), mais la théorie que nous allons développer sera valable plus généralement pour tout espace métrique T compact ou σ-localement compact.

▷ **Fonctions continues sur un espace compact.** On suppose dans un premier temps T compact ; alors, toute fonction f continue sur T est bornée et atteint ses extrema, et $\mathscr{C}(T)$ est un espace de Banach pour la topologie de la convergence uniforme donnée par la norme

$$\|f\|_\infty = \sup_{t\in T} |f(t)| = \max_{t\in T} |f(t)|.$$

En fait :

Proposition 3.1. *L'espace $\mathscr{C}(T)$ est un espace polonais.*

Preuve. On sait déjà que $\mathscr{C}(T)$ est métrisé (normé) par $\|\cdot\|_\infty$ et est complet pour cette distance, donc il reste à montrer qu'il est séparable, c'est-à-dire qu'il admet une sous-suite dense. Soit $(t_n)_{n\in\mathbb{N}}$ une suite dense dans T, $f_n(t) = d(t_n,t)$, et

$$\mathscr{A}_\mathbb{Q} = \mathbb{Q}[f_1, f_2, \ldots] \quad ; \quad \mathscr{A}_\mathbb{R} = \mathbb{R}[f_1, f_2, \ldots]$$

les sous-algèbres de $\mathscr{C}(T)$ engendrées respectivement sur $\mathbb{R}$ et sur $\mathbb{Q}$ par les fonctions constantes et les fonctions f_i. Il est clair que la clôture de $\mathscr{A}_\mathbb{Q}$ dans $\mathscr{C}(T)$ est la même que la clôture de $\mathscr{A}_\mathbb{R}$ (raisonner par double inclusion). Or, l'algèbre $\mathscr{A}_\mathbb{R}$ contient les constantes et sépare les points de T, car si $s \neq t$, choisissant une suite $(t_{n_k})_{k\in\mathbb{N}}$ telle que $\lim_{k\to\infty} t_{n_k} = t$, on a $\lim_{k\to\infty} f_{n_k}(s) = d(s,t) \neq 0$ et $\lim_{k\to\infty} f_{n_k}(t) = d(t,t) = 0$, donc pour k assez grand, $f_{n_k}(s) \neq f_{n_k}(t)$. Par le théorème de Stone–Weierstrass, $\mathscr{A}_\mathbb{R}$ est donc dense dans $\mathscr{C}(T)$, et il en va de même pour $\mathscr{A}_\mathbb{Q}$. D'autre part, $\mathscr{A}_\mathbb{Q}$ est dénombrable, car ses éléments P sont des polynômes à coefficients rationnels en les f_i, donc sont déterminés par une famille finie de multi-indices $\alpha \in (\mathbb{N}^*)^{(\mathbb{N})}$ et de coefficients $c_\alpha \in \mathbb{Q}$ tels que $P = \sum_\alpha c_\alpha f^\alpha$, avec $f^\alpha = (f_1)^{\alpha_1}(f_2)^{\alpha_2}\cdots$.

Comme $\mathbb{Q}$ et $(\mathbb{N}^*)^{(\mathbb{N})}$ sont dénombrables, on conclut que $\mathscr{A}_{\mathbb{Q}}$ est dénombrable, et que $\mathscr{C}(T)$ est bien séparable. □

Plus tard, on s'intéressera à des familles de mesures de probabilité sur $\mathscr{C}(T)$, et l'on essaiera d'établir leur tension. Il faudra alors savoir identifier les parties compactes de $\mathscr{C}(T)$. Le théorème suivant donne un critère simple de relative compacité. Si $\delta > 0$ et $f \in \mathscr{C}(T)$, on définit :

$$\omega(f,\delta) = \sup\{|f(s) - f(t)| \mid s,t \in T \text{ et } d(s,t) \leq \delta\}.$$

Comme les fonctions continues sur un compact sont uniformément continues par le théorème de Heine,

$$\lim_{\delta \to 0} \omega(f,\delta) = 0$$

pour toute fonction $f \in \mathscr{C}(T)$ si T est un espace métrique compact.

Théorème 3.2 (Arzelà–Ascoli)**.** *Soit $\mathscr{F} \subset \mathscr{C}(T)$ une partie de l'espace des fonctions continues sur le compact T. On pose $\omega(\mathscr{F},\delta) = \sup_{f \in \mathscr{F}} \omega(f,\delta)$. La partie $\mathscr{F}$ est relativement compacte si et seulement si les deux conditions suivantes sont vérifiées :*

1. *Pour tout $t \in T$, $\{f(t),\, f \in \mathscr{F}\}$ est une partie bornée de $\mathbb{R}$.*
2. *On a*
$$\lim_{\delta \to 0} \omega(\mathscr{F},\delta) = 0.$$

Ainsi, une partie de $\mathscr{C}(T)$ est relativement compacte si et seulement si elle est localement bornée et uniformément équicontinue.

Preuve. Supposons d'abord $\mathscr{F}$ relativement compacte dans $\mathscr{C}(T)$. Pour tout $t \in T$, l'application $\pi_t : f \mapsto f(t)$ est continue de $\mathscr{C}(T)$ dans $\mathbb{R}$, et même linéaire 1-lipschitzienne. La partie

$$\{f(t),\, f \in \mathscr{F}\} = \pi_t(\mathscr{F}) \subset \pi_t\big(\overline{\mathscr{F}}\big)$$

est donc incluse dans un compact, et donc bornée. Pour démontrer la deuxième implication, raisonnons par l'absurde et supposons que $\lim_{\delta\to 0} \omega(\mathscr{F},\delta) \neq 0$. Alors, on peut trouver $\varepsilon > 0$, une suite $\delta_n \to 0$, et des fonctions $f_n \in \mathscr{F}$ telles que $\omega(f_n,\delta_n) \geq \varepsilon$ pour tout n. Ceci veut dire qu'il existe des points s_n, t_n avec $d(s_n,t_n) \leq \delta_n$ et $|f_n(s_n) - f_n(t_n)| \geq \varepsilon$ pour tout $n \in \mathbb{N}$. Comme T est compact et $\mathscr{F}$ est une partie relativement compacte, quitte à extraire des sous-suites, on peut supposer $s_n \to s$, $t_n \to t$ et $f_n \to f$ pour une certaine fonction $f \in \mathscr{C}(T)$. Comme $\delta_n \to 0$, on a forcément $s = t$. Dans ce cadre, $f(s) = \lim_{n\to\infty} f_n(s_n)$, car

$$|f(s) - f_n(s_n)| \leq |f(s) - f(s_n)| + \|f - f_n\|_\infty \to 0,$$

puisque f est continue en s et $f_n \to f$ dans $\mathscr{C}(T)$. De même, $f(t) = \lim_{n\to\infty} f_n(t_n)$, d'où une contradiction car $|f_n(s_n) - f_n(t_n)|$ reste toujours plus grand que $\varepsilon > 0$. Ceci achève la preuve de l'implication directe.

Pour la réciproque, soit $\mathscr{F}$ une partie de $\mathscr{C}(T)$ qui vérifie les deux hypothèses du théorème, et $(f_n)_{n\in\mathbb{N}}$ une suite dans $\mathscr{F}$. On fixe une suite $(\delta_m)_{m\in\mathbb{N}}$ de nombres réels positifs tendant vers 0, et pour chaque m, un recouvrement du compact T par des boules ouvertes $B_{(t_m^1,\delta_m)}, \ldots, B_{(t_m^{N_m},\delta_m)}$. Pour chaque point t_m^i, la partie $\{f_n(t_m^i),\, n \in \mathbb{N}\}$ est bornée, donc, on peut extraire une sous-suite de $(f_n)_{n\in\mathbb{N}}$ qui converge ponctuellement en t_m^i. Ces points étant en quantité dénombrable, par extraction diagonale, on peut trouver une sous-suite $(f_{n_k})_{k\in\mathbb{N}}$ de $(f_n)_{n\in\mathbb{N}}$ telle que $(f_{n_k}(t_m^i))_{k\in\mathbb{N}}$ ait une limite pour tout point t_m^i avec $m \in \mathbb{N}$ et $i \in [\![1, N_m]\!]$. Montrons alors que la suite $(f_{n_k})_{k\in\mathbb{N}} = (g_k)_{k\in\mathbb{N}}$ est de Cauchy dans $\mathscr{C}(T)$, et donc convergente. Soit $\varepsilon > 0$, et m assez grand tel que $\omega(\mathscr{F}, \delta_m) < \varepsilon$. Comme toutes les suites $(g_k(t_m^i))_{k\in\mathbb{N}}$ avec $i \in [\![1, N_m]\!]$ sont de Cauchy, pour K assez grand, si $k_1, k_2 \geq K$, alors

$$\forall i \in [\![1, N_m]\!],\ \left|g_{k_1}(t_m^i) - g_{k_2}(t_m^i)\right| \leq \varepsilon.$$

Soit t un point quelconque de T. Il existe $i \in [\![1, N_m]\!]$ tel que $t \in B_{(t_m^i, \delta_m)}$. Alors, si $k_1, k_2 \geq K$,

$$\begin{aligned}\left|g_{k_1}(t) - g_{k_2}(t)\right| &\leq \left|g_{k_1}(t) - g_{k_1}(t_m^i)\right| + \left|g_{k_1}(t_m^i) - g_{k_2}(t_m^i)\right| + \left|g_{k_2}(t_m^i) - g_{k_2}(t)\right| \\ &\leq \left|g_{k_1}(t_m^i) - g_{k_2}(t_m^i)\right| + 2\,\omega(\mathscr{F}, \delta_m) \leq 3\varepsilon.\end{aligned}$$

Donc, la suite $(g_k)_{k\in\mathbb{N}}$ est uniformément de Cauchy, et on a trouvé une sous-suite de $(f_n)_{n\in\mathbb{N}}$ qui converge uniformément sur T. □

Remarque 3.3. Si T est compact connexe, on peut affaiblir légèrement les hypothèses du théorème 3.2 et remplacer la première condition par :

1'. Pour un certain $t_0 \in T$, $\{f(t_0),\ f \in \mathscr{F}\}$ est une partie bornée de $\mathbb{R}$.

En effet, la seconde condition montre alors que $\{f(t),\ f \in \mathscr{F}\}$ est une partie bornée pour tout t dans une boule autour de t_0, et de proche en proche on montre le résultat pour tout $t \in T$. En particulier, si l'on considère l'espace $\mathscr{C}([0,1])$, alors on utilisera souvent le théorème d'Ascoli sous la forme suivante : une partie $\mathscr{F}$ de $\mathscr{C}([0,1])$ est relativement compacte si et seulement si :

1. La partie $\{f(0),\ f \in \mathscr{F}\}$ est bornée.
2. On a $\lim_{\delta\to 0} \omega(\mathscr{F}, \delta) = 0$.

▷ **Fonctions continues sur un espace σ-localement compact.** La discussion précédente montre que si T est un espace métrique compact, alors $\mathfrak{X} = \mathscr{C}(T)$ est un espace adéquat pour appliquer la théorie du premier chapitre, car il est polonais et on sait reconnaître ses parties compactes (ce qui sera requis pour déterminer la tension d'une famille de lois sur $\mathfrak{X}$). Si T n'est plus compact, on peut encore dans certains cas équiper $\mathscr{C}(T)$ d'une topologie avec les mêmes propriétés ; la bonne hypothèse est alors celle de σ-locale compacité.

Définition 3.4 (σ-locale compacité). *Un espace métrique T est dit σ-compact s'il existe une famille dénombrable de compacts $(K_n)_{n\in\mathbb{N}}$, qu'on peut supposer croissante pour l'inclusion, et telle que $T = \bigcup_{n\in\mathbb{N}} K_n$. Il est dit σ-localement compact si, de plus, tout point $t \in T$ admet un voisinage compact.*

Exemple 3.5. La droite réelle $\mathbb{R}$ est σ-compacte, car c'est l'union des compacts $[-n, n]$ avec n parcourant $\mathbb{N}$. Elle est même σ-localement compacte, puisque tout $x \in \mathbb{R}$ admet comme voisinage compact $[x - \varepsilon, x + \varepsilon]$.

Lemme 3.6. *Soit T un espace métrique σ-localement compact.*

1. *L'espace T est séparable.*
2. *Il existe une base dénombrable d'ouverts relativement compacts $(U_n)_{n\in\mathbb{N}}$.*
3. *Il existe une famille dénombrable de compacts $(K_n)_{n\in\mathbb{N}}$, croissante pour l'inclusion, telle que tout compact K de T soit inclus dans l'un des K_n.*

Preuve. Soit $(K_n)_{n\in\mathbb{N}}$ une famille dénombrable de compacts de T telle que $T = \bigcup_{n\in\mathbb{N}} K_n$. Tout espace métrique compact est séparable, donc pour tout n, il existe $(x_{n,m})_{m\in\mathbb{N}}$ dense dans K_n. Alors, $\{x_{n,m}\}_{(n,m)\in\mathbb{N}^2}$ est une famille dénombrable dense dans T, donc T est bien séparable.

Dans ce qui suit, on fixe une suite $(t_k)_{k\in\mathbb{N}}$ dense dans T. Chaque point t_k admet un voisinage relativement compact, donc, pour ε_k assez petit, $B_{(t_k,\varepsilon_k)}$ est relativement compact. On note Q_k l'ensemble des nombres rationnels $\varepsilon_k > 0$ tels que $B_{(t_k,\varepsilon_k)}$ soit relativement compact dans T. Notons que si $\varepsilon_k \in Q_k$ et $0 < \eta_k < \varepsilon_k$ avec η_k rationnel, alors η_k est aussi dans Q_k, car la clôture $\overline{B_{(t_k,\eta_k)}}$ est un fermé inclus dans le compact $\overline{B_{(t_k,\varepsilon_k)}}$, donc est compacte elle-même. La famille de boules ouvertes

$$B_{(t_k,\varepsilon_k)},\ k\in\mathbb{N},\ \varepsilon_k\in Q_k$$

est alors une base dénombrable d'ouverts relativement compacts de T. En effet, il reste à montrer que si U est un ouvert de T, alors U s'écrit comme réunion de certaines de ces boules ouvertes. Soit $D = \{k \mid t_k \in U\}$, et pour chaque point $t_k \in U$, $R_k \subset Q_k$ l'ensemble des nombres ε_k tels que $B_{(t_k,\varepsilon_k)} \subset U$. On a :

$$U = \bigcup_{k\in D,\,\varepsilon_k\in R_k} B_{(t_k,\varepsilon_k)}.$$

En effet, l'inclusion $\supset$ est évidente par définition de D et des ensembles R_k. Réciproquement, soit $t\in U$, et $\varepsilon > 0$ rationnel tel que $B_{(t,\varepsilon)}$ soit un voisinage relativement compact de t entièrement inclus dans U. Si t_k est un point de la suite $(t_k)_{k\in\mathbb{N}}$ tel que $d(t_k,t) < \frac{\varepsilon}{2}$, alors la boule $B_{(t_k,\frac{\varepsilon}{2})}$ contient t et est entièrement incluse dans U, donc $\varepsilon_k = \frac{\varepsilon}{2} \in R_k$ et $t \in \bigcup_{k\in D,\,\varepsilon_k\in R_k} B_{(t_k,\varepsilon_k)}$. On a donc bien exhibé une base dénombrable d'ouverts relativement compacts.

Finalement, notons $(U_n)_{n\in\mathbb{N}}$ une base dénombrable d'ouverts relativement compacts dans T, et $K_n = \bigcup_{k=0}^n \overline{U_k}$. Chaque K_n est un compact, et la suite $(K_n)_{n\in\mathbb{N}}$ est croissante pour l'inclusion, avec $\bigcup_{n=0}^{\infty} K_n = \bigcup_{n=0}^{\infty} U_n = T$. Soit $K\subset T$ un compact. Il existe un sous-recouvrement fini $\bigcup_{k=0}^n U_k$ de K par des ouverts de base, et *a fortiori*, $K \subset \bigcup_{k=0}^n \overline{U_k} = K_n$. □

Si T est un espace σ-localement compact, on équipe l'espace des fonctions continues $\mathscr{C}(T)$ de la topologie de la convergence uniforme sur tout compact $K\subset T$. Une base de voisinage d'une fonction $f\in\mathscr{C}(T)$ est formée des parties

$$\left\{ g\in\mathscr{C}(T) \,\middle|\, \|g-f\|_{\infty,K} = \sup_{t\in K} |g(t)-f(t)| \leq \varepsilon \right\},$$

avec $\varepsilon > 0$, et où K parcourt l'ensemble des parties compactes de T.

Proposition 3.7. *Si T est σ-localement compact et si $\mathscr{C}(T)$ est muni de la topologie de la convergence uniforme sur tout compact, alors $\mathscr{C}(T)$ est un espace polonais.*

Lemme 3.8 (Prolongement de Tietze)**.** *Soit $K \subset T$ compact, et f une fonction continue sur K. Il existe une fonction continue bornée g sur T telle que $g_{|K} = f$.*

Preuve. Sans perte de généralité, on peut supposer f à valeurs dans $[0,1]$. Soit $K_- = f^{-1}([0,\frac{1}{3}])$ et $K_+ = f^{-1}([\frac{2}{3},1])$. Ce sont deux fermés disjoints dans K et donc dans T. Tout espace métrique est normal, car on peut utiliser la fonction distance pour séparer les fermés disjoints par des ouverts. Donc, d'après le lemme d'Urysohn (voir les exercices du chapitre 1), il existe une fonction g_1 continue sur T, à valeurs dans $[0,1]$, et telle que $(g_1)_{|K_-} = 0$ et $(g_1)_{|K_+} = 1$. Alors, la fonction

$$f_1 = \frac{3}{2}\left(f - \frac{(g_1)_{|K}}{3}\right)$$

est une fonction continue sur K qui prend de nouveau ses valeurs dans $[0,1]$. On peut donc trouver g_2 telle que

$$f_2 = \frac{3}{2}\left(f_1 - \frac{(g_2)_{|K}}{3}\right)$$

prenne ses valeurs dans $[0,1]$, et telle que g_2 soit continue sur T à valeurs dans $[0,1]$. Notons qu'on a :

$$f = \frac{(g_1)_{|K}}{3} + \frac{2f_1}{3} = \frac{(g_1)_{|K}}{3} + \frac{2(g_2)_{|K}}{9} + \frac{4f_2}{9}.$$

Par récurrence sur n, on peut trouver une suite de fonctions $(g_n)_{n\in\mathbb{N}}$, avec $g_n \in \mathscr{C}(T,[0,1])$ pour tout n, et avec

$$\left(\frac{2}{3}\right)^n \geq \left\| f - \frac{(g_1)_{|K}}{3} - \frac{2(g_2)_{|K}}{9} - \cdots - \frac{2^{n-1}(g_n)_{|K}}{3^n} \right\|_{\infty,K}.$$

Alors, la série de fonctions $\sum_{n=1}^{\infty} \frac{2^{n-1} g_n}{3^n}$ est uniformément convergente sur T, donc elle a une limite g qui est une fonction continue bornée sur T. De plus, $g_{|K} = f$ d'après l'inégalité précédente. $\square$

Preuve de la proposition 3.7. Soit $(K_n)_{n\in\mathbb{N}}$ une suite exhaustive de compacts, c'est-à-dire une suite croissante de compacts telle que $T = \bigcup_{n\in\mathbb{N}} K_n$ et telle que tout compact K de T soit inclus dans un K_n. On équipe $\mathscr{C}(T)$ de la distance suivante :

$$d(f,g) = \sum_{n=0}^{\infty} \frac{\min(1, \|f-g\|_{\infty,K_n})}{2^n}.$$

Comme tout compact K est inclus dans un K_n, on voit facilement que d est bien une distance sur $\mathscr{C}(T)$, et qu'une suite $(f_m)_{m\in\mathbb{N}}$ de fonctions continues sur T converge ou est de Cauchy pour cette distance si et seulement si chaque suite de fonctions $((f_m)_{|K})_{m\in\mathbb{N}}$ converge ou est de Cauchy dans $\mathscr{C}(K)$, K parcourant l'ensemble des compacts de T. Ainsi, d est bien une distance qui métrise la topologie de la convergence uniforme sur tout compact. De plus, comme les espaces $\mathscr{C}(K)$ sont complets, une suite de Cauchy dans $\mathscr{C}(T)$ converge uniformément sur tout compact K, donc converge dans $\mathscr{C}(T)$; ainsi, $(\mathscr{C}(T), d)$ est un espace métrique complet.

Pour la séparabilité, on fixe pour tout $n \in \mathbb{N}$ une suite $(f_{n,m})_{m\in\mathbb{N}}$ dense dans $\mathscr{C}(K_n)$. D'après le lemme précédent, on peut étendre toutes ses fonctions par continuité sur T, donc supposer que chaque fonction $f_{n,m}$ est en fait définie sur T. Alors, la famille $(f_{n,m})_{(n,m)\in\mathbb{N}}$ est dense dans $\mathscr{C}(T)$. En effet, soit

$$V_{(f,\varepsilon,K)} = \left\{ g \in \mathscr{C}(T) \,\middle|\, \|g - f\|_{\infty,K} = \sup_{t\in K} |g(t) - f(t)| \leq \varepsilon \right\}$$

un ouvert de base de cet espace. Si $K \subset K_n$, alors il existe par hypothèse $f_{m,n}$ telle que $\|f - f_{m,n}\|_{\infty,K_n} \leq \varepsilon$. *A fortiori*, $\|f - f_{m,n}\|_{\infty,K} \leq \varepsilon$, donc $f_{m,n} \in V_{(f,\varepsilon,K)}$. □

Proposition 3.9. *Si T est σ-localement compact, alors une partie $\mathscr{F} \subset \mathscr{C}(T)$ est relativement compacte si et seulement si, pour tout compact $K \subset T$, la partie $\mathscr{F}_{|K} = \{f_{|K} \,|\, f \in \mathscr{F}\}$ est relativement compacte dans $\mathscr{C}(K)$.*

Preuve. Comme la projection

$$\begin{aligned} \phi_K : \mathscr{C}(T) &\to \mathscr{C}(K) \\ f &\mapsto f_{|K} \end{aligned}$$

est continue, si $\mathscr{F}$ est relativement compacte, alors $\mathscr{F}_{|K} = \phi_K(\mathscr{F}) \subset \phi_K(\overline{F})$ est une partie incluse dans un compact, donc relativement compacte. Réciproquement, supposons toutes les parties $\mathscr{F}_{|K}$ relativement compactes, et considérons une suite $(f_m)_{m\in\mathbb{N}}$ dans $\mathscr{F}$, ainsi qu'une suite exhaustive de compacts $(K_n)_{n\in\mathbb{N}}$ dans T. Les familles $((f_m)_{|K_n})_{m\in\mathbb{N}}$ sont toutes relativement compactes, donc, par extraction diagonale, on peut trouver une sous-suite $(f_{m_l})_{l\in\mathbb{N}}$ qui converge uniformément sur chaque compact K_n. Comme tout compact K est inclus dans un K_n, $(f_{m_l})_{l\in\mathbb{N}}$ converge donc uniformément sur tout compact K, c'est-à-dire qu'elle converge dans $\mathscr{C}(T)$. Ainsi, $\mathscr{F}$ est une partie relativement compacte. □

Remarque 3.10. La preuve de la proposition précédente montre qu'on peut remplacer dans l'énoncé "pour tout compact K" par "pour tout compact K_n d'une suite exhaustive de compacts", et cette hypothèse est plus simple à vérifier.

Remarque 3.11. La discussion de ce paragraphe servira essentiellement à traiter le cas où $T = \mathbb{R}_+$. Dans ce cas, on peut prendre comme suite exhaustive de compacts $K_n = [0, n]$, et la plupart des preuves précédentes sont un peu plus simples dans ce cadre. Une distance sur $\mathscr{C}(\mathbb{R}_+)$ correspondant à la convergence uniforme sur tout compact est

$$d(f,g) = \sum_{n=1}^{\infty} \frac{\min(1, \|f - g\|_{\infty,[0,n]})}{2^n},$$

et une partie $\mathscr{F}$ de $\mathscr{C}(\mathbb{R}_+)$ est relativement compacte si et seulement si les deux conditions suivantes sont vérifiées :

1. La partie $\{f(0) \,|\, f \in \mathscr{F}\}$ est bornée.
2. Pour tout $n \in \mathbb{N}$, $\lim_{\delta\to 0} \sup\{|f(x) - f(y)| \,|\, x, y \in [0,n],\ |x - y| \leq \delta,\ f \in \mathscr{F}\} = 0$.

▷ **Lois fini-dimensionnelles.** Pour conclure notre présentation de l'espace des fonctions continues $\mathscr{C}(T)$, nous allons donner une caractérisation simple des mesures de probabilité sur cet espace. Dans tout ce qui suit, T est un espace métrique σ-localement compact. Si $t_1, t_2, \ldots, t_d$ sont des points de T, alors l'application

$$\begin{aligned} \mathscr{C}(T) &\to \mathbb{R}^d \\ f &\mapsto (f(t_1), f(t_2), \ldots, f(t_d)) \end{aligned}$$

est continue, donc mesurable. Étant donnée une loi borélienne μ sur $\mathscr{C}(T)$, on note $\mu_{t_1,\ldots,t_d}$ la loi sur $\mathbb{R}^d$ qui est l'image de μ par l'application ci-dessus. On dit que les lois $\mu_{t_1,\ldots,t_d}$ sont les lois fini-dimensionnelles issues de μ. Ces lois vérifient des conditions de compatibilité évidentes : si $\phi_d^D : \mathbb{R}^D \to \mathbb{R}^d$ envoie un vecteur de D coordonnées $(t_1, \ldots, t_D)$ sur le vecteur de $d \leq D$ coordonnées $(t_1, \ldots, t_d)$, alors pour tous points $t_1, \ldots, t_D$ dans T,

$$(\phi_d^D)_* \mu_{t_1,\ldots,t_D} = \mu_{t_1,\ldots,t_d}.$$

Théorème 3.12. *Soit μ et ν deux mesures de probabilité sur $\mathscr{C}(T)$, avec T espace métrique σ-localement compact. Si μ et ν ont les mêmes lois fini-dimensionnelles, alors elles sont égales.*

Preuve. La preuve du théorème repose sur la comparaison de deux tribus sur $\mathscr{C}(T)$. En tant qu'espace métrique muni d'une distance d correspondant à la topologie de la convergence uniforme sur tout compact, $\mathscr{C}(T)$ est muni de la tribu des boréliens $\mathscr{B}(\mathscr{C}(T))$ engendrée par les ouverts

$$V_{(f,\varepsilon,K)} = \{g \in \mathscr{C}(T) \,|\, \|g - f\|_{\infty,K} < \varepsilon\}.$$

D'autre part, $\mathscr{C}(T)$ est une partie de $\mathbb{R}^T$, l'ensemble des familles de nombres réels $(x_t)_{t\in T}$ indexée par T ; et $\mathbb{R}^T$ est muni de la tribu produit $\bigotimes_{t\in T} \mathscr{B}(\mathbb{R})$. La restriction de cette tribu produit à $\mathscr{C}(T)$, que l'on notera $\mathscr{P}$, est la tribu engendrée par les ouverts

$$W_{(f,\varepsilon,t_1,\ldots,t_d)} = \{g \in \mathscr{C}(T) \,|\, \forall i \in [\![1,d]\!],\ |g(t_i) - f(t_i)| < \varepsilon\}.$$

Comme l'inclusion $i : \mathscr{C}(T) \to \mathbb{R}^T$ est continue, et comme $\mathscr{P} = i^{-1}(\bigotimes_{t\in T} \mathscr{B}(\mathbb{R}))$, on a $\mathscr{P} \subset \mathscr{B}(\mathscr{C}(T))$. De plus, l'hypothèse du théorème est équivalente au fait que $\mu(B) = \nu(B)$ pour toute partie $B \in \mathscr{P}$ (on a égalité sur un π-système de parties qui engendre la tribu $\mathscr{P}$). Il suffit donc de démontrer que l'on a en fait :

$$\mathscr{P} = \mathscr{B}(\mathscr{C}(T)).$$

Or, si $(t_k)_{k\in\mathbb{N}}$ est une suite dense dans T, alors

$$V_{(f,\varepsilon,K)} = \bigcup_{l=2}^{\infty} \bigcap_{t_k \in K} W_{(f,(1-\frac{1}{l})\varepsilon,t_k)}.$$

En effet, si $g \in V_{(f,\varepsilon,K)}$, alors il existe $l \geq 2$ tel que $\|g - f\|_{\infty,K} \leq (1 - \frac{1}{l+1})\varepsilon$, et dans ce cas, $|g(t_i) - f(t_i)| < (1 - \frac{1}{l})\varepsilon$ pour tout $t_i \in K$, ce qui démontre l'inclusion $\subset$. Réciproquement, si g est dans un ensemble $\bigcap_{t_k\in K} W_{(f,(1-\frac{1}{l})\varepsilon,t_k)}$, alors $|g(t) - f(t)| < (1 - \frac{1}{l})\varepsilon$ sur un ensemble dense de K, donc par densité de $(t_k)_{k\in\mathbb{N}}$ dans T et par continuité de f et de g,

$$\|g - f\|_{\infty,K} \leq \left(1 - \frac{1}{l}\right)\varepsilon < \varepsilon.$$

Ainsi, on a l'inclusion réciproque $\supset$. Il s'ensuit que $V_{(f,\varepsilon,K)} \in \mathscr{P}$ pour tout ouvert de base $V_{(f,\varepsilon,K)}$, et donc que $\mathscr{P} = \mathscr{B}(\mathscr{C}(T))$. □

Corollaire 3.13. *Soit $(\mu_{n\in\mathbb{N}})_{n\in\mathbb{N}}$ une famille de mesures de probabilité sur $\mathscr{C}(T)$, avec T espace métrique σ-localement compact. On suppose la famille de mesures $(\mu_n)_{n\in\mathbb{N}}$ tendue, et les lois fini-dimensionnelles $(\mu_n)_{t_1,\ldots,t_d}$ qui convergent toutes sur $\mathbb{R}^d$. Alors, $(\mu_n)_{n\in\mathbb{N}}$ a une limite pour la convergence en loi dans $\mathscr{M}^1(\mathscr{C}(T))$.*

Preuve. La limite en loi d'une sous-suite convergente de $(\mu_n)_{n\in\mathbb{N}}$ est unique, puisque ses lois fini-dimensionnelles le sont et la déterminent. Le résultat découle alors comme d'habitude du lemme 1.20, puisque $(\mu_n)_{n\in\mathbb{N}}$ est tendue et donc relativement compacte. □

Remarque 3.14. La réciproque du corollaire est également vraie : si $\mu_n \rightharpoonup \mu$, alors la famille $(\mu_n)_{n\in\mathbb{N}}$ est évidemment tendue, et d'autre part, commes les applications $\pi_{t_1,\ldots,t_d}$ sont continues, on a aussi $(\mu_n)_{t_1,\ldots,t_d} \rightharpoonup \mu_{t_1,\ldots,t_d}$ pour tous temps $t_1,\ldots,t_d$.

3.2 L'espace des fonctions càdlàg

Dans la section précédente, on a étudié en détail la topologie d'espaces tels que $\mathscr{C}([0,1])$ (cas compact) ou $\mathscr{C}(\mathbb{R}_+)$ (cas σ-localement compact), ce qui permettra de manipuler des lois de processus continus à horizon fini ou infini. On souhaite maintenant étendre cette étude à certains processus discontinus. En particulier, on souhaite pouvoir étudier les processus de Poisson, qui sont les processus $(N_t)_{t\in\mathbb{R}_+}$ à valeurs dans $\mathbb{N}$, croissants, tels que chaque ensemble $N^{-1}(\{k\})$ soit un intervalle $[t_k, t_{k+1})$ avec $t_{k+1} - t_k = \xi_k$ suivant une loi exponentielle de paramètre λ, et avec les variables $\xi_{k\geq 0}$ toutes indépendantes.

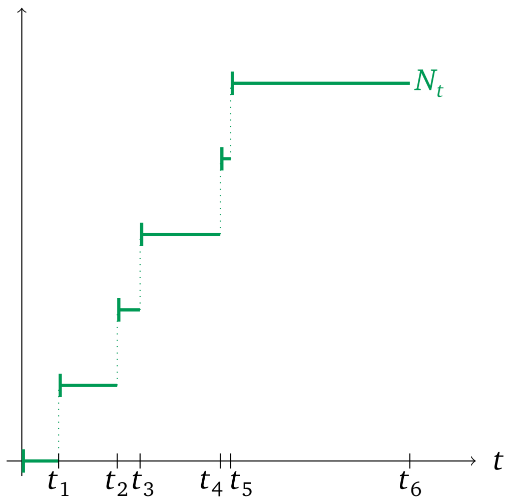


FIGURE 3.2 – Un processus de Poisson de paramètre $\lambda = 1$.

Le bon espace de fonctions pour l'étude de ces processus est l'espace $\mathscr{D}(\mathbb{R}_+)$, défini comme suit :

Définition 3.15 (Fonctions càdlàg). *L'espace $\mathscr{D}(\mathbb{R}_+)$ est l'ensemble des fonctions $f : \mathbb{R}_+ \to \mathbb{R}$ telles que les limites $\lim_{s\to t, s<t} f(s)$ et $\lim_{s\to t, s>t} f(s)$ existent pour tout $t \in \mathbb{R}_+$, et telles que la seconde limite soit toujours*

$$\lim_{\substack{s\to t\\ s>t}} f(s) = f(t).$$

On définit de même $\mathscr{D}(I)$ pour tout intervalle $I \subset \mathbb{R}_+$. Une fonction f appartenant à $\mathscr{D}(\mathbb{R}_+)$ ou $\mathscr{D}(I)$ est dite continue à droite et avec limites à gauche (en abrégé, càdlàg).

L'objectif de cette section est de munir $\mathscr{D}(\mathbb{R}_+)$ d'une topologie qui en fait un espace polonais, et telle que les processus de Poisson par exemple puissent être considérés comme des variables aléatoires dans cet espace.

▷ **Structure des discontinuités.** Soit $0 = t_0 < t_1 < t_2 < \cdots < t_k < \cdots$ une suite de temps dans $\mathbb{R}_+$ qui tend vers l'infini, et $w_0, w_1, w_2, \ldots$ des nombres réels. On note

$$f(t) = \sum_{k \,|\, t_k \leq t} w_k.$$

On obtient une fonction continue à droite et avec limites à gauche, et les fonctions que l'on peut construire de cette façon généralisent celles qui viennent de processus de Poisson (dans ce dernier cas, $w_k = 1$ pour tout k). Une topologie raisonnable sur $\mathscr{D}(\mathbb{R}_+)$ doit être telle que, si f et g sont des fonctions associées à des temps $s_1 < s_2 < \cdots$ et $t_1 < t_2 < \cdots$ et à des sauts $v_0, v_1, v_2, \ldots$ et $w_0, w_1, w_2, \ldots$, alors f est proche de g dans $\mathscr{D}(\mathbb{R}_+)$ si et seulement si les temps s_k sont proches des temps t_k et si les tailles des sauts v_k sont proches des w_k.

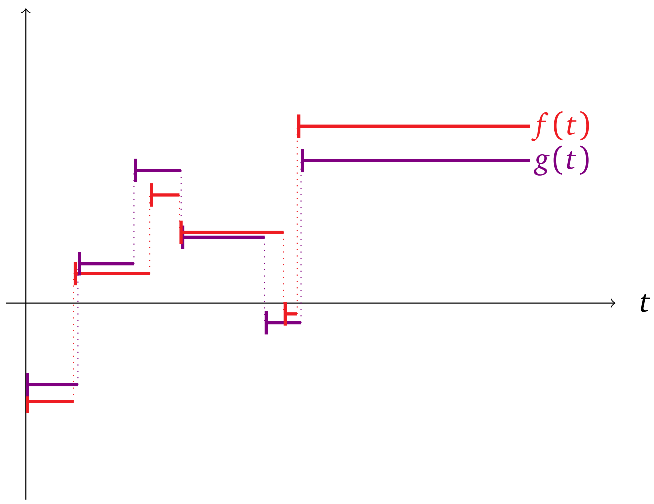


FIGURE 3.3 – Deux fonctions proches dans $\mathscr{D}(\mathbb{R}_+)$.

La topologie de la convergence uniforme sur tout compact est trop fine pour que cette condition soit vérifiée : comme le montre la figure 3.3, si les sauts de deux fonctions $f, g \in \mathscr{D}(\mathbb{R}_+)$ ont lieu à des instants très proches mais distincts, alors la norme infinie $\|f - g\|_{\infty,[0,T]}$ peut être grande, bien que les graphes des deux fonctions f et g soient très proches. Dans ce qui suit, nous allons définir sur $\mathscr{D}(\mathbb{R}_+)$ une topologie un peu plus grossière, mais qui restreinte à $\mathscr{C}(\mathbb{R}_+)$ correspondra à la topologie de la convergence uniforme sur tout compact.

Comme préliminaire, il convient de décrire un peu mieux la structure des discontinuités d'une fonction $f \in \mathscr{D}(\mathbb{R}_+)$. Soit f une fonction dans $\mathscr{D}([0,T])$, où T est un temps fini arbitraire. Une subdivision

$$0 = t_0 < t_1 < t_2 < \cdots < t_n = T$$

de $[0,T]$ est dite δ-espacée si $|t_{k+1} - t_k| \geq \delta$ pour tout $k \in [\![0,n-1]\!]$. D'autre part, on définit pour $f : [0,T] \to \mathbb{R}$:

$$\omega(f,[t_k,t_{k+1})) = \sup\left\{|f(s) - f(t)| \mid s,t \in [t_k,t_{k+1})\right\}.$$

Finalement, on pose

$$\omega'(f,\delta) = \inf_{t_0<t_1<\cdots<t_n} \left(\max_{k\in[\![0,n-1]\!]} \omega(f,[t_k,t_{k+1})) \right),$$

où l'infimum porte sur toutes les subdivisions δ-espacées de $[0,T]$. Si l'on doit préciser l'intervalle $[0,T]$, on notera $\omega'_{[0,T]}(f,\delta)$.

Proposition 3.16. *Une fonction f appartient à $\mathscr{D}([0,T])$ si et seulement si*

$$\lim_{\delta\to 0} \omega'_{[0,T]}(f,\delta) = 0.$$

De plus, pour tout $\delta \leq T$ et toute fonction $f \in \mathscr{D}([0,T])$,

$$\omega'\left(f,\frac{\delta}{2}\right) \leq \omega(f,\delta) \leq 2\,\omega'(f,\delta) + \sup_{s\in[0,T]} |f(s) - f(s_-)|.$$

Preuve. Supposons $f \in \mathscr{D}([0,T])$, et fixons $\varepsilon > 0$. Il s'agit de trouver une subdivision $t_0 < t_1 < \cdots < t_n$ de $[0,T]$ telle que $\max_{k\in[\![0,n-1]\!]} \omega(f,[t_k,t_{k+1})) \leq \varepsilon$. Alors, si $\delta = \min_{k\in[\![0,n-1]\!]} |t_{k+1} - t_k|$, ceci impliquera $\omega'(f,\delta) \leq \varepsilon$, et *a fortiori* $\liminf_{\delta\to 0} \omega'(f,\delta) \leq \varepsilon$. Notons

$$T' = \sup\left\{ t \leq T \;\middle|\; \begin{array}{c}\text{il existe une subdivision } 0 = t_0 < t_1 < \cdots < t_n = t \\ \text{telle que } \max_{k\in[\![0,n-1]\!]} \omega(f,[t_k,t_{k+1}))\leq\varepsilon\end{array} \right\}.$$

Remarquons d'abord que $T' > 0$: comme f est continue en 0, ses variations $|f(0) - f(t)|$ sont bornées par $\frac{\varepsilon}{2}$ sur un petit intervalle $[0,T']$, et

$$\omega(f,[0,T')) = \sup\{|f(s) - f(t)| \mid s,t < T'\} \leq 2\sup\{|f(0) - f(t)| \mid t < T'\} \leq \varepsilon.$$

Supposons maintenant par l'absurde $T' < T$. Comme f admet une limite à gauche de T', sur un intervalle $[T'-\delta,T']$ à gauche de ce point, les fluctuations de f sont bornées et $\omega'(f,[T'-\delta,T')) \leq \varepsilon$. Il en va de même sur un intervalle $[T',T'+\delta)$ (quitte à diminuer la valeur de δ), puisque f est continue à droite de T'. D'autre part, par définition de T', il existe une subdivision $0 = t_0 < t_1 < \cdots < t_n = T' - \delta$ telle que $\omega(f,[t_k,t_{k+1})) \leq \varepsilon$ pour tout $k \in [\![0,n-1]\!]$. Alors,

$$t_0 < t_1 < \cdots < t_n = T' - \delta < t_{n+1} = T' < t_{n+2} = T' + \delta$$

est une subdivision de $[0,T'+\delta]$ telle que les fluctuations de f sur chaque intervalle $[t_k,t_{k+1})$ soient inférieures à ε. Ceci contredit la définition de T', donc $T'=T$. On a donc montré que si $f \in \mathscr{D}([0,T])$, alors $\lim_{\delta\to 0}\omega'(f,\delta)=0$.

Réciproquement, supposons $\lim_{\delta\to 0}\omega'(f,\delta)=0$, et montrons que f est continue à droite et avec limites à gauche. Fixons un point $t\in[0,T)$, et $\varepsilon>0$. Par hypothèse, pour δ assez petit, on peut trouver une subdivision δ-espacée $0=t_0<\cdots<t_n=T$ telle que $\omega(f,[t_k,t_{k+1}))\le\varepsilon$ pour tout k. Soit $[t_k,t_{k+1})$ l'intervalle qui contient t. On a $|f(s)-f(t)|\le\varepsilon$ pour tout $s\in[t,t_{k+1})$, donc f est continue à droite de t. Supposons maintenant $t\in(0,T]$, et montrons que f admet une limite à gauche de t. Par le critère de Cauchy, il suffit de montrer qu'il existe un intervalle $(t-\eta,t)$ tel que si s_1,s_2 appartiennent à cet intervalle, alors $|f(s_1)-f(s_2)|\le\varepsilon$. Si la subdivision précédemment choisie vérifie $t_k<t<t_{k+1}$, alors on peut prendre l'intervalle (t_k,t), et la preuve est achevée. Supposons maintenant que $t=t_k$ est l'un des points de la subdivision. Alors, on peut prendre comme intervalle (t_{k-1},t), donc ce cas aussi est traité. On a donc bien démontré que $f\in\mathscr{D}([0,T])$.

Établissons maintenant la suite d'inégalités. Fixons $\delta\le T$, et considérons la subdivision de $[0,T]$ dont tous les pas $t_{k+1}-t_k$ valent $\frac{\delta}{2}$, sauf le dernier pas $T-t_{n-1}$ qui a une taille comprise entre $\frac{\delta}{2}$ et δ.

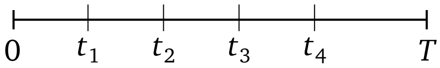


FIGURE 3.4 – Subdivision dont presque tous les pas ont taille $\frac{\delta}{2}$.

C'est une subdivision $\frac{\delta}{2}$-espacée, et sur chaque intervalle $[t_k,t_{k+1})$, les points sont écartés d'au plus δ (dans la dernière part), donc :

$$\omega'\left(f,\frac{\delta}{2}\right)\le\max_{i\in[\![0,n-1]\!]}\omega(f,[t_k,t_{k+1}))\le\omega(f,\delta).$$

Pour la seconde inégalité, on fixe une subdivision δ-espacée $t_0<t_1<\cdots<t_n$ telle que $\max_{k\in[\![0,n-1]\!]}\omega(f,[t_k,t_{k+1}))\le\omega'(f,\delta)+\varepsilon$, avec $\varepsilon>0$ très petit. Si $s<t$ sont deux points de $[0,T]$ écartés de moins que δ, alors :

1. Soit ils tombent dans la même part $[t_k,t_{k+1})$, et dans ce cas
$$|f(s)-f(t)|\le\omega(f,[t_k,t_{k+1}))\le\omega'(f,\delta)+\varepsilon.$$
2. Soit ils tombent dans deux parts consécutives $[t_k,t_{k+1})$ et $[t_{k+1},t_{k+2})$, et dans ce cas,
$$\begin{aligned}|f(s)-f(t)|&\le|f(s)-f((t_{k+1})_-)|+|f((t_{k+1})_-)-f(t_{k+1})|+|f(t_{k+1})-f(t)|\\&\le 2\,\omega'(f,\delta)+2\varepsilon+|f((t_{k+1})_-)-f(t_{k+1})|.\end{aligned}$$

La seconde inégalité de l'énoncé s'en déduit en faisant tendre ε vers 0. □

Corollaire 3.17. *Une fonction $f\in\mathscr{D}([0,T])$ est toujours bornée, et pour tout $\varepsilon>0$, le nombre de discontinuités de f qui sont plus grandes que ε ($|f(t)-f(t_-)|\ge\varepsilon$) est fini. En particulier, f a des discontinuités en quantité dénombrable.*

Preuve. Soit δ tel que $\omega'(f,\delta) \leq 1$. Il existe une subdivision δ-espacée $t_0 < t_1 < \cdots < t_N$ de $[0,T]$ telle que les fluctuations de f sur chaque intervalle $[t_k, t_{k+1})$ soient bornées par 1, ce qui implique en particulier que f soit bornée sur chacun de ces intervalles ; la fonction f est donc bornée sur $[0,T]$. Fixons maintenant $\varepsilon > 0$. Si $\omega'(f,\delta) < \varepsilon$, alors toute subdivision δ-espacée $t_0 < t_1 < \cdots < t_n$ telle que $\max_{k\in[\![0,n-1]\!]} \omega(f,[t_k,t_{k+1})) < \varepsilon$ doit contenir les éléments t tels que $|f(t)-f(t_-)| \geq \varepsilon$. Comme il existe une telle subdivision, on conclut que

$$\mathrm{card}\,\{t \mid |f(t)-f(t_-)| \geq \varepsilon\} \leq n \leq \frac{T}{\delta}.$$

Puis, comme l'ensemble des points de discontinuité de f est l'union dénombrable des ensembles finis $\{t \in [0,T] \,|\, |f(t)-f(t_-)| \geq \frac{1}{k+1}\}$ avec $k \in \mathbb{N}$, cet ensemble est dénombrable. □

Remarque 3.18. La discussion précédente s'étend immédiatement au cas de fonctions $f \in \mathscr{D}(\mathbb{R}_+)$, avec les résultats suivants :

- une fonction $f : \mathbb{R}_+ \to \mathbb{R}$ est dans $\mathscr{D}(\mathbb{R}_+)$ si et seulement si, pour tout $T > 0$, $\lim_{\delta\to 0} \omega'_{[0,T]}(f,\delta) = 0$.
- si $f \in \mathscr{D}(\mathbb{R}_+)$, alors elle est localement bornée et elle a une quantité dénombrable de discontinuités.

▷ **Topologie de Skorohod.** Après ces préliminaires, définissons une topologie sur les espaces $\mathscr{D}(I)$ et $\mathscr{D}(\mathbb{R}_+)$ à l'aide d'une distance. On commencera par le cas des fonctions à horizon fini, c'est-à-dire dans un espace $\mathscr{D}([0,T])$. L'extension des résultats à $\mathscr{D}(\mathbb{R}_+)$ utilise des arguments semblables à ceux qui permettent de passer du cas compact au cas σ-compact pour les espaces de fonctions continues, mais avec des complications liées aux discontinuités des fonctions considérées (voir le dernier paragraphe de cette section).

Définition 3.19 (Distance de Skorohod). *Soit $f,g \in \mathscr{D}([0,T])$, et* $\mathrm{Homeo}_+([0,T])$ *l'ensemble des applications $\psi : [0,T] \to [0,T]$ croissantes, continues et bijectives. La distance de Skorohod entre f et g est définie par*

$$d(f,g) = \inf_{\psi \in \mathrm{Homeo}_+([0,T])} \max(\|\mathrm{id}-\psi\|_\infty, \|f-g\circ\psi\|_\infty).$$

L'idée est que deux fonctions de $\mathscr{D}([0,T])$ sont considérées proches si leurs graphes peuvent être obtenus l'un par rapport à l'autre par de petites modifications (mouvements) selon l'axe des abscisses et selon l'axe des ordonnées. Si l'on n'autorise les modifications du graphe que selon l'axe des ordonnées, ceci revient à regarder la norme uniforme $\|f-g\|_\infty$, et on a vu précédemment qu'elle ne donnait pas une topologie raisonnable sur $\mathscr{D}([0,T])$. Pour autoriser également des modifications selon l'axe des abscisses, on compare en norme uniforme f et $g\circ\psi$, où ψ est un homéomorphisme de $[0,T]$ qui permet de déplacer les points horizontalement ; et on prend en compte ces déplacements avec la norme uniforme $\|\mathrm{id}_{[0,T]}-\psi\|_\infty$.

Proposition 3.20. *L'application d est bien une distance sur $\mathscr{D}([0,T])$, qui en fait un espace métrique séparable.*

Preuve. L'application d est symétrique, car

$$\begin{aligned}
d(g,f) &= \inf_{\psi\in \mathrm{Homeo}_+([0,T])} \max(\|\mathrm{id}-\psi\|_\infty, \|g-f\circ\psi\|_\infty) \\
&= \inf_{\psi\in \mathrm{Homeo}_+([0,T])} \max(\|\mathrm{id}\circ\psi^{-1}-\psi\circ\psi^{-1}\|_\infty, \|g\circ\psi^{-1}-f\circ\psi\circ\psi^{-1}\|_\infty) \\
&= \inf_{\psi\in \mathrm{Homeo}_+([0,T])} \max(\|\mathrm{id}-\psi^{-1}\|_\infty, \|f-g\circ\psi^{-1}\|_\infty) = d(f,g).
\end{aligned}$$

Si $d(f,g)=0$, alors il existe une suite $(\psi_n)_{n\in\mathbb{N}}$ d'homéomorphismes de $[0,T]$ telle que

$$\max(\|\mathrm{id}-\psi_n\|_\infty, \|g-f\circ\psi_n\|_\infty) \to 0.$$

En particulier, $\psi_n \to \mathrm{id}$ en norme infinie. Soit $t\in[0,T]$, et $\varepsilon>0$. Comme f et g sont continues à droite en t, il existe un intervalle $[t,t+\delta)$ tel que pour tout s dans cet intervalle,

$$|f(s)-f(t)|\le\varepsilon \qquad ; \qquad |g(s)-g(t)|\le\varepsilon.$$

Comme $\psi_n\to\mathrm{id}$ en norme infinie, pour n assez grand, $\psi_n(t+\frac{\delta}{2})\in[t,t+\delta)$. Alors,

$$\begin{aligned}
&|g(t)-f(t)| \\
&\le \left|g(t)-g\left(t+\frac{\delta}{2}\right)\right| + \left|g\left(t+\frac{\delta}{2}\right)-f\circ\psi_n\left(t+\frac{\delta}{2}\right)\right| + \left|f\circ\psi_n\left(t+\frac{\delta}{2}\right)-f(t)\right| \\
&\le 2\varepsilon + \|g-f\circ\psi_n\|_\infty
\end{aligned}$$

donc par passage à la limite, $|g(t)-f(t)|\le 2\varepsilon$ pour tout $\varepsilon>0$, ce qui implique $g(t)=f(t)$ et finalement $g=f$. La distance d sépare donc bien les points. Finalement, pour l'inégalité triangulaire, si ψ_1 et ψ_2 sont deux homéomorphismes croissants de $[0,T]$, et si $f,g,h\in\mathscr{D}([0,T])$, alors

$$\begin{aligned}
&\max(\|\mathrm{id}-\psi_1\|_\infty, \|g-f\circ\psi_1\|_\infty) + \max(\|\mathrm{id}-\psi_2\|_\infty, \|h-g\circ\psi_2\|_\infty) \\
&= \max(\|\psi_2-\psi_1\circ\psi_2\|_\infty, \|g\circ\psi_2-f\circ\psi_1\circ\psi_2\|_\infty) + \max(\|\mathrm{id}-\psi_2\|_\infty, \|h-g\circ\psi_2\|_\infty) \\
&\ge \max(\|\mathrm{id}-\psi_2\|_\infty + \|\psi_2-\psi_1\circ\psi_2\|_\infty, \|h-g\circ\psi_2\|_\infty + \|g\circ\psi_2-f\circ\psi_1\circ\psi_2\|_\infty) \\
&\ge \max(\|\mathrm{id}-\psi_1\circ\psi_2\|_\infty, \|h-f\circ\psi_1\circ\psi_2\|_\infty) \ge d(f,h).
\end{aligned}$$

En prenant la borne inférieure sur toutes les paires d'homéomorphismes (ψ_1,ψ_2), on conclut que $d(f,g)+d(g,h)\ge d(f,h)$, donc d est bien une distance.

Pour la séparabilité, on va approcher toute fonction $f\in\mathscr{D}([0,T])$ par une fonction en escalier, puis utiliser des homéomorphismes pour approcher ces fonctions en escalier par des fonctions d'une famille dénombrable. Si $f\in\mathscr{D}([0,T])$, fixons $\varepsilon>0$, et notons $t_1<\cdots<t_{n-1}$ les points de $[0,T]$ où la discontinuité de f est plus grande que ε, avec pour convention $t_0=0$ et $t_n=T$. Par hypothèse, sur chaque intervalle $[t_k,t_{k+1}]$, la plus grande discontinuité de f est plus petite que ε, et d'autre part, $\lim_{\delta\to 0}\omega'_{[t_k,t_{k+1}]}(f,\delta)=0$. L'inégalité

$$\omega_{[t_k,t_{k+1}]}(f,\delta)\le 2\,\omega'_{[t_k,t_{k+1}]}(f,\delta)+\varepsilon$$

montre alors que pour δ assez petit, $\omega_{[t_k,t_{k+1}]}(f,\delta)\le 2\varepsilon$ pour tout $k\in[\![0,n-1]\!]$. On peut donc trouver une sous-subdivision $0=s_0<s_1<\cdots<s_m=T$, dont les pas sont plus petits que δ, et tels que si $a,b\in[s_l,s_{l+1})$, alors $|f(a)-f(b)|\le 2\varepsilon$. La fonction en escalier

$$f_1(t)=f(s_l) \quad \text{si } t\in[s_l,s_{l+1})$$

est alors proche à 2ε près de f en norme infinie. On peut ensuite remplacer les valeurs de f_1 par des nombres rationnels : si

$$f_2(t) = r_l \quad \text{si } t \in [s_l, s_{l+1}), \text{ avec } r_l \in \mathbb{Q} \text{ et } |r_l - f(s_l)| \leq \varepsilon,$$

alors f_2 est une fonction en escalier, à valeurs rationnelles et telle que $\|f_2 - f\|_\infty \leq 3\varepsilon$. Finalement, soit $0 = \widetilde{s_0} < \widetilde{s_1} < \cdots < \widetilde{s_m} = T$ une subdivision de $[0,T]$ telle que tous ses points (sauf éventuellement T) soient rationnels, et telle que $\sup_{l \in [\![0,m-1]\!]} |s_l - \widetilde{s_l}| \leq 3\varepsilon$. Si ψ est l'unique homéomorphisme de $[0,T]$ qui est affine par morceaux et qui envoie $[s_l, s_{l+1}]$ sur $[\widetilde{s_l}, \widetilde{s_{l+1}}]$, alors $\|\psi - \mathrm{id}\|_\infty \leq 3\varepsilon$, et $f_3 = f_2 \circ \psi^{-1}$ est une fonction en escalier, à valeurs rationnelles et dont les sauts ont lieu en des points rationnels. On a

$$d(f, f_3) \leq \max(\|\psi - \mathrm{id}\|_\infty, \|f - f_2\|_\infty) \leq 3\varepsilon,$$

et d'autre part, l'ensemble des fonctions f_3 possibles est dénombrable. Ceci achève la preuve de la séparabilité de $(\mathscr{D}([0,T]), d)$. □

Malheureusement, l'espace métrique $(\mathscr{D}([0,T]), d)$ n'est pas complet (voir les exercices en fin de chapitre). Mais il existe une autre distance d' sur $\mathscr{D}([0,T])$, qui correspond à la même topologie de Skorohod que celle donnée par d, et qui rend l'espace complet. Si ψ est un homéomorphisme croissant de $[0,T]$, posons

$$|||\psi||| = \sup_{0 \leq s < t \leq T} \left| \log\left(\frac{\psi(t) - \psi(s)}{t - s} \right) \right|.$$

Notons que $|||\psi|||$ peut être égal à $+\infty$, par exemple pour un homéomorphisme non lipschitzien. Si $|||\psi||| = 0$, alors la pente de ψ est partout égale à 1, donc $\psi = \mathrm{id}$. Ainsi, la fonction $|||\cdot|||$ fournit une nouvelle façon de mesurer la distance entre un homéomorphisme ψ et la transformation identité. Si $f, g \in \mathscr{D}([0,T])$, on pose :

$$d'(f,g) = \inf_{\psi \in \mathrm{Homeo}_+([0,T])} \max(|||\psi|||, \|f - g \circ \psi\|_\infty).$$

Proposition 3.21. *L'application d' est une distance sur $\mathscr{D}([0,T])$, qui correspond à la même topologie que d.*

Preuve. La symétrie de d' vient des identités $\|g - f \circ \psi\|_\infty = \|g \circ \psi^{-1} - f\|_\infty$ et $|||\psi||| = |||\psi^{-1}|||$. Pour montrer que d' sépare les points dans $\mathscr{D}([0,T])$, considérons deux fonctions f et g telles que $d'(f,g) = 0$. Il existe alors une suite $(\psi_n)_{n \in \mathbb{N}}$ d'homéomorphismes de $[0,T]$ telle que $|||\psi_n||| \to 0$ et $\|g - f \circ \psi_n\|_\infty \to 0$. Si l'on peut montrer que $\psi_n \to \mathrm{id}$ en norme infinie, alors on pourra suivre le même raisonnement que pour la distance d pour établir l'égalité $f = g$. Or, si $|||\psi_n||| \to 0$, alors les pentes des fonctions ψ_n sont uniformément bornées, donc par le théorème d'Arzelà–Ascoli 3.2, $(\psi_n)_{n \in \mathbb{N}}$ est relativement compacte dans $\mathscr{C}([0,T])$ (en fait, c'est le cas de toute famille de fonctions continues croissantes et uniformément bornées, par un argument similaire à celui de la preuve des théorèmes de Dini). De plus, si ψ est une limite d'une sous-suite convergente $(\psi_{n_k})_{k \in \mathbb{N}}$, alors pour tous $s < t$,

$$\log\left| \frac{\psi(t) - \psi(s)}{t - s} \right| = \lim_{k \to \infty} \log\left| \frac{\psi_{n_k}(t) - \psi_{n_k}(s)}{t - s} \right| \leq \liminf_{n \to \infty} |||\psi_n||| = 0.$$

Donc, $\psi = \mathrm{id}$, et l'unicité des limites possibles de sous-suites convergentes prouve que $\psi_n \to \mathrm{id}$ dans $\mathscr{C}([0,T])$. On obtient ainsi l'hypothèse de séparation pour d'. Finalement, si f,g,h sont trois fonctions dans $\mathscr{D}([0,T])$ et si ψ_1,ψ_2 sont deux homéomorphismes croissants, alors

$$\begin{aligned}
&\max(|||\psi_1|||, \|g-f\circ\psi_1\|_\infty) + \max(|||\psi_2|||, \|h-g\circ\psi_2\|_\infty)\\
&= \max(|||\psi_1|||, \|g\circ\psi_2 - f\circ\psi_1\circ\psi_2\|_\infty) + \max(|||\psi_2|||, \|h-g\circ\psi_2\|_\infty)\\
&\geq \max(|||\psi_1|||+|||\psi_2|||, \|h-g\circ\psi_2\|_\infty + \|g\circ\psi_2 - f\circ\psi_1\circ\psi_2\|_\infty)\\
&\geq \max(|||\psi_1\circ\psi_2|||, \|h-f\circ\psi_1\circ\psi_2\|_\infty) \geq d'(f,h),
\end{aligned}$$

d'où $d'(f,g)+d'(g,h) \geq d'(f,h)$ par passage à la borne inférieure sur (ψ_1,ψ_2). La fonction d' est donc bien une distance sur $\mathscr{D}([0,T])$.

Montrons maintenant que d' est topologiquement équivalente à d. Pour tout homéomorphisme croissant ψ,

$$|\psi(t)-t| = t\left|\frac{\psi(t)-\psi(0)}{t-0} - 1\right| \leq T\left(\mathrm{e}^{|||\psi|||}-1\right),$$

donc $\|\psi-\mathrm{id}\|_\infty \leq T(\mathrm{e}^{|||\psi|||}-1)$. Par conséquent, si $d'(f_n,f)$ tend vers 0, alors $d(f_n,f)$ tend également vers 0. Réciproquement, supposons $d(f_n,f)\to 0$, où f est une fonction fixée dans $\mathscr{D}([0,T])$. Il existe alors une suite ψ_n d'homéomorphismes de $[0,T]$ telle que $\|\psi_n-\mathrm{id}\|_\infty \to 0$ et $\|f\circ\psi_n - f_n\|_\infty \to 0$. Dans ce qui suit, on fixe une subdivision $0 = s_0 < s_1 < \cdots < s_m = T$ comme dans la preuve de la proposition 3.20 : si $a,b \in [s_l, s_{l+1})$, alors $|f(a)-f(b)| \leq \varepsilon$. On note alors θ_n l'homéomorphisme de $[0,T]$ qui est affine sur chaque segment $[\psi_n^{-1}(s_l), \psi_n^{-1}(s_{l+1})]$, avec $\theta_n(\psi_n^{-1}(s_l)) = s_l$ pour tout $l \in [\![0,m]\!]$. Comme $\lim_{n\to\infty}\psi_n(s_l) = s_l$ pour tout l,

$$\lim_{n\to\infty} |||\theta_n||| = \lim_{n\to\infty} \sup_{l\in[\![0,m-1]\!]} \left|\log\left(\frac{s_{l+1}-s_l}{\psi_n^{-1}(s_{l+1})-\psi_n^{-1}(s_l)}\right)\right| = 0.$$

Évaluons alors

$$\|f\circ\theta_n - f_n\|_\infty \leq \|f\circ\theta_n - f\circ\psi_n\|_\infty + \|f\circ\psi_n - f_n\|_\infty.$$

Pour prouver que cette suite tend vers 0, comme $\|f\circ\psi_n - f_n\|_\infty \to 0$, il suffit de montrer que $\|f\circ\theta_n\circ\psi_n^{-1} - f\|_\infty$ tend vers 0. Soit $t\in[0,T]$, et $[s_l,s_{l+1}]$ l'intervalle qui contient t. Notons que l'homéomorphisme $\theta_n\circ\psi_n^{-1}$ stabilise chaque intervalle $[s_l,s_{l+1}]$. Par conséquent, t et $\theta_n\circ\psi_n^{-1}(t)$ sont dans le même intervalle de la subdivision, donc

$$\left|f\circ\theta_n\circ\psi_n^{-1}(t) - f(t)\right| \leq \varepsilon.$$

On conclut que $\|f\circ\theta_n\circ\psi_n^{-1} - f\|_\infty \leq \varepsilon$, et donc $d'(f_n,f)\to 0$. □

Théorème 3.22. *L'espace $\mathscr{D}([0,T])$ muni de la topologie de Skorohod est polonais.*

Preuve. On sait déjà que $\mathscr{D}([0,T])$ est métrisable séparable, et on va montrer qu'il est complet pour la distance d'. Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions dans $\mathscr{D}([0,T])$, qui est de Cauchy pour la distance d'. Il suffit de montrer que $(f_n)_{n\in\mathbb{N}}$ a une sous-suite convergente, car une suite de Cauchy et relativement compacte est toujours convergente. Soit $(n_k)_{k\in\mathbb{N}}$ une suite croissante d'entiers telle que

$$\sup_{m_1,m_2\geq n_k} d'(f_{m_1}, f_{m_2}) < \frac{1}{2^k}.$$

Alors, en particulier, $(g_k)_{k\in\mathbb{N}} = (f_{n_k})_{k\in\mathbb{N}}$ est une sous-suite telle que $d'(g_k, g_{k+1}) < \frac{1}{2^k}$ pour tout k, donc il existe une suite d'homéomorphismes $(\psi_k)_{k\in\mathbb{N}}$ telle que

$$|||\psi_k||| \leq \frac{1}{2^k} \qquad ; \qquad \|g_k \circ \psi_k - g_{k+1}\|_\infty \leq \frac{1}{2^k}.$$

Pour $k, m \geq 1$, posons $\theta_{k,m} = \psi_k \circ \psi_{k+1} \circ \cdots \circ \psi_{k+m}$. Nous allons d'abord montrer que $(\theta_{k,m})_{m\geq 1}$ converge uniformément vers un homéomorphisme θ_k de $[0, T]$. On a

$$\begin{aligned}\|\theta_{k,m+1} - \theta_{k,m}\|_\infty &= \sup_{t\in[0,T]} |\theta_{k,m}(\psi_{k+m+1}(t)) - \theta_{k,m}(t)| \\ &\leq \left(\sup_{t\in[0,T]} |\psi_{k+m+1}(t) - t|\right) \mathrm{e}^{|||\theta_{k,m}|||} \leq \|\psi_{k+m+1} - \mathrm{id}\|_\infty \, \mathrm{e}^{|||\theta_{k,m}|||}.\end{aligned}$$

Or, $|||\theta_{k,m}||| \leq |||\psi_k||| + \cdots + |||\psi_{k+m}||| \leq \frac{1}{2^{k-1}}$, et $\|\psi_{k+m+1} - \mathrm{id}\|_\infty \leq T\,(\mathrm{e}^{\frac{1}{2^{k+m+1}}} - 1)$, donc il existe une constante $C = C(T)$ telle que, pour tous $k, m \geq 1$,

$$\|\theta_{k,m+1} - \theta_{k,m}\|_\infty \leq \frac{C}{2^{k+m}}.$$

On en déduit que les suites $(\theta_{k,m})_{m\geq 1}$ sont uniformément de Cauchy dans $\mathscr{C}([0, T])$, et convergent donc uniformément vers des fonctions continues θ_k. Ces fonctions θ_k sont des homéomorphismes de $[0, T]$, car

$$|||\theta_k||| \leq |||\psi_k||| + |||\psi_{k+1}||| + \cdots \leq \frac{1}{2^{k-1}}$$

pour tout $k \geq 1$, et la finitude de $|||\theta_k|||$ implique que la fonction θ_k est un homéomorphisme. D'autre part,

$$\|g_k \circ \theta_k - g_{k+1} \circ \theta_{k+1}\|_\infty = \|g_k \circ \psi_k - g_{k+1}\|_\infty \leq \frac{1}{2^k}$$

pour tout $k \geq 1$. La suite $(g_k \circ \theta_k)_{k\in\mathbb{N}}$ est donc de Cauchy pour la norme uniforme, et elle admet une limite g, qui est aussi la limite de $(g_k)_{k\in\mathbb{N}}$ pour la distance d' :

$$d'(g_k, g) \leq \max(|||\theta_k|||, \|g_k \circ \theta_k - g\|_\infty) \to 0.$$

La limite g appartient bien à $\mathscr{D}([0, T])$, car c'est la limite uniforme de fonctions càdlàg. On a donc bien exhibé une sous-suite convergente de $(f_n)_{n\in\mathbb{N}}$, ce qui démontre la complétude. □

Ainsi, on a équipé l'espace de fonctions $\mathscr{D}([0, T])$ d'une topologie adéquate pour pouvoir y faire des probabilités et y considérer des (suites de) processus.

▷ **Parties compactes et lois fini-dimensionnelles.** Pour conclure l'étude des espaces de trajectoires discontinues $\mathscr{D}([0,T])$, il nous faut comme pour $\mathscr{C}([0,T])$ identifier les parties relativement compactes, et caractériser les lois sur $\mathscr{D}([0,T])$. Concernant les parties relativement compactes, on a un analogue du théorème d'Arzelà–Ascoli :

Théorème 3.23. *Soit $\mathscr{F} \subset \mathscr{D}([0,T])$ une partie de l'espace des fonctions càdlàg sur $[0,T]$. On pose $\omega'(\mathscr{F},\delta) = \sup_{f\in\mathscr{F}} \omega'(f,\delta)$. La partie $\mathscr{F}$ est relativement compacte si et seulement si les deux conditions suivantes sont vérifiées :*

1. *On a $\sup_{f\in\mathscr{F}} \|f\|_{\infty,[0,T]} < +\infty$.*
2. *On a*
$$\lim_{\delta\to 0} \omega'(\mathscr{F},\delta) = 0.$$

Notons que l'on a le même critère pour les parties relativement compactes de $\mathscr{C}([0,T])$, mais avec les fonctions $\omega(\cdot,\delta)$ au lieu de $\omega'(\cdot,\delta)$.

Preuve. Si $\mathscr{F}$ n'est pas bornée pour la norme infinie, alors on peut trouver une suite de fonctions $(f_n)_{n\in\mathbb{N}}$ telle que $\|f_{n+1}\|_\infty \geq \max(\|f_0\|_\infty, \|f_1\|_\infty, \ldots, \|f_n\|_\infty) + 1$ pour tout n. Alors, aucune sous-suite $(f_{n_k})_{k\in\mathbb{N}} = (g_k)_{k\in\mathbb{N}}$ ne peut converger dans $\mathscr{D}([0,T])$. En effet, pour tout k et tout homéomorphisme ψ de $[0,T]$, si $t_{k+1} \in [0,T]$ est un point tel que $|g_{k+1}(t_{k+1})| \geq \|g_k\|_\infty + \frac{1}{2}$, alors

$$\|g_{k+1} - g_k \circ \psi\|_\infty \geq |g_{k+1}(t_{k+1}) - g_k \circ \psi(t_{k+1})| \geq |g_{k+1}(t_{k+1})| - \|g_k\|_\infty \geq \frac{1}{2},$$

donc $d(g_{k+1}, g_k) \geq \frac{1}{2}$. Ceci montre que la première condition est nécessaire à la relative compacité. Pour la seconde condition, raisonnons également par l'absurde et supposons $\lim_{\delta\to 0} \omega'(\mathscr{F},\delta) \neq 0$; alors, on peut trouver une suite de fonctions $(f_n)_{n\in\mathbb{N}}$ dans $\mathscr{F}$, une suite $(\delta_n)_{n\in\mathbb{N}}$ tendant vers 0, et $\varepsilon > 0$ tels que $\omega'(f_n, \delta_n) \geq \varepsilon$ pour tout $n \geq 0$. Sous ces hypothèses, aucune sous-suite $(f_{n_k})_{k\in\mathbb{N}} = (g_k)_{k\in\mathbb{N}}$ ne peut converger dans $\mathscr{D}([0,T])$. En effet, si $g_k \to g$ pour la topologie de Skorohod, fixons $\delta > 0$, et une subdivision δ-espacée $0 = t_0 < t_1 < \cdots < t_n = T$ de $[0,T]$. Pour chaque k, notons ψ_k un homéomorphisme de $[0,T]$ tel que $\|g_k \circ \psi_k - g\|_\infty \leq 2\,d(g_k, g)$, et tel que $\|\psi_k - \mathrm{id}\|_\infty \to 0$. Pour k assez grand, la subdivision $\psi_k(t_0) < \psi_k(t_1) < \cdots < \psi_k(t_n)$ est $\frac{\delta}{2}$-espacée, et donc δ_{n_k}-espacée. Alors,

$$\max_{i\in[\![0,n-1]\!]} \omega(g_k, [\psi_k(t_i), \psi_k(t_{i+1}))) \geq \varepsilon.$$

On peut donc trouver des points x_k et y_k dans le même intervalle $[t_{i_k}, t_{i_k+1})$, et tels que $|g_k(\psi_k(x_k)) - g_k(\psi_k(y_k))| \geq \frac{\varepsilon}{2}$. Supposons k assez grand, tel que $d(g_k, g) \leq \frac{\varepsilon}{16}$. On a alors

$$|g(x_k) - g(y_k)| \geq |g_k(\psi_k(x_k)) - g_k(\psi_k(y_k))| - 2\,\|g_k \circ \psi_k - g\|_\infty \geq \frac{\varepsilon}{2} - 4\,d(g_k, g) \geq \frac{\varepsilon}{4},$$

et donc $\max_{i\in[\![0,n-1]\!]} \omega(g, [t_i, t_{i+1})) \geq \frac{\varepsilon}{4}$. Par conséquent, $\omega'(g,\delta) \geq \frac{\varepsilon}{4}$, et comme ceci est vrai pour tout $\delta > 0$, $g \notin \mathscr{D}([0,T])$.

On sait donc que les deux conditions sont nécessaires à la relative compacité de $\mathscr{F}$. Montrons maintenant qu'elles sont suffisantes. Comme $(\mathscr{D}([0,T]), d')$ est un espace complet, il suffit de montrer qu'une partie $\mathscr{F}$ qui vérifie les deux conditions est précompacte, c'est-à-dire que pour tout $\varepsilon > 0$, elle peut être recouverte par un nombre fini de boules de rayon ε pour d'. Fixons donc $\varepsilon > 0$, et notons $M = \sup_{f\in\mathscr{F}} \|f\|_\infty$. Soit δ suffisamment petit, tel que $\omega'(\mathscr{F},\delta) < \varepsilon$. Pour toute fonction $f \in \mathscr{F}$, il existe une subdivision $t_0 < t_1 < \cdots < t_n$ de $[0,T]$ qui est δ-espacée et telle que $\omega(f,[t_k,t_{k+1})) \leq \varepsilon$ pour tout k. Si $\frac{T}{2^k} \leq \delta$, notons alors $s_0 < s_1 < \cdots < s_n$ la subdivision de $[0,T]$ dont tous les points appartiennent au réseau régulier $\{0, \frac{T}{2^k}, \frac{2T}{2^k}, \ldots, T\}$, et telle que $t_i - \frac{T}{2^k} < s_i \leq t_i$ pour tout i. Si ψ est l'homéomorphisme affine par morceaux qui envoie chaque intervalle $[s_i, s_{i+1}]$ sur $[t_i, t_{i+1}]$, alors

$$\frac{\delta}{\delta + \frac{T}{2^k}} \leq \frac{t_{i+1}-t_i}{s_{i+1}-s_i} \leq \frac{\delta}{\delta - \frac{T}{2^k}}$$

pour tout i, d'où

$$|||\psi||| \leq \frac{T}{2^k\delta - T} \leq \varepsilon \text{ pour } k \text{ assez grand (indépendamment de } f).$$

Considérons alors les fonctions g qui sont continues à droite, prennent leurs valeurs dans un réseau fini de pas inférieur à ε et couvrant $[-M,M]$, et qui ont leurs sauts en des points du réseau régulier $\{0, \frac{T}{2^k}, \frac{2T}{2^k}, \ldots, T\}$. Il y a un nombre fini de telles fonctions, et l'une d'entre elles est à d'-distance de f plus petite que 2ε. En effet, si g est une fonction dont les sauts ont lieu aux points s_i, avec $|g(s_i) - f(t_i)| \leq \varepsilon$, alors

$$\begin{aligned}\|g\circ\psi^{-1} - f\|_\infty &= \max_{i\in[\![0,n-1]\!]} \sup_{s\in[t_i,t_{i+1})} |g\circ\psi^{-1}(s) - f(s)| \\ &\leq \max_{i\in[\![0,n-1]\!]} (\omega(f,[t_i,t_{i+1})) + |g(s_i) - f(t_i)|) \leq 2\varepsilon.\end{aligned}$$

Ceci achève la preuve de la précompacité et donc relative compacité de $\mathscr{F}$. □

Examinons maintenant les propriétés des lois fini-dimensionnelles issues de lois sur $\mathscr{D}([0,T])$. Une différence majeure entre le cas de $\mathscr{D}([0,T])$ et le cas de $\mathscr{C}([0,T])$ est que les applications

$$\begin{aligned}\pi_{t_1<t_2<\cdots<t_d} : \mathscr{D}([0,T]) &\to \mathbb{R}^d \\ f &\mapsto (f(t_1), f(t_2), \ldots, f(t_d))\end{aligned}$$

ne sont pas continues pour la topologie de Skorohod. En effet, si l'on considère par exemple l'application $\pi_{\frac{1}{2}} : f \in \mathscr{D}([0,1]) \mapsto f(\frac{1}{2}) \in \mathbb{R}$, alors elle n'est pas continue en la fonction $f = 1_{[\frac{1}{2},1]}$: les fonctions $f_n = 1_{[\frac{1}{2}+\frac{1}{n},1]}$ convergent vers f dans $\mathscr{D}([0,1])$, mais $\pi_{\frac{1}{2}}(f_n) = 0$ tandis que $\pi_{\frac{1}{2}}(f_n) = 1$. On a néanmoins le résultat plus faible suivant :

Proposition 3.24. *Les applications $\pi_{t_1,t_2,\ldots,t_d}$ sont mesurables pour la topologie de Skorohod. De plus, π_t est continue en $f \in \mathscr{D}([0,T])$ si l'une des conditions suivantes est vérifiée :*

1. *$t = 0$ ou $t = T$.*
2. *$t \in (0,T)$ et f est continue en t : $f(t) = f(t_-)$.*

Preuve. Remarquons tout de suite que les applications π_0 et π_T sont continues sur $\mathscr{D}([0,T])$, car

$$|\pi_0(f)-\pi_0(g)| = \inf_{\psi\in\mathrm{Homeo}_+([0,T])} |f(0)-g\circ\psi(0)| \leq d(f,g)$$

et de même pour π_T (on utilise le fait que tous les homéomorphismes croissants de $[0,T]$ envoient 0 sur 0 et T sur T). Dans les autres cas, pour montrer que π_t est mesurable, on introduit des approximations

$$\pi_{t,\varepsilon} : \mathscr{D}([0,T]) \to \mathbb{R}$$
$$f \mapsto \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f(s)\,ds.$$

Pour tout $f \in \mathscr{D}([0,T])$ et tout $t \in (0,T)$, comme la fonction f est continue à droite de t, $\lim_{\varepsilon\to 0}\pi_{t,\varepsilon}(f) = f(t) = \pi_t(f)$. D'autre part, les applications $\pi_{t,\varepsilon}$ sont continues pour la topologie de Skorohod. En effet, supposons que $f_n \to f$ pour cette topologie, et notons ψ_n une suite d'homéomorphismes tels que $|||\psi_n||| \to 0$ et $\|f_n - f\circ\psi_n\|_\infty \to 0$. Si ψ est un homéomorphisme de $[0,T]$ avec $|||\psi||| < +\infty$, alors en particulier, la mesure image $\psi_*(ds)$ de la mesure de Lebesgue par ψ est absolument continue par rapport à la mesure de Lebesgue, de densité $m_\psi(s)$ encadrée par

$$\mathrm{e}^{-|||\psi|||} \leq m_\psi(s) \leq \mathrm{e}^{|||\psi|||}.$$

On peut alors écrire :

$$\begin{aligned}
\pi_{t,\varepsilon}(f_n) &= \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f_n(s)\,ds = \frac{1}{\varepsilon}\int_{\psi_n(t)}^{\psi_n(t+\varepsilon)} (f_n\circ\psi_n^{-1})(s)\,m_{\psi_n}(s)\,ds \\
&= \frac{1}{\varepsilon}\int_{\psi_n(t)}^{\psi_n(t+\varepsilon)} f(s)\,m_{\psi_n}(s)\,ds + O\left(\frac{|\psi_n(t+\varepsilon)-\psi_n(t)|\,\|f_n - f\circ\psi_n\|_\infty}{\varepsilon}\right) \\
&= \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f(s)\,ds + O\left(\mathrm{e}^{|||\psi_n|||}\,\|f_n - f\circ\psi_n\|_\infty\right).
\end{aligned}$$

Par hypothèse, le reste est un $o(1)$ lorsque n tend vers l'infini, donc $\pi_{t,\varepsilon}$ est continue en f. Alors, π_t s'écrit comme limite ponctuelle de fonctions continues sur $\mathscr{D}([0,T])$, donc elle est mesurable.

Montrons finalement que si f est continue en t, alors l'application π_t est continue en f. Soit $(f_n)_{n\in\mathbb{N}}$ une suite convergeant vers f dans $\mathscr{D}([0,T])$, avec f continue en t. On fixe une suite d'homéomorphismes $(\psi_n)_{n\in\mathbb{N}}$ comme précédemment, et $\varepsilon > 0$. Pour n assez grand, $\psi_n(t)$ est très proche de t, et comme f est continue en t, $|f(t) - f(\psi_n(t))| \leq \varepsilon$. Alors,

$$|\pi_t(f) - \pi_t(f_n)| \leq \varepsilon + \|f_n - f\circ\psi_n\|_\infty,$$

ce qui prouve que $\pi_t(f_n) \to \pi_t(f)$. □

Compte tenu de la non-continuité des applications $\pi_{t_1,\ldots,t_d}$, si $(\mu_n)_{n\in\mathbb{N}}$ est une famille de mesures de probabilité sur $\mathscr{D}([0,T])$, alors $\mu_n \rightharpoonup \mu$ n'implique pas la convergence des lois fini-dimensionnelles $(\mu_n)_{t_1,\ldots,t_d} \rightharpoonup \mu_{t_1,\ldots,t_d}$. On peut néanmoins énoncer un critère de convergence semblable au corollaire 3.13 :

Proposition 3.25. *Soit $(\mu_n)_{n\in\mathbb{N}}$ est une famille de mesures de probabilité sur $\mathscr{D}([0,T])$, et μ une autre loi. On pose*

$$D_\mu = \{t \in [0,T] \mid \mu(\{f(t) \neq f(t_-)\}) > 0\}.$$

1. *L'ensemble D_μ est dénombrable.*
2. *Si $(\mu_n)_{n\in\mathbb{N}}$ est tendue et si $(\mu_n)_{t_1,\ldots,t_d} \rightharpoonup \mu_{t_1,\ldots,t_d}$ pour tous $t_1,\ldots,t_d$ en dehors de D_μ, alors $\mu_n \rightharpoonup \mu$ (et réciproquement).*

Preuve. Soit $\varepsilon,\delta > 0$. S'il existe une suite $(t_n)_{n\in\mathbb{N}}$ telle que $\mu(\{f \mid |f(t_n)-f((t_n)_-)| \geq \varepsilon\}) \geq \delta$, alors, notant $A_n = \{f \mid |f(t_n)-f((t_n)_-)| \geq \varepsilon\}$, on a

$$\mu(\limsup_n A_n) = \lim_{k\to\infty} \mu\left(\bigcup_{n\geq k} A_n\right) \geq \delta,$$

donc

$$\mu(\text{il y a une infinité de points } t_n \text{ tels que } |f(t_n)-f((t_n)_-)| \geq \varepsilon) > 0,$$

ce qui est impossible car les fonctions de $\mathscr{D}([0,T])$ n'ont qu'un nombre fini de discontinuités plus grandes que ε. Donc, pour tout $\varepsilon,\delta > 0$,

$$\{t \in [0,T] \mid \mu(\{f \mid |f(t)-f(t_-)| \geq \varepsilon\}) \geq \delta\}$$

est fini. En prenant des réunions dénombrables sur les $\varepsilon = \frac{1}{n}$ et les $\delta = \frac{1}{k}$, on conclut que l'ensemble

$$D_\mu = \{t \in [0,T] \mid \mu(\{f \mid |f(t)-f(t_-)| > 0\}) > 0\}$$

est dénombrable.

Pour la seconde partie de la proposition, montrons d'abord que la donnée des lois fini-dimensionnelles $\mu_{t_1,\ldots,t_d}$ pour tous temps $t_1,\ldots,t_d$ en dehors d'une partie dénombrable D déterminent entièrement la loi μ. Comme dans la preuve du théorème 3.12, considérons l'application

$$\begin{aligned} i : \mathscr{D}([0,T]) &\to \mathbb{R}^{[0,T]\setminus D} \\ f &\mapsto (f(t))_{t\in[0,T]\setminus D}; \end{aligned}$$

d'après la proposition précédente, cette application est mesurable, donc la tribu produit $\mathscr{P} = i^{-1}(\bigotimes_{t\in[0,T]\setminus D} \mathscr{B}(\mathbb{R}))$ est incluse dans la tribu borélienne $\mathscr{B}(\mathscr{D}([0,T]))$. Réciproquement, notons

$$V_{(f,\varepsilon)} = \{g \in \mathscr{D}([0,T]) \mid d'(f,g) < \varepsilon\};$$

nous allons montrer que toutes ces boules ouvertes sont dans $\mathscr{P}$. On a

$$V_{(f,\varepsilon)} = \bigcup_{\psi \,\mid\, |||\psi||| < \varepsilon} \{g \in \mathscr{D}([0,T]) \mid \|g - f\circ\psi\|_\infty < \varepsilon\}$$

et il n'est pas très difficile de montrer qu'il existe en fait une famille dénombrable $(\psi_n)_{n\in\mathbb{N}}$ d'homéomorphismes de $[0,T]$ telle que $|||\psi_n||| < \varepsilon$ pour tout $n \in \mathbb{N}$, et telle que

$$V_{(f,\varepsilon)} = \bigcup_{n\in\mathbb{N}} \{g \in \mathscr{D}([0,T]) \mid \|g - f\circ\psi_n\|_\infty < \varepsilon\}$$

(on laisse ce point en exercice ; la famille $(\psi_n)_{n\in\mathbb{N}}$ peut dépendre de f). Maintenant, comme les fonctions considérées sont continues à droite en tout point,

$$\begin{aligned}&\{g \in \mathscr{D}([0,T]) \mid \|g - f\circ\psi_n\|_\infty < \varepsilon\}\\ &= \bigcap_{k\in\mathbb{N}}\bigcup_{l\geq 1}\left\{g \in \mathscr{D}([0,T]) \mid |\pi_{t_k}(g) - \pi_{t_k}(f\circ\psi_n)| \leq \left(1-\frac{1}{l}\right)\varepsilon\right\},\end{aligned}$$

où $(t_k)_{k\in\mathbb{N}}$ est une suite dense dans $[0,T]$ et qui évite D. Comme la partie à droite est une intersection dénombrable de réunions dénombrable de parties $\mathscr{P}$-mesurables, la preuve de l'inclusion $\mathscr{B}(\mathscr{D}([0,T])) \subset \mathscr{P}$ est achevée. Maintenant, si μ et ν sont deux lois sur $\mathscr{D}([0,T])$ telles que $\mu_{t_1,\ldots,t_d} = \nu_{t_1,\ldots,t_d}$ pour tous temps en dehors de D, alors elles correspondent sur un π-système qui engendre la tribu $\mathscr{P} = \mathscr{B}(\mathscr{D}([0,T]))$, donc $\mu = \nu$. Démontrons maintenant le second point de la proposition. Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite tendue de mesures sur $\mathscr{D}([0,T])$, telle que $(\mu_n)_{t_1,\ldots,t_d} \rightharpoonup \mu_{t_1,\ldots,t_d}$ pour tous $t_1,\ldots,t_d$ en dehors de D_μ. Si ν est la limite en loi d'une sous-suite convergente $(\mu_{n_k})_{k\in\mathbb{N}}$, alors comme les applications $\pi_{t_1,\ldots,t_d}$ avec $t_1,\ldots,t_d \notin D_\nu$ sont continues presque partout sur le support de ν, on a :

$$\forall t_1 < t_2 < \cdots < t_d \notin D_\nu,\ \ (\mu_{n_k})_{t_1,\ldots,t_d} \rightharpoonup \nu_{t_1,\ldots,t_d}.$$

Par conséquent, si les temps $t_1,\ldots,t_d$ ne sont pas dans $D_\mu \cup D_\nu$, alors $\mu_{t_1,\ldots,t_d} = \nu_{t_1,\ldots,t_d}$. Par la discussion précédente, ceci implique $\mu = \nu$. □

Un cas particulier de la proposition précédente est celui où la limite μ porte sur l'ensemble des fonctions continues $\mathscr{C}([0,T])$. Dans ce cas, $D_\mu = \emptyset$ et on retrouve le même critère de convergence en loi que dans le corollaire 3.13, mais avec des lois μ_n portant sur $\mathscr{D}([0,T])$.

▷ **Trajectoires discontinues à horizon infini.** Pour conclure cette section, expliquons comment étendre la théorie précédente à l'espace des fonctions discontinues $\mathscr{D}(\mathbb{R}_+)$. La différence essentielle entre ce cas et celui des fonctions continues sur $\mathbb{R}_+$ est que les projections

$$\begin{aligned}\mathscr{D}([0,T+U]]) &\to \mathscr{D}([0,T])\\ f &\mapsto f_{|[0,T]}\end{aligned}$$

ne sont pas continues pour les topologies de Skorohod : en effet, dans $\mathscr{D}([0,1])$, les fonctions $f_n = 1_{[\frac{1}{2}+\frac{1}{n},1]}$ convergent vers $f = 1_{[\frac{1}{2},1]}$, mais dans $\mathscr{D}([0,\frac{1}{2}])$, la distance entre $(f_n)_{|[0,\frac{1}{2}]} = 0$ et $f_{|[0,\frac{1}{2}]} = \delta_{\frac{1}{2}}$ est toujours égale à 1. Par conséquent, on ne peut pas directement équiper $\mathscr{D}(\mathbb{R}_+)$ d'une topologie limite projective des topologies des espaces $\mathscr{D}([0,T])$. Pour pallier cette difficulté liée aux discontinuités des fonctions $f \in \mathscr{D}(\mathbb{R}_+)$, on peut utiliser des projections qui forcent la continuité. Plus précisément, si $m \in \mathbb{N}^*$ et si $f \in \mathscr{D}([0,M])$ avec $M > m$, notons

$$\pi_m^M(f)(t) = c_m(t)\,f(t),\quad \text{avec } c_m(t) = \begin{cases}1 & \text{si } t \in [0,m-1],\\ m-t & \text{si } t \in [m-1,m],\\ 0 & \text{si } t \geq m,\end{cases}$$

le résultat $\pi_m^M(f)$ étant considéré comme un élément de $\mathscr{D}([0,m])$.

Proposition 3.26. *Pour tous $m < M$, la projection π_m^M est continue de l'espace $\mathscr{D}([0,M])$ vers $\mathscr{D}([0,m])$. De plus, si $m_1 < m_2 < m_3$, alors $\pi_{m_1}^{m_2} \circ \pi_{m_2}^{m_3} = \pi_{m_1}^{m_3}$.*

Preuve. La compatibilité des projections π_m^M est évidente, car $c_m(t)\,c_M(t) = c_m(t)$ pour tous $m < M$. Montrons maintenant la continuité pour les topologies de Skorohod. Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions dans $\mathscr{D}([0,M])$, qui converge vers une fonction f. Si $\varepsilon > 0$, pour tout $n \geq n_0$, il existe un homéomorphisme ψ_n de $[0,M]$ tel que $\|\psi_n - \mathrm{id}_{[0,M]}\|_\infty \leq \varepsilon$ et $\|f \circ \psi_n - f_n\|_{\infty,[0,M]} \leq \varepsilon$. Notons $\theta_n : [0,\psi_n(m)] \to [0,m]$ l'homéomorphisme affine par morceaux qui est défini comme suit :

— c'est l'identité sur $[0,\psi_n(m-2\varepsilon)]$;

— c'est la fonction affine sur $[\psi_n(m-2\varepsilon),\psi_n(m)]$ qui envoie cet intervalle sur l'intervalle $[\psi_n(m-2\varepsilon), m]$.

On a

$$\|\theta_n - \mathrm{id}\|_\infty \leq |m - \psi_n(m)| \leq \varepsilon,$$

et donc, si $\phi_n = \theta_n \circ \psi_n$, alors ϕ_n est un homéomorphisme croissant de $[0,m]$ tel que

$$\|\phi_n - \mathrm{id}\|_\infty \leq \|\theta_n - \mathrm{id}\|_\infty + \|\psi_n - \mathrm{id}\| \leq 2\varepsilon.$$

Évaluons alors $\Delta_n(t) = |((\pi_m^M(f)) \circ \phi_n)(t) - (\pi_m^M(f_n))(t)|$. Si $t \in [0, m-2\varepsilon]$, alors comme c_m est 1-lipschitzienne,

$$\begin{aligned}\Delta_n(t) &= |(c_m \circ \phi_n)(t)\,(f \circ \psi_n)(t) - c_m(t)\,f_n(t)| \\ &\leq \|f\|_\infty \|\phi_n - \mathrm{id}\|_\infty + |c_m(t)\,(f\circ\psi_n)(t) - c_m(t)\,f_n(t)| \leq (2\|f\|_\infty + 1)\varepsilon.\end{aligned}$$

Supposons maintenant $t \in [m-2\varepsilon, m]$. Alors, $|c_m(t)| \leq 2\varepsilon$ et

$$|c_m \circ \phi_n(t)| \leq m - \psi_n(m-2\varepsilon) \leq 3\varepsilon,$$

donc

$$\begin{aligned}\Delta_n(t) &\leq |(c_m \circ \phi_n)(t)\,(f \circ \phi_n)(t) - c_m(t)\,(f \circ \psi_n)(t)| + |c_m(t)\,(f \circ \psi_n - f_n)(t)| \\ &\leq (5\|f\|_\infty + 1)\varepsilon.\end{aligned}$$

On conclut que $d(\pi_m^M(f), \pi_m^M(f_n)) \leq (5\|f\|_\infty + 1)\varepsilon$ pour n assez grand, d'où la continuité. □

Si $m \in \mathbb{N}^*$, notons $\pi_m : \mathscr{D}(\mathbb{R}_+) \to \mathscr{D}([0,m])$ l'application de multiplication par la fonction de coupure c_m. On a $\pi_m = \pi_m^M \circ \pi_M$ pour tous $M > m$. On munit $\mathscr{D}(\mathbb{R}_+)$ de la distance

$$d(f,g) = \sum_{m=1}^{\infty} \frac{\min(1, d(\pi_m(f), \pi_m(g)))}{2^m},$$

et de la topologie qui en découle.

Théorème 3.27. *L'espace $\mathscr{D}(\mathbb{R}_+)$ est un espace polonais pour la distance d, et une suite de fonctions $(f_n)_{n\in\mathbb{N}}$ converge vers f dans $\mathscr{D}(\mathbb{R}_+)$ si et seulement si toutes les suites de projections $(\pi_m(f_n))_{n\in\mathbb{N}}$ convergent vers $\pi_m(f)$. Une partie $\mathscr{F}$ de $\mathscr{D}(\mathbb{R}_+)$ est relativement compacte si et seulement si toutes les projections $\pi_m(\mathscr{F})$ le sont dans les espaces $\mathscr{D}([0,m])$.*

Preuve. La définition de la distance d implique immédiatement que $f_n \to f$ dans $(\mathscr{D}(\mathbb{R}_+), d)$ si et seulement si toutes les projections $\pi_m(f_n)$ convergent vers $\pi_m(f)$. Montrons que $\mathscr{D}(\mathbb{R}_+)$ est complet. Si $(f_n)_{n\in\mathbb{N}}$ est une suite de Cauchy dans cet espace, alors les suites $(\pi_m(f_n))_{m\in\mathbb{N}}$ sont de Cauchy dans les espaces $\mathscr{D}([0,m])$, donc elles admettent des limites f^m :

$$\forall m \geq 1,\ \ \pi_m(f_n) \to_{n\to\infty} f^m \in \mathscr{D}([0,m]).$$

Comme π_m^M est une application continue et $\pi_m = \pi_m^M \circ \pi_M$, $f^m = \pi_m^M(f^M)$ pour tous $m < M$. Soit $t \in \mathbb{R}_+$. Pour tout $M > m > t$, $f^m(t) = f^M(t)$. On peut donc définir une fonction $f : \mathbb{R}_+ \to \mathbb{R}$ par la formule

$$f(t) = \lim_{m\to\infty} f^m(t).$$

Cette fonction appartient à $\mathscr{D}(\mathbb{R}_+)$, et elle a la propriété que $\pi_m(f) = f^m$ pour tout $m \geq 1$. Alors, $d(f_n, f) \to 0$, puisque c'est vrai pour toutes les projections ; on a donc démontré qu'une suite de Cauchy dans $\mathscr{D}(\mathbb{R}_+)$ admettait toujours une limite.

Pour la séparabilité, introduisons des suites $(f_n^m)_{n\in\mathbb{N}}$ denses dans les espaces $\mathscr{D}([0,m])$ (c'est possible, car ils sont polonais). On prolonge f_n^m en une fonction de $\mathscr{D}(\mathbb{R}_+)$ en posant $f_n^m(t > m) = f_n^m(m)$. Montrons alors que que la famille $(f_n^m)_{n,m\in\mathbb{N}}$ est dense dans $\mathscr{D}(\mathbb{R}_+)$. Soit $f \in \mathscr{D}(\mathbb{R}_+)$, $\varepsilon > 0$ et M tel que $\frac{1}{2^M} \leq \varepsilon$. On peut trouver f_n^M proche de $\pi_M(f)$ dans $\mathscr{D}([0,M])$ à $\frac{\varepsilon}{5\,\|f\|_{\infty,[0,M]}+1}$ près. Alors, pour tout $m < M$, d'après la preuve de la proposition précédente,

$$d(\pi_m(f), \pi_m(f_n^M)) = d(\pi_m^M(\pi_M(f)), \pi_m^M(f_n^M)) \leq (5\,\|\pi_M(f)\|_\infty + 1)\, d(\pi_M(f), f_n^M) \leq \varepsilon.$$

Alors,

$$d(f, f_n^M) \leq \sum_{m=1}^{M-1} \frac{\varepsilon}{2^m} + \sum_{m=M}^{\infty} \frac{1}{2^m} \leq \varepsilon + 2\varepsilon = 3\varepsilon.$$

Examinons finalement la relative compacité. Si $\mathscr{F}$ est relativement compacte dans l'espace $\mathscr{D}(\mathbb{R}_+)$, alors comme les projections π_m sont continues, les parties $\pi_m(\mathscr{F})$ sont relativement compactes dans les espaces $\mathscr{D}([0,m])$. Réciproquement, si ces parties sont relativement compactes, soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions dans $\mathscr{F}$. Chaque suite de projections $(\pi_m(f_n))_{n\in\mathbb{N}}$ est relativement compacte, donc par extraction diagonale, on peut trouver une sous-suite $(f_{n_k})_{k\in\mathbb{N}}$ telle que $(\pi_m(f_{n_k}))_{k\in\mathbb{N}}$ converge pour tout m. D'après la discussion au début de la preuve, ceci implique que $(f_{n_k})_{k\in\mathbb{N}}$ converge, donc $\mathscr{F}$ est bien relativement compacte. □

Ainsi, on peut bien équiper $\mathscr{D}(\mathbb{R}_+)$ d'une topologie d'espace polonais, mais la connection avec les espaces $\mathscr{D}([0,T])$ avec T fini n'est pas donnée par les projections usuelles (si l'on veut que ces projections soient continues).

3.3 Critères de tension

Dans les sections précédentes, on a équipé les espaces $\mathscr{C}(\mathbb{R}_+)$ et $\mathscr{D}(\mathbb{R}_+)$ de topologies d'espaces polonais, et on a identifié les parties relativement compactes de ces espaces. Dans cette section, on utilise ces résultats pour identifier les familles tendues de mesures sur ces espaces. Une première étape simple est la réduction au cas compact :

Proposition 3.28. *Soit $\mathscr{P}$ une famille de mesures de probabilité sur $\mathscr{C}(\mathbb{R}_+)$. La famille est relativement compacte si et seulement si les images $(\phi_m)_*\mathscr{P}$ sont relativement compactes dans les espaces $\mathscr{M}^1(\mathscr{C}([0,m]))$, où ϕ_m est la projection*

$$\begin{aligned}\mathscr{C}(\mathbb{R}_+) &\to \mathscr{C}([0,m])\\ f &\mapsto f_{|[0,m]}.\end{aligned}$$

De même, soit $\mathscr{P}$ une famille de mesures de probabilité sur $\mathscr{D}(\mathbb{R}_+)$. La famille $\mathscr{P}$ est relativement compacte si et seulement si les images $(\pi_m)_*\mathscr{P}$ sont relativement compactes dans les espaces $\mathscr{M}^1(\mathscr{D}([0,m]))$.

Preuve. Traitons par exemple le cas des fonctions continues, l'autre cas étant identique. Si $\mathscr{P}$ est une famille relativement compacte, comme $\phi_m : \mathscr{C}(\mathbb{R}_+) \to \mathscr{C}([0,m])$ est continue, l'application induite $(\phi_m)_* : \mathscr{M}^1(\mathscr{C}(\mathbb{R}_+)) \to \mathscr{M}^1(\mathscr{C}([0,m]))$ est également continue, donc $(\phi_m)_*\mathscr{P}$ est relativement compacte (image par une application continue d'une partie incluse dans un compact). Passons maintenant à la réciproque. Si toutes les parties $(\phi_m)_*\mathscr{P}$ sont relativement compactes, alors elles sont tendues, donc pour tout $\varepsilon > 0$ et tout $m \geq 1$, on peut trouver un compact $K_{m,\varepsilon} \subset \mathscr{C}([0,m])$ tel que $((\phi_m)_*\mu)(K_{m,\varepsilon}) \geq 1-\frac{\varepsilon}{2^m}$ pour tout $\mu \in \mathscr{P}$. Posons

$$K = \bigcap_{m\geq 1}(\phi_m)^{-1}(K_{m,\varepsilon}).$$

On a $\phi_m(K) \subset K_{m,\varepsilon}$ pour tout m, donc, K est une partie relativement compacte de $\mathscr{C}(\mathbb{R}_+)$. De plus, pour tout $\mu \in \mathscr{P}$,

$$\mu(K^c) \leq \sum_{m=1}^{\infty}((\phi_m)_*\mu)(K_{m,\varepsilon}) \leq \sum_{m=1}^{\infty}\frac{\varepsilon}{2^m} = \varepsilon.$$

Donc, $\mathscr{P}$ est une partie tendue, et elle est relativement compacte par le théorème de Prohorov. □

▷ Familles tendues dans $\mathscr{M}^1(\mathscr{C}([0,T]))$. Dans ce qui suit, on donne deux critères de tension simples pour une famille $\mathscr{P}$ de mesures de probabilité sur $\mathscr{C}([0,T])$; comme $\mathscr{C}([0,T])$ est polonais, par le théorème de Prohorov, ces critères caractérisent la relative compacité.

Théorème 3.29. *Une famille $\mathscr{P}$ est tendue dans $\mathscr{M}^1(\mathscr{C}([0,T]))$ si et seulement si :*

$$\lim_{A\to+\infty}\sup_{\mu\in\mathscr{P}}\mu(\{f\,|\,|f(0)|\geq A\}) = 0 \qquad ; \qquad \forall \varepsilon>0,\ \lim_{\delta\to 0}\sup_{\mu\in\mathscr{P}}\mu(\{f\,|\,\omega(f,\delta)\geq\varepsilon\}) = 0.$$

Preuve. Soit $\mathscr{P}$ une famille de mesures qui vérifie les deux hypothèses limites, et $\varepsilon > 0$. On peut trouver A tel que $\mu(\{f\,|\,|f(0)| \leq A\}) \geq 1-\varepsilon$, puis, pour tout $n \geq 1$, δ_n tel que $\mu(\{f\,|\,\omega(f,\delta_n) \leq \frac{1}{n}\}) \geq 1-\frac{\varepsilon}{2^n}$ si $\mu \in \mathscr{P}$. Considérons alors la partie

$$K = \left\{f\,\,|\,\,|f(0)| \leq A \text{ et } \forall n \geq 1,\ \omega(f,\delta_n) \leq \frac{1}{n}\right\}.$$

Pour tout $\mu \in \mathscr{P}$, $\mu(K) \geq 1-2\varepsilon$. De plus, K est une partie relativement compacte de $\mathscr{C}([0,T])$, car elle vérifie bien les deux hypothèses du théorème d'Ascoli 3.2. Donc, $\mathscr{P}$ est tendue.

Réciproquement, supposons $\mathscr{P}$ tendue. Soit $\varepsilon > 0$, et K une partie relativement compacte de $\mathscr{C}([0,T])$ telle que $\inf_{\mu\in\mathscr{P}} \mu(K) \geq 1-\varepsilon$. Par le théorème d'Ascoli, $K \subset \{f \,|\, |f(0)| \leq A\}$ pour une certaine constante A, donc il existe A tel que $\mu(\{f \,|\, |f(0)| \leq A\}) \geq 1-\varepsilon$ pour tout $\mu \in \mathscr{P}$. On en déduit que

$$\liminf_{A\to+\infty} \sup_{\mu\in\mathscr{P}} \mu(\{f \,|\, |f(0)| \geq A\}) \leq \varepsilon,$$

et comme ceci est vrai pour tout $\varepsilon > 0$, $\lim_{A\to+\infty} \sup_{\mu\in\mathscr{P}} \mu(\{f \,|\, |f(0)| \geq A\}) = 0$. La première limite est donc établie, et la seconde provient d'un raisonnement similaire et de la seconde condition du théorème d'Ascoli. □

Remarque 3.30. Le théorème se généralise immédiatement à tout espace de fonctions $\mathscr{C}(T)$ avec T espace compact ; dans ce cas, la première condition du critère de tension s'écrit

$$\lim_{A\to+\infty} \sup_{\mu\in\mathscr{P}} \mu(\{f \,|\, |f(t_0)| \geq A\}) = 0,$$

où t_0 est un point arbitraire de T.

Les conditions du théorème précédent ne sont pas forcément pratiques à vérifier : en effet, il faut calculer les probabilités des ensembles $\{f \,|\, \omega(f,\delta) \leq \varepsilon\}$, ce qui souvent n'est pas aisé. Le critère suivant est seulement suffisant, mais nettement plus pratique :

Théorème 3.31 (Kolmogorov–Chentsov). *Soit $(X^{(n)})_{n\in\mathbb{N}}$ une suite de processus aléatoires dans $\mathscr{C}([0,T])$, qui vérifie les deux conditions suivantes :*

1. *La suite de variables aléatoires $(X^{(n)}(0))_{n\in\mathbb{N}}$ est tendue dans $\mathbb{R}$.*
2. *Il existe des constantes $a \geq 1$, $b > 1$ et $C > 0$ telles que, pour tout $n \in \mathbb{N}$, et tous $s,t \in [0,T]$,*
$$\mathbb{E}\left[|X^{(n)}(s)-X^{(n)}(t)|^a\right] \leq C\,|s-t|^b.$$

Alors, la suite $(X^{(n)})_{n\in\mathbb{N}}$ est tendue.

Preuve. La première partie est équivalente à la première condition du théorème 3.29, donc, il suffit de montrer que pour tout $\varepsilon > 0$, $\lim_{\delta\to 0}\sup_{n\in\mathbb{N}} \mathbb{P}[\omega(X^{(n)},\delta) \geq \varepsilon] = 0$. Sans perte de généralité, on supposera $T = 1$, puisqu'il y a une isométrie évidente entre $\mathscr{C}([0,T])$ et $\mathscr{C}([0,1])$. Posons

$$\xi_k^{(n)} = \max\left\{|X^{(n)}(s)-X^{(n)}(t)| \;\Big|\; s,t \in \left\{0,\frac{1}{2^k},\frac{2}{2^k},\ldots,1\right\},\ |s-t| = \frac{1}{2^k}\right\}.$$

Si $s,t \in [0,1]$ avec $|s-t| \leq \frac{1}{2^l}$, alors on peut utiliser le développement dyadique de ces points pour trouver des suites $(s_k)_{k\geq l}$ et $(t_k)_{k\geq l}$ avec :

$$2^k s_k \in \mathbb{N},\ |s_k - s_{k+1}| \in \left\{0,\frac{1}{2^{k+1}}\right\},\ s_k \to s$$

et de même pour $(t_k)_{k\geq n}$ et t. Alors, par continuité de $X^{(n)}$,

$$|X^{(n)}(s)-X^{(n)}(t)| = \lim_{k\to\infty} |X^{(n)}(s_k)-X^{(n)}(t_k)| \leq |X^{(n)}(s_l)-X^{(n)}(t_l)| + 2\sum_{k\geq l+1} \xi_k^{(n)} \leq 2\sum_{k\geq l} \xi_k^{(n)}.$$

Donc, $\omega(X^{(n)}, \frac{1}{2^l}) \leq 2\sum_{k\geq l} \xi_k^{(n)}$, et par conséquent,

$$\mathbb{E}\left[\left(\omega\left(X^{(n)}, \frac{1}{2^l}\right)\right)^a\right]^{\frac{1}{a}} \leq 2\sum_{k\geq l} \mathbb{E}\left[(\xi_k^{(n)})^a\right]^{\frac{1}{a}}.$$

Or,

$$\mathbb{E}\left[(\xi_k^{(n)})^a\right] \leq \sum_{j=1}^{2^k} \mathbb{E}\left[\left|X^{(n)}\left(\frac{j}{2^k}\right) - X^{(n)}\left(\frac{j-1}{2^k}\right)\right|^a\right] \leq 2^k \times C\,\frac{1}{2^{kb}} = \frac{C}{2^{k(b-1)}}.$$

Il suit :

$$\mathbb{E}\left[\left(\omega\left(X^{(n)}, \frac{1}{2^l}\right)\right)^a\right] \leq K\,2^{-(b-1)(l-1)},$$

et par l'inégalité de Markov,

$$\forall n \in \mathbb{N},\ \mathbb{P}\left[\omega\left(X^{(n)}, \frac{1}{2^l}\right) \geq \varepsilon\right] \leq \frac{K\,2^{-(b-1)(l-1)}}{\varepsilon^a} \to_{l\to\infty} 0. \qquad \square$$

Remarque 3.32. On a utilisé dans la preuve l'inégalité de Hölder–Minkowski pour les normes $\mathscr{L}^a$ avec $a \geq 1$. En utilisant son analogue pour $a \in (0,1]$, on peut étendre le théorème au cas plus général où $a > 0$ (en revanche, il faut toujours $b > 1$). Si l'on regarde des fonctions continues à valeurs dans $\mathbb{R}^d$ au lieu de $\mathbb{R}$, on a le même résultat mais avec l'hypothèse $b > d$. Enfin, on peut montrer que si X est la limite dans $\mathscr{C}([0,T])$ de fonctions vérifiant les hypothèses du théorème de Kolmogorov–Chentsov, alors X est presque sûrement localement Hölderienne d'exposant c pour tout $c < \frac{b-1}{a}$.

▷ Familles tendues dans $\mathscr{M}^1(\mathscr{D}([0,T]))$. L'analogue du théorème d'Arzelà–Ascoli pour les fonctions càdlàg est le théorème 3.23, et il mène à un analogue dans $\mathscr{M}^1(\mathscr{D}([0,T]))$ du critère de tension 3.29 :

Théorème 3.33. *Une famille $\mathscr{P}$ est tendue dans $\mathscr{M}^1(\mathscr{D}([0,T]))$ si et seulement si :*

$$\lim_{A\to+\infty} \sup_{\mu\in\mathscr{P}} \mu(\{f \mid \|f\|_\infty \geq A\}) = 0 \quad ; \quad \forall \varepsilon > 0,\ \lim_{\delta\to 0} \sup_{\mu\in\mathscr{P}} \mu(\{f \mid \omega'(f,\delta) \geq \varepsilon\}) = 0.$$

Preuve. C'est exactement la même preuve que dans le cas des fonctions continues, mais avec la nouvelle caractérisation des parties relativement compactes, qui repose sur les quantités $\|\cdot\|_\infty$ et ω'. $\square$

De nouveau, il existe des critères numériques suffisants qui rendent plus pratique la vérification de la tension. Pour la démonstration du théorème de Donsker dans $\mathscr{D}(\mathbb{R}_+)$, on utilisera le critère suivant :

Théorème 3.34. *Soit X un processus aléatoire dans $\mathscr{D}([0,T])$, qui est presque sûrement continu en T. On suppose que $(X^{(n)})_{n\in\mathbb{N}}$ est une suite de processus aléatoires dans $\mathscr{D}([0,T])$, qui converge en lois fini-dimensionnelles vers X. S'il existe $a>0$, $b>1$ et une fonction continue croissante $F:[0,T]\to\mathbb{R}_+$ telle que*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\Big[|X^{(n)}(t)-X^{(n)}(s)|^a\,|X^{(n)}(s)-X^{(n)}(r)|^a\Big]\leq (F(t)-F(r))^b$$

pour tous $r\leq s\leq t$, alors $(X^{(n)})_{n\in\mathbb{N}}$ converge en loi vers X dans $\mathscr{D}([0,T])$.

On admettra ce critère numérique, dont la preuve est vraiment pénible (voir les références données en fin de chapitre).

3.4 Théorèmes de Donsker

Dans cette dernière section, on démontre que les processus aléatoires $C^{(n)}$ et $D^{(n)}$, qui ont été introduits au début du chapitre et qui appartiennent respectivement à $\mathscr{C}(\mathbb{R}_+)$ et à $\mathscr{D}(\mathbb{R}_+)$, convergent en loi vers le même processus aléatoire continu.

▷ **Lois fini-dimensionnelles.** D'après la discussion des paragraphes précédents, si μ_n est la loi de $C^{(n)}$ dans $\mathscr{C}(\mathbb{R}_+)$ et si ν_n est la loi de $D^{(n)}$ dans $\mathscr{D}(\mathbb{R}_+)$, alors pour démontrer la convergence en loi des processus, il suffit de montrer que :

1. Pour tous temps $t_1<t_2<\cdots<t_d$, les lois fini-dimensionnelles $(\mu_n)_{t_1,\ldots,t_d}$ et $(\nu_n)_{t_1,\ldots,t_d}$ convergent.
2. Les familles de mesures de probabilité $(\mu_n)_{n\in\mathbb{N}}$ et $(\nu_n)_{n\in\mathbb{N}}$ sont tendues.

Le premier point est une simple conséquence du théorème central limite :

Proposition 3.35. *Pour tous temps $t_1<t_2<\cdots<t_d$, les lois fini-dimensionnelles $(\mu_n)_{t_1,\ldots,t_d}$ et $(\nu_n)_{t_1,\ldots,t_d}$ convergent vers la loi d'un vecteur gaussien de matrice de covariance* $\mathrm{cov}(Y_i,Y_j)=\min(t_i,t_j)$.

Preuve. La proposition est équivalente à la convergence en loi des vecteurs

$$\begin{gathered}\big(C^{(n)}(t_1),C^{(n)}(t_2)-C^{(n)}(t_1),\ldots,C^{(n)}(t_d)-C^{(n)}(t_{d-1})\big)\\ \text{et }\big(D^{(n)}(t_1),D^{(n)}(t_2)-D^{(n)}(t_1),\ldots,D^{(n)}(t_d)-D^{(n)}(t_{d-1})\big)\end{gathered}$$

vers un vecteur gaussien $(Z_1,\ldots,Z_d)$ dont les composantes sont indépendantes, et tel que $\mathrm{var}(Z_i)=t_i-t_{i-1}$. Pour chaque temps t_i, notons $n_i=\lfloor nt_i\rfloor$, de sorte que

$$\begin{aligned}C^{(n)}(t_i)-C^{(n)}(t_{i-1})&=\frac{1}{\sqrt{n}}\sum_{k=n_{i-1}+1}^{n_i}X_k+\frac{\{nt_i\}X_{n_i+1}-\{nt_{i-1}\}X_{n_{i-1}+1}}{\sqrt{n}};\\ D^{(n)}(t_i)-D^{(n)}(t_{i-1})&=\frac{1}{\sqrt{n}}\sum_{k=n_{i-1}+1}^{n_i}X_k.\end{aligned}$$

Notons que les suites $\frac{\{nt_i\}X_{n_i+1}-\{nt_{i-1}\}X_{n_{i-1}+1}}{\sqrt{n}}$ convergent toutes en probabilité vers 0, et même en norme $\mathscr{L}^2$: pour n assez grand, $n_i \neq n_{i-1}$ pour tout i, et donc

$$\mathbb{E}\left[\left(\frac{\{nt_i\}X_{n_i+1}-\{nt_{i-1}\}X_{n_{i-1}+1}}{\sqrt{n}}\right)^2\right] = \frac{\{nt_i\}^2+\{nt_{i-1}\}^2}{n}\,\mathbb{E}[(X_1)^2] \leq \frac{2}{n}\,\mathbb{E}[(X_1)^2] \to 0.$$

Par le lemme de Slutsky, il suffit donc de démontrer la convergence en loi des vecteurs $(D^{(n)}(t_i)-D^{(n)}(t_{i-1}))_{i\in[\![1,d]\!]}$; la limite en loi des vecteurs $(C^{(n)}(t_i)-C^{(n)}(t_{i-1}))_{i\in[\![1,d]\!]}$ sera la même. Par le théorème de Lévy, on peut utiliser les transformées de Fourier pour établir la convergence en loi ; fixons donc un vecteur $(\xi_1,\ldots,\xi_d)$ dans $\mathbb{R}^d$. On a :

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\sum_{i=1}^d \xi_i(D^{(n)}(t_i)-D^{(n)}(t_{i-1}))}\right] &= \prod_{i=1}^d \left(\mathrm{e}^{\frac{\mathrm{i}\xi_i}{\sqrt{n}}X_1}\right)^{n_i-n_{i-1}} = \prod_{i=1}^d \left(1-\frac{(\xi_i)^2}{n}+o\left(\frac{1}{n}\right)\right)^{n(t_i-t_{i-1})+O(1)}\\ &\to_{n\to\infty} \prod_{i=1}^d \mathrm{e}^{-\frac{(t_i-t_{i-1})(\xi_i)^2}{2}},\end{aligned}$$

ce que l'on voulait démontrer. □

▷ **Tension de la suite $(C^{(n)})_{n\in\mathbb{N}}$.** Pour démontrer la tension de la suite $(C^{(n)})_{n\in\mathbb{N}}$, on supposera que les variables X_k ont des moments d'ordre 4. On peut alors utiliser le critère de Kolmogorov–Chentsov avec $a=4$ et $b=2$ (il est par ailleurs évident que $(C^{(n)}(0))_{n\in\mathbb{N}}$ est tendue, puisque $C^{(n)}(0)=0$ pour tout n). Calculons donc $\mathbb{E}\left[\left(C^{(n)}(t)-C^{(n)}(s)\right)^4\right]$ avec $s<t$. Si l'on suppose d'abord que $t-s\geq \frac{1}{n}$, alors

$$\begin{aligned}\mathbb{E}\left[\left(C^{(n)}(t)-C^{(n)}(s)\right)^4\right] &= \frac{1}{n^2}\,\mathbb{E}\left[\left(\sum_{k=m+1}^{M} X_k + \{nt\}X_{M+1}-\{ns\}X_{m+1}\right)^4\right]\\ &= \frac{1}{n^2}\left(\sum_{k=m+1}^{M}\mathbb{E}[(X_1)^4]+6\sum_{m+1\leq k<l\leq M}(\mathbb{E}[(X_1)^2])^2+O(n(t-s))\right)\end{aligned}$$

où $m=\lfloor ns\rfloor$ et $M=\lfloor nt\rfloor$, et où le reste $O(n(t-s))$ provient des termes mettant en jeu $\{nt\}X_{M+1}-\{ns\}X_{m+1}$, que l'on n'a pas comptés dans le développement de la puissance 4. On obtient donc :

$$\mathbb{E}\left[\left(C^{(n)}(t)-C^{(n)}(s)\right)^4\right] = 3(t-s)^2+O\left(\frac{t-s}{n}\right) = O((t-s)^2).$$

Supposons maintenant $t-s\leq\frac{1}{n}$. Si $s<t$ appartiennent au même intervalle $[\frac{k}{n},\frac{k+1}{n}]$, alors

$$\mathbb{E}\left[\left(C^{(n)}(t)-C^{(n)}(s)\right)^4\right] = \mathbb{E}\left[\left(\sqrt{n}(t-s)X_{k+1}\right)^4\right] = n^2\,(t-s)^4\,\mathbb{E}[(X_1)^4] \leq \mathbb{E}[(X_1)^4]\,(t-s)^2.$$

De même, si $s\in[\frac{k-1}{n},\frac{k}{n}]$ et $t\in[\frac{k}{n},\frac{k+1}{n}]$, alors

$$\begin{aligned}\mathbb{E}\left[\left(C^{(n)}(t)-C^{(n)}(s)\right)^4\right] &= \frac{1}{n^2}\mathbb{E}\left[((nt-k)X_{k+1}+(k-ns)X_k)^4\right]\\ &= \frac{(nt-k)^4+(k-ns)^4}{n^2}\,\mathbb{E}[(X_1)^4]+\frac{6\,(nt-k)^2\,(k-ns)^2}{n^2}\\ &\leq n^2\,(t-s)^4\,(\mathbb{E}[(X_1)^4]+6) \leq (\mathbb{E}[(X_1)^4]+6)\,(t-s)^2.\end{aligned}$$

Ainsi, on peut bien employer le théorème 3.31. Il démontre la tension dans tout espace de fonctions $\mathscr{C}([0,T])$, et donc dans $\mathscr{C}(\mathbb{R}_+)$. Ainsi :

Théorème 3.36 (Donsker). *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées, telles que $\mathbb{E}[X_1]=0$, $\mathbb{E}[(X_1)^2]=1$ et $\mathbb{E}[(X_1)^4]<+\infty$. On considère les marches aléatoires continues*

$$C^{(n)}(t)=\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{\lfloor nt\rfloor}X_i+(nt-\lfloor nt\rfloor)X_{\lfloor nt\rfloor+1}\right).$$

Lorsque n tend vers l'infini, $(C^{(n)}(t))_{t\in\mathbb{R}_+}$ converge en loi dans $\mathscr{C}(\mathbb{R}_+)$ vers l'unique processus continu gaussien $(B_t)_{t\in\mathbb{R}_+}$ tel que $\mathbb{E}[B_t]=0$ et $\mathrm{cov}(B_s,B_t)=\min(s,t)$ pour tous temps s et t dans $\mathbb{R}_+$. Ce processus est appelé mouvement brownien.

Remarque 3.37. En fait, le résultat reste vrai avec des variables qui ont des moments d'ordre 2, mais la preuve est nettement plus difficile. Elle repose sur une inégalité assez subtile due à Etemadi : si $S_n=\sum_{k=1}^n X_k$ est une somme de variables i.i.d. centrées, alors

$$\mathbb{P}\left[\max_{k\in[\![1,n]\!]}|S_k|\geq 3\lambda\right]\leq 3\max_{k\in[\![1,n]\!]}\mathbb{P}[|S_k|\geq\lambda]$$

pour tout $\lambda>0$. Ceci implique $\lim_{\lambda\to\infty}\limsup_{n\to\infty}\lambda^2\,\mathbb{P}[\max_{k\in[\![1,n]\!]}|S_k|\geq\lambda\sqrt{n}]=0$, et comme les accroissements de $(S_n)_{n\in\mathbb{N}}$ sont indépendants, cette limite permet de contrôler les lois des modules de continuité $\omega(C^{(n)},\delta)$.

▷ **Tension de la suite $(D^{(n)})_{n\in\mathbb{N}}$.** Pour démontrer la tension de la suite $(D^{(n)})_{n\in\mathbb{N}}$ dans $\mathscr{D}([0,T])$, on utilise le théorème 3.34, avec $X=B$ mouvement brownien, qui d'après le paragraphe précédent existe et est continu partout, et qui est la limite en lois fini-dimensionnelles des processus $D^{(n)}$. Il reste donc à calculer

$$\mathbb{E}\big[(D^{(n)}(t)-D^{(n)}(s))^2\,(D^{(n)}(s)-D^{(n)}(r))^2\big]$$

pour $r\leq s\leq t$. Notons que contrairement à précédemment, on n'a pas besoin de l'hypothèse de moments d'ordre 4, car les variables $(D^{(n)}(t)-D^{(n)}(s))^2$ et $(D^{(n)}(s)-D^{(n)}(r))^2$ sont toujours indépendantes. On peut donc factoriser les espérances :

$$\begin{aligned}\mathbb{E}\big[(D^{(n)}(t)-D^{(n)}(s))^2\,(D^{(n)}(s)-D^{(n)}(r))^2\big]&=\mathbb{E}\big[(D^{(n)}(t)-D^{(n)}(s))^2\big]\,\mathbb{E}\big[(D^{(n)}(s)-D^{(n)}(r))^2\big]\\&=\frac{1}{n^2}(\lfloor nt\rfloor-\lfloor ns\rfloor)(\lfloor ns\rfloor-\lfloor nr\rfloor).\end{aligned}$$

Si $|t-s|\geq\frac{1}{n}$ et $|s-r|\geq\frac{1}{n}$, alors l'expression ci-dessus est plus petite que

$$\left(t-s+\frac{1}{n}\right)\left(s-r+\frac{1}{n}\right)\leq 4(t-s)(s-r)\leq 4(t-r)^2.$$

Dans les autres cas, supposons par exemple $|t-s|<\frac{1}{n}$. L'espérance vaut alors 0 sauf si $s<\frac{k}{n}\leq t$ pour un certain k. Dans ce cas, $\lfloor nt\rfloor-\lfloor ns\rfloor=1$, et

$$\mathbb{E}\big[(D^{(n)}(t)-D^{(n)}(s))^2\,(D^{(n)}(s)-D^{(n)}(r))^2\big]\leq\frac{\lfloor ns\rfloor-\lfloor nr\rfloor}{n^2}.$$

Supposons $s - r \geq \frac{1}{n}$. Alors,

$$\mathbb{E}[(D^{(n)}(t) - D^{(n)}(s))^2(D^{(n)}(s) - D^{(n)}(r))^2] \leq \frac{2(s-r)}{n} \leq 2(t-r)^2.$$

Le seul cas qui reste est celui où $|t-s| < \frac{1}{n}$ et $|s-r| < \frac{1}{n}$. Pour que l'espérance soit non nulle, il faut que $r < \frac{k}{n} \leq s < \frac{k+1}{n} \leq t$ pour un certain entier k. Alors,

$$\mathbb{E}\Big[(D^{(n)}(t) - D^{(n)}(s))^2\,(D^{(n)}(s) - D^{(n)}(r))^2\Big] = \frac{1}{n^2} \leq (t-r)^2.$$

Le théorème 3.34 s'applique donc avec $a = 2$, $b = 2$ et la fonction $F(t) = 2t$. On conclut :

Théorème 3.38 (Donsker). *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées, telles que $\mathbb{E}[X_1] = 0$ et $\mathbb{E}[(X_1)^2] = 1$. On considère les marches aléatoires discontinues*

$$D^{(n)}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i.$$

Lorsque n tend vers l'infini, $(D^{(n)}(t))_{t\in\mathbb{R}_+}$ converge en loi dans $\mathscr{D}(\mathbb{R}_+)$ vers le mouvement brownien.

Preuve. Le critère numérique de tension et la convergence des lois fini-dimensionnelles impliquent la convergence en loi dans chaque espace $\mathscr{D}([0,T])$ des restrictions $(D^{(n)}_{|[0,T]})_{n\in\mathbb{N}}$. Or, l'application $\pi_m : \mathscr{D}([0,m+1]) \to \mathscr{D}([0,m])$ est continue, donc les processus aléatoires $\pi_m(D^{(n)})$ convergent en loi dans $\mathscr{D}([0,m])$ vers $\pi_m B$ pour tout $m \in \mathbb{N}$. Ceci implique que $D^{(n)}$ converge en loi vers B dans $\mathscr{D}(\mathbb{R}_+)$. □

Références

Pour la structure de $\mathscr{C}(T)$ avec T espace compact, le théorème de Stone–Weierstrass et celui de Heine, on renvoie à [S. Lang, *Real and Functional Analysis*, Graduate Texts in Mathematics vol. 142, Springer-Verlag, 1993, Chapters II, III]. Le théorème de prolongement de Tietze est également issu de ce même ouvrage. L'extension au cas σ-localement compact est classique, voir la proposition 16.6 dans [O. Kallenberg, *Foundations of Modern Probability*, 2nd edition, Probability and Its Applications, Springer-Verlag, 2001].

La topologie de Skorohod et l'espace de fonctions $\mathscr{D}$ sont introduits et traités en détail dans [P. Billingsley, *Convergence of Probability Measures*, 2nd edition, Wiley, 1999, Chapters 12-16]. En particulier, on renvoie aux théorèmes 13.3 et 13.5 de *loc. cit.* pour une preuve complète de notre théorème 3.34, qui repose sur l'introduction d'un autre module de continuité

$$\omega''(f,\delta) = \sup_{\substack{t-r\leq\delta \\ r\leq s\leq t}} \min(|f(t) - f(s)|, |f(s) - f(r)|).$$

Notre preuve du critère de Kolmogorov–Chentsov est issue de [O. Kallenberg, *Foundations of Modern Probability*, 2nd edition, Probability and Its Applications, Springer-Verlag, 2001, Théorèmes 3.23 et 16.9].

Exercices

1. Soit T un réel positif, et $f : [0,T] \to \mathbb{R}_+$. Montrer que $f \in \mathscr{C}([0,T])$ si et seulement si
$$\lim_{\delta\to 0} \omega(f,\delta) = 0$$
(le cours contient le critère analogue avec ω' pour l'espace $\mathscr{D}([0,T])$).

2. Vérifier qu'une suite de fonctions càdlàg dans $\mathscr{D}([0,T])$ qui converge uniformémement a une limite qui est càdlàg.

3. Montrer que la partie
$$\{\psi \in \mathrm{Homeo}_+([0,T]) \mid |||\psi||| \leq K\}$$
est compacte dans $\mathscr{C}([0,T])$.

4. Soit U une variable de loi uniforme sur $[0,1]$, et $X^{(n)}(t \in [0,1])$ le processus continu défini par
$$X^{(n)}(t) = \min(1, n|U-t|).$$
Montrer que $X^{(n)}$ converge en lois fini-dimensionnelles, mais pas dans $\mathscr{C}([0,1])$.

5. En considérant les suites de fonctions $(1_{[0,1-\frac{1}{n})})_{n\in\mathbb{N}}$ et $(1_{[1+\frac{1}{n},T]})_{n\in\mathbb{N}}$, montrer que l'addition n'est pas une opération continue sur $\mathscr{D}([0,T])$. On n'a donc pas de structure d'espace vectoriel topologique.

6. Dans $\mathscr{D}([0,1])$, on considère la suite de fontions $(f_n)_{n\in\mathbb{N}}$ définie par $f_n = 1_{[0,\frac{1}{2^n})}$. Montrer que la distance $d(f_n, f_{n+1})$ est égale à $\frac{1}{2^{n+1}}$. En déduire que l'espace métrique $(\mathscr{D}([0,1]), d)$ n'est pas complet.

7. On restreint la topologie de Skorohod de $\mathscr{D}([0,T])$ à $\mathscr{C}([0,T])$.
   (a) Montrer que la topologie qu'on obtient sur $\mathscr{C}([0,T])$ est celle de la convergence uniforme.
   (b) Montrer que $\mathscr{C}([0,T])$ est une partie fermée de $\mathscr{D}([0,T])$.
   (c) En déduire que le critère de tension du théorème 3.34 s'applique à des processus aléatoires dans $\mathscr{C}([0,T])$ (au lieu de $\mathscr{D}([0,T])$). Redémontrer avec ce critère le théorème de Donsker pour les suites $(C^{(n)})_{n\in\mathbb{N}}$, avec une hypothèse de moments d'ordre 4.

8. Soit $(U_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes, toutes de loi uniforme sur $[0,1]$. On pose
$$X^{(n)}(t) = \sum_{k=1}^{\lfloor nt\rfloor} 1_{(U_k \leq \frac{1}{n})}.$$

(a) Montrer que les fonctions $X^{(n)}$ sont des variables aléatoires à valeurs dans l'espace $\mathscr{D}([0,1])$.

(b) Montrer que le nombre $\kappa(n)$ de discontinuités de $X^{(n)}$ sur le segment $[0,1]$ suit asymptotiquement une loi de Poisson de paramètre 1 :

$$\forall k \in \mathbb{N}, \quad \lim_{n\to\infty} \mathbb{P}[\kappa(n) = k] = \frac{1}{\mathrm{e}\,k!}.$$

(c) Conditionnellement à l'événement $\kappa(n) = k$, montrer que la loi des temps $t_1 < t_2 < \cdots < t_k$ des sauts de $X^{(n)}$ est donnée par :

$$\text{loi de } (n_1 t, n_2 t, \ldots, n_k t) = \text{loi uniforme sur les parties de taille } k \text{ dans } [\![1,n]\!].$$

En déduire que conditionnellement à l'événement $\kappa(n) = k$,

$$\mathbb{P}[\exists i \,|\, t_i - t_{i-1} \leq \delta] \leq C(k)\,\delta,$$

où $C(k)$ est une constante explicite qui ne dépend que de k (raffinement : montrer qu'on peut prendre $C(k) = k^3$).

(d) Montrer que

$$\lim_{\delta\to 0}\left(\sup_{n\in\mathbb{N}} \mathbb{P}[X^{(n)} \text{ a deux sauts séparés de moins de } \delta]\right) = 0.$$

On pourra calculer ces probabilités en conditionnant par rapport à la valeur de $\kappa(n)$, et en mettant de côté les valeurs trop grandes, qui ont une faible probabilité. En déduire que $(X^{(n)})_{n\in\mathbb{N}}$ est un processus tendu dans $\mathscr{D}([0,1])$.

(e) Montrer que $(X^{(n)})_{n\in\mathbb{N}}$ converge en loi dans $\mathscr{D}([0,1])$, et décrire sa limite.

(f) Redémontrer la tension en utilisant le critère de Kolmogorov pour les fonctions dans $\mathscr{D}([0,1])$.

9. On se propose de démontrer le théorème de Donsker pour les suites

$$C^{(n)}(t) = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{\lfloor nt\rfloor} X_i + (nt - \lfloor nt\rfloor) X_{\lfloor nt\rfloor+1}\right),$$

avec des variables i.i.d. $(X_i)_{i\geq 1}$ qui vérifient seulement les hypothèses $\mathbb{E}[X_i] = 0$ et $\mathbb{E}[(X_i)^2] = 1$.

(a) On pose $S_n = X_1 + \cdots + X_n$, et on introduit le temps d'arrêt $\tau = \inf\{k \,|\, |S_k| \geq 3M\}$, où M est un palier fixé. Montrer que

$$\mathbb{P}\left[\max_{k\in[\![1,n]\!]} |S_k| \geq 3M\right] \leq \mathbb{P}[|S_n| \geq M] + \sum_{k=1}^{n-1} \mathbb{P}[\tau = k]\,\mathbb{P}[|S_n - S_k| > 2M].$$

En déduire l'inégalité d'Etemadi :

$$\mathbb{P}\left[\max_{k\in[\![1,n]\!]} |S_k| \geq 3M\right] \leq 3 \max_{k\in[\![1,n]\!]} \mathbb{P}[|S_k| \geq M].$$

(b) À l'aide du théorème central limite, montrer que

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \left(\lambda^2\, \mathbb{P}\left[\max_{k\in[\![1,n]\!]} |S_k| \geq \lambda\sqrt{n} \right] \right) = 0.$$

(c) Montrer que pour toute subdvision δ-espacée $0 = t_0 < t_1 < \cdots < t_r = 1$ de $[0,T]$,

$$\mathbb{P}[\omega(C^{(n)},\delta) \geq 3\varepsilon] \leq \sum_{i=1}^{r} \mathbb{P}\left[\sup_{u\in[t_{i-1},t_i]} |C^{(n)}(u) - C^{(n)}(t_i)| \geq \varepsilon \right].$$

(d) En choisissant convenablement la subdivision, montrer que la suite $(C^{(n)})_{n\in\mathbb{N}}$ est tendue dans $\mathscr{C}([0,T])$, et conclure.

10. Soit $\phi_T : \mathscr{D}(\mathbb{R}_+) \to \mathscr{D}([0,T])$ l'application qui à une fonction f associe sa restriction sur $[0,T]$. On a vu dans le cours que cette application n'était pas continue.
    (a) Montrer que ϕ_T est tout de même mesurable.
    (b) Montrer que si $f \in \mathscr{D}(\mathbb{R}_+)$ est continue en T, alors ϕ_T est continue au point f.

11. Montrer qu'une suite de fonctions càdlàg $(f_n)_{n\in\mathbb{N}}$ converge dans $\mathscr{D}(\mathbb{R}_+)$ vers une fonction f si et seulement s'il existe une suite d'homéomorphismes croissants $\psi_n : \mathbb{R}_+ \to \mathbb{R}_+$ tels que $\|\psi_n - \mathrm{id}_{\mathbb{R}_+}\|_{\infty,\mathbb{R}_+} \to 0$, et tels que pour tout $m \in \mathbb{N}$,

$$\|f_n \circ \psi_n - f\|_{\infty,[0,m]} \to 0.$$

Chapitre 4

Mesures aléatoires de Poisson

Soit $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$ un espace topologique mesuré, pas forcément de masse finie ; par exemple, $\mathbb{R}$ muni de la mesure de Lebesgue. Si la mesure μ n'est pas de masse finie, il n'est pas possible de donner un sens à l'expression "un point choisi dans $\mathfrak{X}$ suivant la mesure μ". Sous certaines hypothèses, on peut tout de même faire des probabilités par rapport à la mesure μ en considérant une *infinité* de points aléatoires répartis suivant la mesure μ. Ceci mène à la notion de nuage poissonnien de points, ou mesure aléatoire de Poisson.

Dans le premier paragraphe 4.1, on rappelle les propriétés des variables de Poisson, et la théorie des processus de Poisson, qui sont des cas particuliers de processus de Lévy (voir le chapitre 5). Les sauts d'un processus de Poisson forment un nuage poissonnien sur $\mathbb{R}_+$, et dans la section 4.2, on généralise cette construction en définissant pour tout espace polonais mesuré $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$ qui est s-fini un nuage de Poisson sur $\mathfrak{X}$ d'intensité μ. Ces mesures atomiques aléatoires N ont une caractérisation fonctionnelle par la formule de Campbell, et elles sont stables pour de nombreuses opérations listées dans la section 4.3. Finalement, dans le paragraphe 4.4, on étudie des nuages poissoniens sur $\mathfrak{X} \times \mathbb{R}_+$, qui peuvent être réinterprétés comme processus indexés par une variable temporelle, et permettent la construction de processus càdlàg généralisant les processus de Poisson. Ces processus de Poisson composés constitueront l'un des blocs de base dans la construction des processus de Lévy.

4.1 Variables et processus de Poisson

Dans cette section, on rappelle la théorie des variables et processus de Poisson.

▷ **Variables de Poisson.** On fixe dans tout ce qui suit un paramètre $\lambda \in \mathbb{R}_+$.

Définition 4.1 (Loi de Poisson). *Une variable de Poisson de paramètre λ est une variable aléatoire entière X telle que*

$$\forall k \in \mathbb{N},\ \mathbb{P}[X = k] = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}.$$

On note dans ce cas $X \sim \mathscr{P}(\lambda)$.

Remarque 4.2. Si $\lambda = +\infty$, on convient qu'une variable de Poisson $X \sim \mathscr{P}(+\infty)$ vaut $+\infty$ presque sûrement. De même, si $\lambda = 0$, une variable de Poisson $X \sim \mathscr{P}(0)$ vaut 0 presque sûrement.

Les moments factoriels d'une variable de Poisson $X \sim \mathscr{P}(\lambda)$ sont

$$\mathbb{E}[X(X-1)\cdots(X-k+1)] = \sum_{l=k}^{\infty} \mathrm{e}^{-\lambda}\, \frac{\lambda^l}{(l-k)!} = \lambda^k.$$

On en déduit que $\mathbb{E}[X] = \mathrm{var}(X) = \lambda$. D'autre part, la transformée de Fourier d'une variable de Poisson X de paramètre λ est

$$\phi_X(\xi) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X}] = \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda}\, \frac{(\lambda \mathrm{e}^{\mathrm{i}\xi})^k}{k!} = \mathrm{e}^{-\lambda}\, \mathrm{e}^{\lambda \mathrm{e}^{\mathrm{i}\xi}} = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\xi}-1)}.$$

On déduit de ce calcul plusieurs résultats importants :

Proposition 4.3 (Principe de superposition). *Soit $(\lambda_i)_{i\in I}$ une famille sommable de paramètres positifs : $\lambda = \sum_{i\in I} \lambda_i < +\infty$. Étant donnée une famille de variables indépendantes $(X_i)_{i\in I}$ avec $X_i \sim \mathscr{P}(\lambda_i)$, la somme $X = \sum_{i\in I} X_i$ est finie presque sûrement, et elle suit une loi de Poisson de paramètre λ.*

Preuve. La variable X est positive et a son premier moment fini :

$$\mathbb{E}[X] = \sum_{i\in I} \mathbb{E}[X_i] = \sum_{i\in I} \lambda_i = \lambda < +\infty.$$

Elle est donc finie presque sûrement, et on peut calculer (en utilisant le théorème de convergence dominée lorsque I est infini) :

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X}] = \prod_{i\in I} \mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X_i}] = \prod_{i\in I} \mathrm{e}^{\lambda_i(\mathrm{e}^{\mathrm{i}\xi}-1)} = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\xi}-1)}.$$

Donc, X suit une loi de Poisson de paramètre λ. □

Remarque 4.4. Si $\sum_{i\in I} \lambda_i = +\infty$, alors $X = \sum_{i\in I} X_i$ suit encore une loi de Poisson $\mathscr{P}(+\infty)$, c'est-à-dire que $X = +\infty$ presque sûrement. En effet, les événements $\{X_i \geq 1\}$ sont indépendants et ont probabilité

$$\mathbb{P}[X_i \geq 1] = 1 - \mathrm{e}^{-\lambda_i},$$

donc $\sum_{i\in I} \mathbb{P}[X_i \geq 1] = +\infty$, et par Borel–Cantelli, une infinité de variables X_i est ≥ 1. Ainsi, $X = +\infty$ presque sûrement.

Remarque 4.5. La transformée de Laplace $\mathbb{E}[\mathrm{e}^{zX}]$ d'une variable de Poisson $X \sim \mathscr{P}(\lambda)$ est en fait bien définie et convergente sur tout le plan complexe, et égale à $\mathrm{e}^{\lambda(\mathrm{e}^z-1)}$.

Proposition 4.6 (Principe de raréfaction). *Soit $X \sim \mathscr{P}(\lambda)$, et $(\nu_k)_{k\geq 1}$ une suite de variables indépendantes de X, i.i.d. et à valeurs dans $\mathbb{N}$, avec $\mathbb{P}[\nu_k = i] = p_i$. Les variables*

$$X_i = \sum_{k=1}^{X} 1_{\nu_k = i}$$

sont indépendantes, avec $X_i \sim \mathscr{P}(\lambda\, p_i)$.

Preuve. Notons que $X = \sum_{i=0}^{\infty} X_i$. Si l'on montre que pour tout $j \geq 1$, la famille

$$\left(X_0, \ldots, X_j, X - \sum_{i=0}^{j} X_i\right)$$

est constituée de variables de Poisson indépendantes, on aura établi le résultat ; cette remarque permet de se ramener au cas où les p_i forment une mesure de support fini $[\![0, j]\!]$ sur $\mathbb{N}$.

On calcule maintenant facilement, avec $k = k_0 + k_1 + \cdots + k_j$ et $1 = p_0 + p_1 + \cdots + p_j$:

$$\begin{aligned}\mathbb{P}[X_0 = k_0, X_1 = k_1, \ldots, X_j = k_j] &= \mathbb{P}\left[X = k \text{ et } \sum_{l=1}^{k} 1_{v_l=0} = k_0 \text{ et } \cdots \text{ et } \sum_{l=1}^{k} 1_{v_l=j} = k_j\right] \\ &= \mathbb{P}[X = k] \binom{k}{k_0, k_1, \ldots, k_j} (p_0)^{k_0} (p_1)^{k_1} \cdots (p_j)^{k_j} \\ &= \mathrm{e}^{-\lambda} \lambda^k \prod_{i=0}^{j} \frac{(p_i)^{k_i}}{(k_i)!} = \prod_{i=0}^{j} \mathrm{e}^{-\lambda p_i} \frac{(\lambda p_i)^{k_i}}{(k_i)!}.\end{aligned}$$

□

▷ **Processus de Poisson.** La version dépendante en temps des variables de Poisson est le processus càdlàg vérifiant les propriétés suivantes :

Définition 4.7 (Processus de Poisson). *Soit $\lambda \in \mathbb{R}_+^*$. Le processus de Poisson de paramètre λ est le processus aléatoire $(N_t)_{t \in \mathbb{R}_+}$ à valeurs dans $\mathscr{D}(\mathbb{R}_+)$, tel que $N_0 = 0$ et, pour tous temps $t_1 < t_2 < \ldots < t_d$, $(N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_d} - N_{t_{d-1}})$ est une famille de variables de Poisson indépendantes de paramètres $\lambda t_1, \lambda(t_2 - t_1), \ldots, \lambda(t_d - t_{d-1})$. En particulier, $t \mapsto N_t$ est croissant, à valeurs entières et à sauts tous de taille* 1.

L'unicité en loi du processus de Poisson de paramètre λ est claire, car les lois fini-dimensionnelles de $(N_t)_{t \geq 0}$ sont entièrement déterminées, et elles caractérisent une loi sur $\mathscr{D}(\mathbb{R}_+)$. Pour l'existence, on peut procéder comme suit. Soit $(\xi_k)_{k \geq 1}$ une famille de variables indépendantes, suivant toutes une loi exponentielle de paramètre λ : $\mathbb{P}[\xi_k \geq t] = \mathrm{e}^{-\lambda t}$ pour tout $t \geq 0$. On pose alors

$$N_t = \max\left\{k \in \mathbb{N} \,\middle|\, \sum_{j=1}^{k} \xi_j \leq t\right\}.$$

C'est bien défini, car la somme $\sum_{j=1}^{\infty} \xi_j$ est infinie presque sûrement (utiliser par exemple Borel–Cantelli). On obtient ainsi un processus croissant, càdlàg et dont les sauts ont lieu aux instants presque sûrement distincts $T_1 < T_2 < \cdots < T_k < \cdots$, avec

$$T_k = \sum_{j=1}^{k} \xi_j.$$

Déterminons la loi de N_t. La loi de T_k est une loi γ de paramètres (k,λ) (convolée de k lois exponentielles) :

$$T_k \sim 1_{x\geq 0}\,\frac{\lambda^k\,x^{k-1}\,\mathrm{e}^{-\lambda x}}{\Gamma(k)}\,dx.$$

On a alors :

$$\begin{aligned}\mathbb{P}[N_t=k] &= \mathbb{P}[T_k\leq t \text{ et } \xi_{k+1}>t-T_k]\\ &= \iint_{\mathbb{R}_+\times\mathbb{R}_+} 1_{0\leq x\leq t}\,\frac{\lambda^k\,x^{k-1}\,\mathrm{e}^{-\lambda x}}{(k-1)!}\,1_{y>t-x}\,\lambda\,\mathrm{e}^{-\lambda y}\,dx\,dy\\ &= \int_0^t \frac{\lambda^k\,x^{k-1}}{(k-1)!}\,\mathrm{e}^{-\lambda t}\,dx = \mathrm{e}^{-\lambda t}\,\frac{(\lambda t)^k}{k!}.\end{aligned}$$

Chaque variable N_t suit donc une loi de Poisson de paramètre λt. Pour achever la preuve du fait que $(N_t)_{t\geq 0}$ est un processus de Poisson, on va utiliser le principe de raréfaction, et le lemme suivant :

Lemme 4.8. *Dans la construction précédente, conditionnellement à l'événement $N_t = k$, la distribution des instants de saut $T_1 < T_2 < \cdots < T_k$ est uniforme sur le k-simplexe*

$$S^k(t)=\{(t_1,\ldots,t_d)\,|\,0\leq t_1\leq t_2\leq\cdots\leq t_k\leq t\}.$$

Preuve. Soit f une fonction mesurable bornée de $S^d(t)$ vers $\mathbb{R}$. On calcule

$$\begin{aligned}&\mathbb{E}[f(T_1,T_2,\ldots,T_k)\,|\,N_t=k]\\ &= \frac{k!}{(\lambda t)^k}\,\mathrm{e}^{\lambda t}\int_{S^k(t)}\int_{u=t-t_k}^{\infty} f(t_1,t_2,\ldots,t_k)\,\lambda^{k+1}\,\mathrm{e}^{-\lambda t_1}\,\mathrm{e}^{-\lambda(t_2-t_1)}\cdots\mathrm{e}^{-\lambda(t_k-t_{k-1})}\,\mathrm{e}^{-\lambda u}\,dt_1\cdots dt_k\,du\\ &= \frac{k!}{t^k}\int_{S^k(t)} f(t_1,t_2,\ldots,t_k)\,dt_1\cdots dt_k.\end{aligned}$$

□

Concluons maintenant la preuve du fait que $(N_t)_{t\geq 0}$ est un processus de Poisson. Fixons des temps $t_1 < t_2 < \cdots < t_d$, et notons $(U_k)_{k\geq 1}$ une famille de variables indépendantes de loi uniforme sur $[0,t_d]$. D'après le lemme précédent, la loi de l'ensemble aléatoire de points

$$\left\{T_1,T_2,\ldots,T_{N_{t_d}}\right\}$$

est la même que la loi de l'ensemble aléatoire de points

$$\{U_1,U_2,\ldots,U_X\},$$

où X suit une loi de Poisson de paramètre t_d (ces ensembles aléatoires sont des variables aléatoires à valeurs dans $\bigsqcup_{k=0}^{\infty} S^k(T)$). Introduisons alors les variables

$$\nu_k = \text{unique } i\in[\![1,d]\!] \text{ tel que } U_k\in(t_{i-1},t_i].$$

Les $\nu_{k\geq 1}$ sont des variables indépendantes, avec $p_i=\mathbb{P}[\nu_k=i]=(t_i-t_{i-1})/t_d$. De plus, si $X_i=\sum_{k=1}^{X} 1_{\nu_k=i}$, alors $(X_1,X_2,\ldots,X_d)$ a la même loi que $(N_{t_1},N_{t_2}-N_{t_1},\ldots,N_{t_d}-N_{t_{d-1}})$.

Par le principe de raréfaction, on conclut que $(N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_d} - N_{t_{d-1}})$ est un vecteur de variables de Poisson indépendantes de paramètres $\lambda t_1, \lambda(t_2 - t_1), \ldots, \lambda(t_d - t_{d-1})$. On a donc achevé la construction d'un processus de Poisson de paramètre λ, et par construction, ce processus n'a que des sauts de taille 1, qui se produisent à des instants espacés par des variables indépendantes de loi exponentielle $\mathscr{E}(\lambda)$.

Dans la théorie des mesures aléatoires de Poisson, on s'intéressera aux positions des sauts $T_{k\geq 1}$, en oubliant totalement le processus $(N_t)_{t\geq 0}$ (il peut être entièrement reconstruit si l'on connaît la position des sauts). Pour tout intervalle ouvert $(a,b) \subset \mathbb{R}_+$, le nombre de sauts T_k qui tombent dans (a,b) suit une loi de Poisson de paramètre $\lambda(b-a)$. Comme tout ouvert de $\mathbb{R}$ est une réunion dénombrable d'intervalles ouverts disjoints, en utilisant le principe de superposition, on en déduit que pour tout ouvert $U \subset \mathbb{R}_+$, le nombre N_U de points T_k dans U suit une loi de Poisson de paramètre $\lambda \times$ (mesure de Lebesgue de U). De plus, si U et V sont disjoints, alors N_U et N_V sont indépendants. Il n'est pas difficile d'étendre ce résultat aux nombres

$$N_A = \text{nombre de points } T_k \text{ dans } A,$$

avec A partie mesurable quelconque de $\mathscr{B}(\mathbb{R}_+)$. L'objectif de ce chapitre est d'étudier en toute généralité des mesures atomiques aléatoires N sur des espaces topologiques mesurés $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$, telles que pour toute partie mesurable A, $N(A)$ suit une loi de Poisson de paramètre $\mu(A)$, avec indépendance de $N(A)$ et $N(B)$ si A et B sont disjoints.

4.2 Mesures aléatoires de Poisson

▷ **Espace des mesures atomiques.** Un préliminaire à la construction de mesures aléatoires poissonniennes est la définition d'un espace adéquat contenant toutes les mesures atomiques, qu'on choisira plus tard aléatoirement. Dans tout ce qui suit, $\mathfrak{X}$ est un espace topologique polonais, et on note $\mathscr{M}^{+,s}(\mathfrak{X})$ l'ensemble des mesures boréliennes μ positives et s-finies sur $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}))$, c'est-à-dire telles qu'existe une famille dénombrable $(\mu_i)_{i\in I}$ de mesures finies sur $\mathfrak{X}$ avec $\mu = \sum_{i\in I} \mu_i$. Cette hypothèse de s-finitude est légèrement plus faible que l'hypothèse plus classique de σ-finitude, mais elle a l'avantage d'être toujours stable par somme dénombrable de mesures.

Exemple 4.9. La mesure de Lebesgue sur $\mathbb{R}$ est s-finie, puisque c'est la somme de mesures finies

$$dx = \sum_{n\in\mathbb{Z}} (1_{x\in[n,n+1)}\, dx).$$

De même, pour toute famille dénombrable $(x_i)_{i\in I}$ de points de $\mathbb{R}$, et toute famille de poids $(c_i)_{i\in I}$ dans $\mathbb{R}_+ \sqcup \{+\infty\}$, la mesure atomique $\sum_{i\in I} c_i\,\delta_{x_i}$ est s-finie. En effet, il suffit de le montrer pour une mesure $c_i\,\delta_{x_i}$: c'est alors évident si $c_i < +\infty$, et si $c_i = +\infty$, alors $c_i\,\delta_{x_i} = \sum_{j=1}^{\infty} \delta_{x_i}$. Notons que la mesure de Lebesgue est en fait σ-finie, car on peut la décomposer comme somme dénombrable de mesures finies *à supports disjoints* ; ce n'est pas le cas de $\sum_{i\in I} c_i\,\delta_{x_i}$ si l'un des poids c_i est infini.

Remarque 4.10. Si $(\mu_j)_{j\in J}$ est une famille dénombrable de mesures s-finies, alors la somme $\mu = \sum_{j\in J} \mu_j$ est encore une mesure s-finie. Cette stabilité par somme dénombrable n'est plus vraie pour la σ-finitude, et c'est l'une des raisons qui poussent à utiliser l'hypothèse plus confortable de s-finitude.

On équipe $\mathscr{M}^{+,s}(\mathfrak{X})$ de la plus petite tribu qui rend mesurable les applications

$$\mu \mapsto \mu(A) \in \mathbb{R}_+ \sqcup \{+\infty\},$$

avec A parcourant $\mathscr{B}(\mathfrak{X})$. Notons que si f est une fonction mesurable positive sur $\mathfrak{X}$ et si $\mu \in \mathscr{M}^{+,s}(\mathfrak{X})$, alors on peut toujours donner un sens à $\mu(f) = \int_{\mathfrak{X}} f(x)\,\mu(dx)$: c'est la somme finie ou infinie $\sum_{i\in I} \int_{\mathfrak{X}} f(x)\,\mu_i(dx)$ si $\mu = \sum_{i\in I} \mu_i$ est une décomposition de μ en mesures positives finies, et chaque terme $\int_{\mathfrak{X}} f(x)\,\mu_i(dx)$ est par convergence monotone la limite des intégrales finies $\int_{\mathfrak{X}} f(x)\,1_{f(x)\leq m}\,\mu_i(dx)$ lorsque m tend vers l'infini.

Proposition 4.11. *Notons $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$ l'ensemble des sommes dénombrables d'atomes, c'est-à-dire les sommes dénombrables*

$$N = \sum_{i\in I} \delta_{x_i} \text{ avec } I \text{ dénombrable et } (x_i)_{i\in I} \in \mathfrak{X}^I.$$

Si $\mathfrak{X}$ est un espace polonais, alors $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$ est une partie mesurable de $\mathscr{M}^{+,s}(\mathfrak{X})$.

Preuve. La proposition est liée à l'identité :

$$\mathscr{M}^{\mathrm{atom}}(\mathfrak{X}) = \left\{N \in \mathscr{M}^{+,s}(\mathfrak{X}) \mid \forall A \in \mathscr{B}(\mathfrak{X}),\ N(A) \in \mathbb{N} \sqcup \{+\infty\}\right\}.$$

Si $\sum_{i\in I} \delta_{x_i} = N \in \mathscr{M}^{\mathrm{atom}}$, alors $N(A) = \mathrm{card}\,\{i \in I \,|\, x_i \in A\}$ pour toute partie mesurable A, donc $N(A) \in \mathbb{N} \sqcup \{+\infty\}$. Réciproquement, supposons $N(A) \in \mathbb{N} \sqcup \{+\infty\}$ pour toute partie mesurable A. En particulier, pour tout $x \in \mathfrak{X}$, $c_x = N(\{x\})$ est un poids dans $\mathbb{N} \sqcup \{+\infty\}$. De plus, comme N est s-finie, l'ensemble S des points $x \in \mathfrak{X}$ tels que $c_x > 0$ est dénombrable. Posons $M = \sum_{x\in S} c_x\,\delta_x$, et montrons que $M = N$. Notons que $Z = N - M$ est encore une mesure positive s-finie sur $\mathfrak{X}$, telle que $Z(A) \in \mathbb{N} \sqcup \{+\infty\}$ pour toute partie mesurable A. En effet, la partie S est dénombrable dans un espace métrique, donc mesurable, et

$$Z_{|\mathfrak{X}\setminus S} = N_{|\mathfrak{X}\setminus S} = \sum_{i\in I} (N_i)_{|\mathfrak{X}\setminus S}$$

si $N = \sum_{i\in I} N_i$ est une décomposition de N comme somme de mesures finies. La mesure $Z_{|\mathfrak{X}\setminus S}$ est donc s-finie, et d'autre part, par construction, $Z_{|S} = 0$, donc $Z = Z_{|\mathfrak{X}\setminus S}$ est s-finie. De plus, la restriction d'une mesure qui prend des valeurs entières à une partie mesurable est encore uniquement à valeurs entières, donc $Z = N_{|\mathfrak{X}\setminus S}$ est bien uniquement à valeurs entières. Montrons maintenant qu'on a en fait $Z = 0$. Si $Z \neq 0$, alors on peut trouver une partie mesurable A telle que $Z(A) \in [\![1, +\infty]\!]$. Fixons une suite $(x_m)_{m\in\mathbb{N}}$ dense dans $\mathfrak{X}$. Au moins l'une des intersections $A \cap \overline{B}_{(x_m, \frac{1}{2})}$ a Z-mesure non nulle, donc on peut trouver $m_1 \in \mathbb{N}$ tel que $Z(A \cap \overline{B}_{(x_{m_1}, \frac{1}{2})}) \in [\![1, +\infty]\!]$; *a fortiori*, $Z(\overline{B}_{(x_{m_1}, \frac{1}{2})}) \in [\![1, +\infty]\!]$. Puis, on peut trouver

$m_2 \in \mathbb{N}$ tel que $Z(\overline{B}_{(x_{m_1},\frac{1}{2})} \cap \overline{B}_{(x_{m_2},\frac{1}{4})}) \in [\![1,+\infty]\!]$. Par récurrence, on construit une suite $(x_{m_k})_{k\geq 1}$ telle que

$$Z\left(\overline{B}_{(x_{m_1},\frac{1}{2})} \cap \cdots \cap \overline{B}_{(x_{m_k},\frac{1}{2^k})}\right) \in [\![1,+\infty]\!].$$

La suite de fermés

$$F_k = \overline{B}_{(x_{m_1},\frac{1}{2})} \cap \cdots \cap \overline{B}_{(x_{m_k},\frac{1}{2^k})}$$

est décroissante et de diamètre tendant vers 0 ; comme $\mathfrak{X}$ est complet, $F_\infty = \bigcap_{k=1}^{\infty} F_k$ est non vide et restreint à un unique point $x \in \mathfrak{X}$, et on a $Z(\{x\}) \in [\![1,+\infty]\!]$. Or, ceci n'est pas possible, car dans ce cas $N(\{x\}) > M(\{x\})$ et $x \in S$; or, $N_{|S} = M_{|S}$, ce qui est une contradiction.

Remarquons finalement qu'on peut modifier la preuve de l'identité précédente pour montrer qu'on a également la caractérisation suivante :

$$\mathscr{M}^{\mathrm{atom}}(\mathfrak{X}) = \left\{N \in \mathscr{M}^{+,s}(\mathfrak{X}) \mid \forall m,n,\ N(B_{(x_m,\frac{1}{2^n})}) \in \mathbb{N} \sqcup \{+\infty\}\right\}.$$

Comme le terme de droite est une intersection dénombrable de parties mesurables de l'espace $\mathscr{M}^{+,s}(\mathfrak{X})$, $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$ est bien mesurable. □

▷ Construction et caractérisation des mesures aléatoires poissonniennes.

Définition 4.12 (Mesure aléatoire poissonnienne)**.** *Soit $\mathfrak{X}$ un espace topologique polonais, et μ une mesure s-finie sur $\mathfrak{X}$. Une mesure aléatoire poissonnienne, ou nuage de points poissonnien d'intensité μ sur $\mathfrak{X}$ est une variable aléatoire N à valeurs dans $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$, qui vérifie les deux propriétés suivantes :*

1. *Pour toute partie mesurable $A \subset \mathfrak{X}$, $N(A)$ suit une loi de Poisson de paramètre $\mu(A)$.*
2. *Si $(A_i)_{i\in I}$ est une famille de parties mesurables disjointes de $\mathfrak{X}$, alors les variables $(N(A_i))_{i\in I}$ sont indépendantes.*

Théorème 4.13 (Existence des nuages poissonniens)**.** *Pour toute mesure $\mu \in \mathscr{M}^{+,s}(\mathfrak{X})$, il existe une mesure aléatoire poissonnienne N d'intensité μ, et elle est unique en loi dans $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$.*

Lemme 4.14 (Stabilité par somme)**.** *Si $(N_j)_{j\in J}$ est une famille dénombrable de nuages poissonniens indépendants et d'intensités $(\mu_j)_{j\in J}$, alors $N = \sum_{j\in J} N_j$ est un nuage poissonnien d'intensité $\mu = \sum_{j\in J} \mu_j$.*

Preuve. Soit $A \in \mathscr{B}(\mathfrak{X})$. Comme les variables $N_j(A)$ sont des variables de Poisson indépendantes de paramètres $\mu_j(A)$, par le principe de superposition, $N(A)$ suit une loi de Poisson de paramètre $\mu(A) = \sum_{j\in J} \mu_j(A)$. De plus, si $(A_i)_{i\in[\![1,l]\!]}$ est une famille de parties mesurables disjointes de $\mathfrak{X}$, alors

$$\begin{aligned}
\mathbb{P}[N(A_1)=k_1 \text{ et } \cdots \text{ et } N(A_l)=k_l] &= \sum_{\substack{k_1=\sum_{j\in J} k_{1j}\\ k_2=\sum_{j\in J} k_{2j}\\ \vdots \\ k_l=\sum_{j\in J} k_{lj}}} \mathbb{P}[\forall i \in [\![1,l]\!],\ \forall j \in J,\ N_j(A_i)=k_{ij}] \\
&= \prod_{i=1}^{l} \left(\sum_{k_i=\sum_{j\in J} k_{ij}} \prod_{j\in J} \mathbb{P}[N_j(A_i)=k_{ij}] \right) \\
&= \prod_{i=1}^{l} \mathbb{P}[N(A_i)=k_i],
\end{aligned}$$

d'où la condition d'indépendance. □

Démonstration du théorème 4.13. Par le lemme précédent, pour établir l'existence des nuages poissonniens, il suffit de traiter le cas où l'intensité μ est une mesure finie. Soit M une variable de Poisson de paramètre $\mu(\mathfrak{X})$, et $(X_k)_{k\geq 1}$ une famille de variables indépendantes à valeurs dans $\mathfrak{X}$ et de loi $\frac{\mu(\cdot)}{\mu(\mathfrak{X})}$. On pose

$$N = \sum_{k=1}^{M} \delta_{X_k}.$$

Vérifions que N est une mesure poissonnienne d'intensité μ. Soient $A_1,\ldots,A_l$ des parties mesurables disjointes de $\mathfrak{X}$, et $B = \mathfrak{X} \setminus \bigsqcup_{i=1}^{l} A_i$ leur complémentaire. On note

$$\nu_k = \begin{cases} i & \text{si } X_k \in A_i \\ 0 & \text{si } X_k \in B. \end{cases}$$

Les variables $(\nu_k)_{k\geq 1}$ sont indépendantes et de même loi, donc, par le principe de raréfaction, les variables

$$N(A_i) = \sum_{k=1}^{M} 1_{X_k \in A_i} = \sum_{k=1}^{M} 1_{\nu_k = i}$$

sont indépendantes et suivent des lois de Poisson de paramètres $\mu(\mathfrak{X}) \times \mathbb{P}[\nu_k = i] = \mu(A_i)$. Ainsi, N est bien un nuage poissonnien d'intensité μ sur $\mathfrak{X}$.

Pour l'unicité, considérons l'ensemble des parties mesurables de $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$ qui s'écrivent :

$$\left\{N \in \mathscr{M}^{\mathrm{atom}} \mid N(A_1)=k_1 \text{ et } N(A_2)=k_2 \text{ et } \cdots \text{ et } N(A_l)=k_l\right\}$$

où les A_i sont des parties mesurables et les k_i dans $\mathbb{N} \sqcup \{+\infty\}$. C'est un ensemble stable par intersections et qui engendre la tribu de $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$, donc, par le lemme de classe monotone, si deux mesures de probabilités sont égales sur ces parties, alors elles sont globalement égales. Or, toute partie de ce type peut se réécrire sous la forme

$$\left\{N \in \mathscr{M}^{\mathrm{atom}} \mid N(B_1)=j_1 \text{ et } N(B_2)=j_2 \text{ et } \cdots \text{ et } N(B_l)=j_l\right\}$$

avec des B_i mesurables et disjoints ; et la probabilité d'une telle partie pour une mesure aléatoire poissonnienne d'intensité μ est égale à $\prod_{i=1}^{l} \frac{(\mu(B_i))^{j_i}}{(j_i)!} \,\mathrm{e}^{-\mu(B_i)}$. Donc, la loi d'un nuage de points d'intensité μ est uniquement déterminée. □

Proposition 4.15 (Formule de Campbell, cas positif)**.** *Soit N une mesure aléatoire de Poisson d'intensité μ, et $f : \mathfrak{X} \to \mathbb{R}_+$ une fonction mesurable positive. La variable aléatoire positive $N(f) = \int_{x\in\mathfrak{X}} f(x)\,N(dx)$ a pour espérance $\mathbb{E}[N(f)] = \mu(f) = \int_{x\in\mathfrak{X}} f(x)\,\mu(dx)$. De plus,*

$$\mathbb{E}\left[\mathrm{e}^{-N(f)}\right] = \exp\left(\int_{x\in\mathfrak{X}} (\mathrm{e}^{-f(x)} - 1)\,\mu(dx)\right).$$

Preuve. On peut écrire f comme limite ponctuelle croissante de fonctions en escalier $f_n = \sum_{i\in I_n} c_i\, 1_{A_i}$, et pour chacune de ces fonctions,

$$N(f_n) = \sum_{i\in I_n} c_i\, N(A_i)$$

est une combinaison linéaire de variables de Poisson indépendantes, d'espérance

$$\mathbb{E}[N(f_n)] = \sum_{i\in I_n} c_i\, \mathbb{E}[N(A_i)] = \sum_{i\in I_n} c_i\, \mu(A_i) = \mu(f_n).$$

Si $(\Omega, \mathscr{F}, \mathbb{P})$ est l'espace de probabilité sous-jacent à la mesure aléatoire $N = N_\omega$, lorsque n tend vers l'infini, pour tout $\omega \in \Omega$, par le théorème de convergence monotone, $N_\omega(f) = \lim_{n\to\infty} N_\omega(f_n)$. En particulier, $N(f)$ est bien une variable aléatoire positive définie sur $(\Omega, \mathscr{F}, \mathbb{P})$, en tant que limite ponctuelle de variables aléatoires. En utilisant de nouveau le théorème de convergence monotone, on obtient

$$\mathbb{E}[N(f)] = \lim_{n\to\infty} \mathbb{E}[N(f_n)] = \lim_{n\to\infty} \mu(f_n) = \mu(f).$$

Calculons maintenant la transformée de Laplace. Pour chaque fonction f_n,

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{-N(f_n)}\right] &= \prod_{i\in I_n} \mathbb{E}\left[\mathrm{e}^{-c_i\, N(A_i)}\right] = \prod_{i\in I_n} \mathrm{e}^{\mu(A_i)(\mathrm{e}^{-c_i}-1)} \\ &= \exp\left(\sum_{i\in I_n} \mu(A_i)\,(\mathrm{e}^{-c_i} - 1)\right) = \exp\left(\int_{x\in\mathfrak{X}} (\mathrm{e}^{-f_n(x)} - 1)\,\mu(dx)\right).\end{aligned}$$

Par convergence monotone appliquée aux suites décroissantes de fonctions $\mathrm{e}^{-N_\omega(f_n)} \searrow \mathrm{e}^{-N_\omega(f)}$ et $(\mathrm{e}^{-f_n(x)} - 1) \searrow (\mathrm{e}^{-f(x)} - 1)$, on obtient bien

$$\mathbb{E}\left[\mathrm{e}^{-N(f)}\right] = \exp\left(\int_{x\in\mathfrak{X}} (\mathrm{e}^{-f(x)} - 1)\,\mu(dx)\right).$$ □

Proposition 4.16 (Formule de Campbell, cas intégrable)**.** *Soit N une mesure aléatoire de Poisson d'intensité μ, et $f \in \mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$. La fonction f appartient à $\mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), N)$ presque sûrement, et la variable aléatoire $N(f) = \int_{x\in\mathfrak{X}} f(x)\,N(dx)$ a pour espérance $\mathbb{E}[N(f)] = \mu(f)$. De plus,*

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}N(f)}\right] = \exp\left(\int_{x\in\mathfrak{X}} (\mathrm{e}^{\mathrm{i}f(x)} - 1)\,\mu(dx)\right).$$

Preuve. En utilisant le cas positif, comme $|N(f)| \leq N(|f|)$,

$$\mathbb{E}[|N(f)|] \leq \mathbb{E}[N(|f|)] = \mu(|f|) < +\infty.$$

Par conséquent, f est intégrable contre N presque sûrement, et $N(f)$ est intégrable. En décomposant f comme différence $f_+ - f_-$ de deux fonctions positives, on montre ensuite en utilisant le cas continu que $\mathbb{E}[N(f)] = \mu(f)$.

Pour la formule donnant la transformée de Fourier, notons qu'elle est vraie pour les fonctions $f = \sum_{i\in I} c_i\, 1_{A_i}$ en escalier. Ces fonctions sont denses dans $\mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$, et si f est une fonction μ-intégrable générale, on peut trouver une suite de fonctions étagées $(f_n)_{n\in\mathbb{N}}$ telle que $f_n \to f$ en norme $\mathscr{L}^1$ et μ-presque partout. On a alors $(N(f_n))_{n\in\mathbb{N}}$ qui converge en norme dans $\mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ vers $N(f)$, car

$$\mathbb{E}[|N(f) - N(f_n)|] \leq \mathbb{E}[N(|f - f_n|)] = \mu(|f - f_n|) \to 0.$$

Quitte à extraire une sous-suite, on peut aussi supposer $N(f_n) \to_{\text{p.s.}} N(f)$. Alors, par convergence dominée,

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}N(f)}\right] = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{e}^{\mathrm{i}N(f_n)}\right] = \exp\left(\lim_{n\to\infty} \int_{\mathfrak{X}} (\mathrm{e}^{\mathrm{i}f_n(x)} - 1)\,\mu(dx)\right).$$

Or,

$$\left|\int_{\mathfrak{X}} (\mathrm{e}^{\mathrm{i}f(x)} - 1)\,\mu(dx) - \int_{\mathfrak{X}} (\mathrm{e}^{\mathrm{i}f_n(x)} - 1)\,\mu(dx)\right| \leq \int_{\mathfrak{X}} \left|\mathrm{e}^{\mathrm{i}f(x)} - \mathrm{e}^{\mathrm{i}f_n(x)}\right|\,\mu(dx) \leq \mu(|f - f_n|) \to 0.$$

La formule de Campbell est donc également établie dans le cas intégrable. □

Remarque 4.17. La formule de Campbell fournit en fait une caractérisation des mesures aléatoires de Poisson. En effet, si N est une variable aléatoire à valeurs dans $\mathscr{M}^{\text{atom}}(\mathfrak{X})$ qui vérifie la formule de Campbell pour toute fonction f mesurable positive, alors la transformée de Laplace du vecteur $(N(A_1), \ldots, N(A_k))$ avec $A_1, \ldots, A_k$ parties disjointes est entièrement déterminée, et est celle d'un vecteur de variables de Poisson indépendantes ; donc, N est un nuage poissonnien de points.

▷ **Existence de points multiples.** Étant donnée une mesure d'intensité μ qui est s-finie, la mesure aléatoire poissonnienne N d'intensité μ est une somme de Dirac $\sum_{i\in I} \delta_{x_i}$, avec des points aléatoires x_i qui ne sont pas forcément tous distincts. On néanmoins a un critère simple sur μ pour que presque sûrement tous les atomes de N soient distincts :

Proposition 4.18. *Soit $\mu \in \mathscr{M}^{+,s}(\mathfrak{X})$ une mesure sans atome, c'est-à-dire telle que $\mu(\{x\}) = 0$ pour tout singleton x. Presque sûrement, la mesure poissonnienne N d'intensité μ a ses atomes qui sont tous distincts :*

$$\mathbb{P}[\forall x \in \mathfrak{X},\ N(\{x\}) \in \{0, 1\}] = 1.$$

Preuve. Supposons d'abord μ de masse finie. Alors, la mesure poissonnienne N peut être construite en tirant au hasard M de loi $\mathscr{P}(\lambda)$, puis M points indépendants $X_1, \ldots, X_M$ dans $\mathfrak{X}$ de loi $\frac{\mu(\cdot)}{\mu(\mathfrak{X})}$. Or, étant donnée une suite $(X_n)_{n\in\mathbb{N}}$ de telles variables indépendantes,

$$\mathbb{P}[X_i = X_j] = \frac{1}{(\mu(\mathfrak{X}))^2} \int_{\mathfrak{X}^2} 1_{x=y}\, \mu(dx)\, \mu(dy) = \frac{1}{(\mu(\mathfrak{X}))^2} \int_{\mathfrak{X}} 0\; \mu(dx) = 0,$$

donc tous les points de la suite sont distincts avec probabilité 1. Par conséquent, $N = \sum_{i=1}^{M} \delta_{X_i}$ a avec probabilité 1 tous ses atomes distincts.

Traitons maintenant le cas général. On décompose μ comme somme de mesures finies $\mu = \sum_{i\in I} \mu_i$, et on note $N = \sum_{i\in I} N_i$, où les N_i sont des nuages poissonniens indépendants de loi μ_i. On a vu que N était un nuage poissonnien d'intensité μ. Pour montrer que N n'a pas de point multiple, il suffit de montrer que si N_i et N_j sont des nuages poissonniens indépendants associés à des mesures finies sans atomes μ_i et μ_j, alors presque sûrement, N_i et N_j n'ont pas de point en commun. Or, c'est une conséquence du cas fini, car $N_i + N_j$ est un nuage poissonnien d'intensité $\mu_i + \mu_j$, donc sans point multiple d'après ce qui précède. $\square$

Exemple 4.19. La mesure aléatoire poissonnienne sur $\mathbb{R}^d$ d'intensité la mesure de Lebesgue a presque sûrement tous ses points distincts.

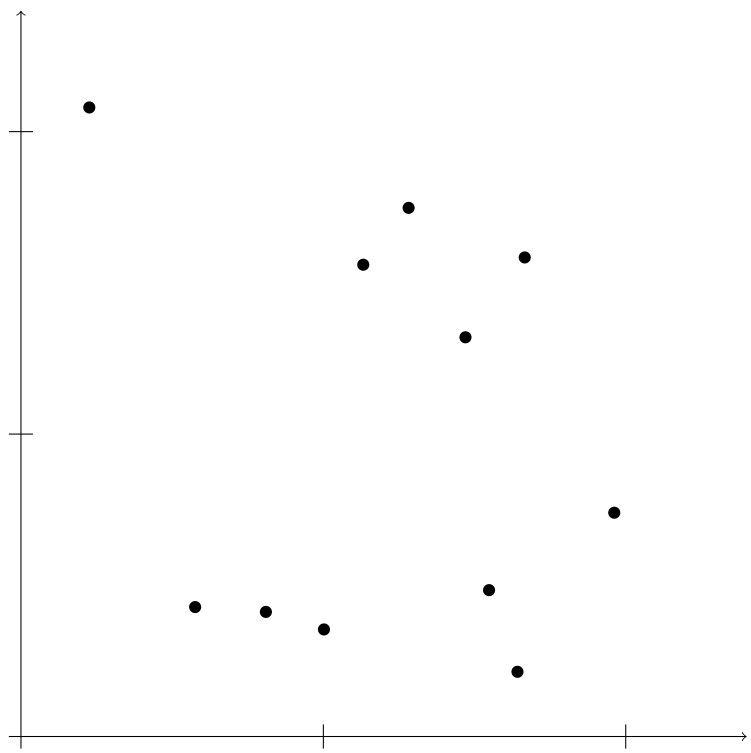


FIGURE 4.1 – Nuage poissonnien de points sur $\mathbb{R}^2$.

4.3 Transformations de nuages de Poisson

Dans cette section, on liste les transformations possibles des mesures aléatoires poissonniennes, qui conservent ce caractère. Un premier résultat simple est la conservation par image :

Proposition 4.20 (Stabilité par image)**.** *Soit μ une mesure s-finie sur un espace polonais $\mathfrak{X}$, et $f : \mathfrak{X} \to \mathfrak{Y}$ une application mesurable. Si N est un nuage poissonnien sur $\mathfrak{X}$ d'intensité μ, alors $f_*(N)$ est un nuage poissonnien sur $\mathfrak{Y}$ d'intensité $f_*(\mu)$.*

Preuve. Si $f : \mathfrak{X} \to \mathfrak{Y}$ est une application mesurable, alors $f_* : \mathscr{M}^{\mathrm{atom}}(\mathfrak{X}) \to \mathscr{M}^{\mathrm{atom}}(\mathfrak{Y})$ est une application mesurable, car si $B \in \mathscr{B}(\mathfrak{Y})$ et $k \in \mathbb{N}$, alors

$$\begin{aligned}(f_*)^{-1}(\{\nu \in \mathscr{M}^{\mathrm{atom}}(\mathfrak{Y}) \,|\, \nu(B) = k\}) &= \{\mu \in \mathscr{M}^{\mathrm{atom}}(\mathfrak{X}) \,|\, \mu(f^{-1}(B)) = k\}) \\ &= \{\mu \in \mathscr{M}^{\mathrm{atom}}(\mathfrak{X}) \,|\, \mu(A) = k\}\end{aligned}$$

avec $A = f^{-1}(B)$ dans $\mathscr{B}(\mathfrak{X})$. Le terme de droite est bien mesurable dans $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$, ce qui prouve que l'image réciproque par f_* d'une partie mesurable de $\mathscr{M}^{\mathrm{atom}}(\mathfrak{Y})$ est une partie mesurable de $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$.

Soit N nuage de Poisson d'intensité μ. Par la discussion précédente, la mesure image $f_*(N)$ est bien une variable aléatoire dans $\mathscr{M}^{\mathrm{atom}}(\mathfrak{Y})$, et si $B \in \mathscr{B}(\mathfrak{Y})$, alors

$$f_*(N)(B) = N(f^{-1}(B)) \sim \mathscr{P}(\mu(f^{-1}(B))) = \mathscr{P}(f_*(\mu)(B)).$$

De plus, si $B_1, \ldots, B_k$ sont des parties mesurables disjointes de $\mathfrak{Y}$, alors les parties $A_1 = f^{-1}(B_1), \ldots, A_k = f^{-1}(B_k)$ sont disjointes dans $\mathfrak{X}$, et donc, les variables

$$f_*(N)(B_i) = N(A_i)$$

sont indépendantes. On a donc montré que $f_*(N)$ est un nuage d'intensité $f_*(\mu)$. □

On a déjà vu que les mesures aléatoires de Poisson étaient stables par somme dénombrable. Un autre résultat de stabilité très simple est :

Proposition 4.21 (Stabilité par restriction)**.** *Soit N un nuage poissonnien d'intensité μ sur $\mathfrak{X}$. Si $\mathfrak{Y} \subset \mathfrak{X}$ est une partie mesurable, alors la restriction $N_{|\mathfrak{Y}}$ est un nuage poissonnien d'intensité $\mu_{|\mathfrak{Y}}$. De plus, si $(\mathfrak{Y}_i)_{i \in I}$ est une famille de parties disjointes, alors les mesures restreintes $N_{|\mathfrak{Y}_i}$ sont indépendants.*

Une propriété plus subtile est la stabilité par marquage mesurable. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathfrak{X}$, et M une variable aléatoire à valeurs dans $\mathbb{N} \sqcup \{\infty\}$. On suppose que

$$N = \sum_{i=1}^{M} \delta_{X_i},$$

qui est une variable aléatoire dans $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X})$, est une mesure aléatoire de Poisson d'intensité $\mu \in \mathscr{M}^{+,s}(\mathfrak{X})$. Par exemple, si μ est une mesure finie, et si $M \sim \mathscr{P}(\mu(\mathfrak{X}))$ et les X_n sont des variables indépendantes de loi $\frac{\mu(\cdot)}{\mu(\mathfrak{X})}$, alors on sait que N est un nuage poissonnien d'intensité μ. Soit $(Y_n)_{n\in\mathbb{N}}$ une suite de variables à valeurs dans un espace polonais $\mathfrak{Y}$, qui sont indépendantes de $(X_n)_{n\in\mathbb{N}}$ et de M, indépendantes entre elles et toutes de même loi μ'.

Proposition 4.22 (Stabilité par marquage). *Dans le contexte précédent, le nuage marqué*

$$\widetilde{N} = \sum_{i=1}^{M} \delta_{(X_i,Y_i)}$$

est une mesure aléatoire de Poisson sur $\mathfrak{X}\times\mathfrak{Y}$, d'intensité $\mu\otimes\mu'$.

Preuve. Soit $f : \mathfrak{X}\times\mathfrak{Y} \to \mathbb{R}_+$ une fonction mesurable positive. On calcule

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{-\widetilde{N}(f)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{-\sum_{i=1}^{M} f(X_i,Y_i)} \mid M,\,(X_n)_{n\in\mathbb{N}}\right]\right]\\ &= \mathbb{E}\left[\prod_{i=1}^{M}\int_{\mathfrak{Y}} \mathrm{e}^{-f(X_i,y)}\,\mu'(dy)\right] = \mathbb{E}\left[\exp\left(\sum_{i=1}^{M}\log\left(\int_{\mathfrak{Y}} \mathrm{e}^{-f(X_i,y)}\,\mu'(dy)\right)\right)\right].\end{aligned}$$

L'application $F : x \mapsto -\log\left(\int_{\mathfrak{Y}} \mathrm{e}^{-f(x,y)}\,\mu'(dy)\right)$ est mesurable par rapport à x, et positive ; et par la formule de Campbell,

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{-\widetilde{N}(f)}\right] = \mathbb{E}\left[\mathrm{e}^{-N(F)}\right] &= \exp\left(\int_{\mathfrak{X}} (\mathrm{e}^{-F(x)}-1)\,\mu(dx)\right)\\ &= \exp\left(\int_{\mathfrak{X}\times\mathfrak{Y}} (\mathrm{e}^{-f(x,y)}-1)\,\mu(dx)\,\mu'(dy)\right).\end{aligned}$$

Comme la formule de Campbell caractérise les nuages poissonniens, on conclut que $\widetilde{N}$ est un nuage poissonnien de points d'intensité $\mu\otimes\mu'$. □

Exemple 4.23. Soit $N = \sum_{i=1}^{M} \delta_{X_i}$ une mesure aléatoire de Poisson d'intensité μ sur $\mathfrak{X}$, et $(Y_n)_{n\in\mathbb{N}}$ une suite de variables de Bernoulli indépendantes de paramètre p. Comme $\widetilde{N} = \sum_{i=1}^{M} \delta_{(X_i,Y_i)}$ est une mesure aléatoire de Poisson d'intensité $\mu\otimes\mathscr{B}(p)$ sur $\mathfrak{X}\times\{0,1\}$, si

$$N_p = \sum_{i=1}^{M} Y_i\,\delta_{X_i}$$

alors N_p est la restriction de $\widetilde{N}$ à $\mathfrak{X}\times\{1\} = \mathfrak{X}$, donc N_p est une mesure aléatoire de Poisson d'intensité $(\mu\otimes\mathscr{B}(p))_{|\mathfrak{X}\times\{1\}} = p\mu$.

4.4 Processus ponctuels de Poisson

▷ **Mesures aléatoires à paramètre temporel.** Soit $\mathfrak{X}$ un espace polonais, et μ une mesure s-finie sur $\mathfrak{X}$. On considère le nuage poissonnien N sur $\mathfrak{X}\times\mathbb{R}_+$, d'intensité $\mu\otimes dt$. On parlera aussi de mesure aléatoire de Poisson à paramètre temporel, d'intensité μ sur $\mathfrak{X}$. Comme μ et dt sont s-finies, il en va de même pour $\mu\otimes dt$, qui est par ailleurs sans atome, car dt est sans atome. Le nuage poissonnien est donc bien défini, et avec tous ses points (X,T) qui sont simples. En fait, on a le résultat plus fort suivant :

Proposition 4.24. *Avec probabilité* 1*, pour tout* $t \geq 0$*,* $N(\mathfrak{X} \times \{t\}) \in \{0,1\}$.

Preuve. Supposons d'abord μ de masse finie sur $\mathfrak{X}$. Dans ce cas, pour tout $n \geq 1$ et tout $T \geq 0$,

$$\begin{aligned}\mathbb{P}[\exists t \in [0,T],\ N(\mathfrak{X} \times \{t\}) \geq 2] &\leq \mathbb{P}\left[\exists j \in [\![1,n]\!],\ N\left(\mathfrak{X} \times \left[\frac{j-1}{n}T, \frac{j}{n}T\right]\right) \geq 2\right] \\ &\leq \sum_{j=1}^{n} (1 - \mathrm{e}^{-\lambda}(1+\lambda)) \quad \text{avec } \lambda = \frac{\mu(\mathfrak{X})\,T}{n} \\ &\leq \sum_{j=1}^{n} \lambda^2 = \frac{(\mu(\mathfrak{X})\,T)^2}{n}.\end{aligned}$$

Comme ceci est vrai pour tout n, $\mathbb{P}[\exists t \in [0,T],\ N(\mathfrak{X} \times \{t\}) \geq 2] = 0$. En faisant tendre T vers l'infini, on conclut que si μ est de masse finie, alors

$$\mathbb{P}[\exists t \in \mathbb{R}_+,\ N(\mathfrak{X} \times \{t\}) \geq 2] = 0.$$

Dans le cas général, décomposons μ comme somme de mesures finies : $\mu = \sum_{i=1}^{\infty} \mu_i$. On peut alors écrire $N = \sum_{i=1}^{\infty} N_i$, où les N_i sont des mesures aléatoires poissonniennes indépendantes d'intensité $\mu_i \times dx$ sur $\mathfrak{X} \times \mathbb{R}_+$. Comme

$$\mathbb{P}\left[\exists t \in \mathbb{R}_+,\ \left(\sum_{i=1}^{I} N_i\right)(\mathfrak{X} \times \{t\}) \geq 2\right] = 0$$

pour tout $I \geq 1$, on en déduit le même résultat pour $N = \sup_{I \geq 1} \sum_{i=1}^{I} N_i$; c'est ce que l'on voulait démontrer. □

Ainsi, si N est une mesure aléatoire de Poisson à paramètre temporel, alors pour tout $t \in \mathbb{R}_+$, on a soit $N_{|\mathfrak{X}\times\{t\}} = 0$, soit $N_{|\mathfrak{X}\times\{t\}} = \delta_{(x,t)}$ pour un certain point $x \in \mathfrak{X}$. Dans ce qui suit, on définit X_t comme l'application à valeurs dans $\mathfrak{X} \sqcup \{\dagger\}$ donnée par :

$$X_t = \begin{cases} x & \text{si } N_{|\mathfrak{X}\times\{t\}} = \delta_{(x,t)}, \\ \dagger & \text{si } N \text{ ne charge pas } \mathfrak{X} \times \{t\}. \end{cases}$$

On appelle $(X_t)_{t \in \mathbb{R}_+}$ le processus ponctuel de Poisson d'intensité μ. Ce processus n'a aucune régularité, si ce n'est :

Proposition 4.25. *Pour tout* $t \in \mathbb{R}_+$*,* X_t *est une variable aléatoire à valeurs dans* $\mathfrak{X} \sqcup \{\dagger\}$*, c'est-à-dire une fonction mesurable sur l'espace de probabilité* $(\Omega, \mathscr{F}, \mathbb{P})$ *sur lequel la mesure aléatoire à paramètre temporel* N *est définie.*

Preuve. Pour toute partie $A \in \mathscr{B}(\mathfrak{X})$,

$$\{X_t \in A\} = \{N(A \times \{t\}) = 1\},$$

tandis que $\{X_t = \dagger\}$ est le même événement que $\{N(\mathfrak{X} \times \{t\}) = 0\}$. Par hypothèse, N est une variable aléatoire à valeurs dans $\mathscr{M}^{\mathrm{atom}}(\mathfrak{X} \times \mathbb{R}_+)$, donc ces événements sont bien mesurables. □

Le processus ponctuel de Poisson $(X_t)_{t\in\mathbb{R}_+}$ d'intensité μ peut être reconstruit à partir d'un processus de Poisson usuel lorsque μ est une mesure finie sur $\mathfrak{X}$. Plus précisément :

Proposition 4.26. *Si $\mu(\mathfrak{X}) = +\infty$, alors $\{t \,|\, X_t \neq \dagger\}$ est dense dans $\mathbb{R}_+$. Si $0 < \mu(\mathfrak{X}) < +\infty$, alors l'ensemble aléatoire $\{t \,|\, X_t \neq \dagger\}$ est presque sûrement infini discret. Il s'écrit dans ce cas $\{T_1 < T_2 < \cdots < T_n < \cdots\}$, avec :*

1. *les variables $T_1, T_2 - T_1, \ldots, T_n - T_{n-1}, \ldots$ indépendantes et suivant la loi exponentielle de paramètre $\mu(\mathfrak{X})$;*
2. *les variables $X_{T_1}, X_{T_2}, \ldots, X_{T_n}, \ldots$ indépendantes des temps $T_1, T_2, \ldots$ et indépendantes entre elles, et suivant toutes la loi $\frac{\mu(\cdot)}{\mu(\mathfrak{X})}$.*

Preuve. Si $\mu(\mathfrak{X}) = +\infty$, alors la variable $N(\mathfrak{X}\times[s,s+\varepsilon])$ suit une loi de Poisson de paramètre infini pour tout $s \in \mathbb{R}_+$ et $\varepsilon > 0$, donc est égale à $+\infty$ presque sûrement. Par conséquent, $\{t \in [s,s+\varepsilon] \,|\, X_t \neq \dagger\}$ est non vide, et même infini presque sûrement : ceci démontre la densité de l'ensemble $\{t \in \mathbb{R}_+ \,|\, X_t \neq \dagger\}$.

Supposons maintenant $\mu(\mathfrak{X}) < \infty$. Alors, une réalisation possible de N est par marquage de la mesure aléatoire de Poisson d'intensité $\mu(\mathfrak{X})\,dt$ sur $\mathbb{R}_+$. En effet, si $P = \sum_{i=1}^{\infty} \delta_{T_i}$ est une telle mesure aléatoire, et si $(Y_i)_{i\in\mathbb{N}}$ est une famille de variables aléatoires indépendantes de loi $\frac{\mu(\cdot)}{\mu(\mathfrak{X})}$, alors d'après la proposition 4.22, $N = \sum_{i=1}^{\infty} \delta_{(Y_i,T_i)}$ est une mesure poissonnienne d'intensité $\mu \otimes dt$. Remarquons de plus que les points de la mesure P peuvent être indexés de sorte que $T_1 < T_2 < \cdots < T_n < \cdots$, car P est presque sûrement une mesure infinie discrète sur $\mathbb{R}_+$. En effet, comme dt est sans atome, tous les T_i sont p.s. distincts, et dans tout intervalle $[a,b]$, le nombre de points T_i suit une loi de Poisson de paramètre $(b-a)\,\mu(\mathfrak{X})$, donc est fini. Avec cette convention, on a $Y_i = X_{T_i}$, et il reste donc à démontrer le premier point de la proposition, qui concerne une mesure aléatoire de Poisson sur $\mathbb{R}_+$ d'intensité $\lambda\,dt$ avec $\lambda = \mu(\mathfrak{X})$.

Si $(T_n)_{n\geq 1}$ est la suite croissante des atomes d'une mesure aléatoire poissonnienne P sur $\mathbb{R}_+$ d'intensité $\mu(\mathfrak{X})\,dt$, considérons d'abord le temps T_1. On a

$$\mathbb{P}[T_1 \geq t] = \mathbb{P}[P([0,t)) = 0] = \mathrm{e}^{-t\mu(\mathfrak{X})},$$

donc T_1 suit bien une loi $\mathscr{E}(\mu(\mathfrak{X}))$. Montrons maintenant que pour tout $n \geq 1$, la variable $T_n - T_{n-1}$ est indépendant de $(T_i)_{i\leq n-1}$ et de même loi que T_1. Si $f : (\mathbb{R}_+)^{n-1} \to \mathbb{R}$ et $g : \mathbb{R}_+ \to \mathbb{R}$ sont des fonctions continues à support compact avec $f_{|([0,K]^{n-1})^c} = 0$, alors

$$\begin{aligned}
&\mathbb{E}[f(T_1, T_2, \ldots, T_{n-1})\,g(T_n - T_{n-1})] \\
&= \lim_{m\to\infty} \left(\sum_{p=1}^{m} \mathbb{E}\Big[f(T_1,\ldots,T_{n-1})\,1_{T_{n-1}\in(\frac{p-1}{m}K,\frac{p}{m}K]}\,1_{T_n>\frac{p}{m}K}\, g\Big(T_n - \frac{p}{m}K\Big)\Big]\right) \\
&= \lim_{m\to\infty} \left(\sum_{p=1}^{m} \mathbb{E}\Big[f(T_1,\ldots,T_{n-1})\,1_{T_{n-1}\in(\frac{p-1}{m}K,\frac{p}{m}K]}\,1_{T_n>\frac{p}{m}K} \times \right. \\
&\qquad\qquad \left. g\Big(\text{premier atome de } P \text{ sur } \Big[\frac{p}{m}K, +\infty\Big) - \frac{p}{m}K\Big)\Big]\right).
\end{aligned}$$

Remarquons que la variable $f(T_1,\ldots,T_{n-1})\,1_{T_{n-1}\in(\frac{p-1}{m}K,\frac{p}{m}K]}\,1_{T_n>\frac{p}{m}K}$ est mesurable par rapport à la tribu engendrée par les $P(A)$, $A \subset [0, \frac{p}{M}K]$, tandis que la variable

$$g\left(\text{premier atome de } P \text{ sur } \left[\frac{p}{m}K, +\infty\right) - \frac{p}{m}K\right)$$

est mesurable par rapport à la tribu engendrée par les $P(A)$, $A \subset (\frac{p}{M}K, +\infty)$. Par conséquent, les deux variables sont indépendantes, et de plus, la loi de

$$\text{premier atome de } P \text{ sur } \left[\frac{p}{m}K, +\infty\right) - \frac{p}{m}K$$

est exponentielle de paramètre $\mu(\mathfrak{X})$, pour les mêmes raisons que pour T_1. Donc,

$$\begin{aligned}
&\mathbb{E}[f(T_1, T_2, \ldots, T_{n-1})\, g(T_n - T_{n-1})] \\
&= \lim_{m\to\infty} \left(\sum_{p=1}^{m} \mathbb{E}\Big[f(T_1, \ldots, T_{n-1})\, 1_{T_{n-1} \in (\frac{p-1}{m}K, \frac{p}{m}K]}\, 1_{T_n > \frac{p}{m}K} \Big]\, \mathbb{E}[g(T_1)] \right) \\
&= \mathbb{E}[f(T_1, \ldots, T_{n-1})]\, \mathbb{E}[g(T_1)].
\end{aligned}$$

Ceci implique le résultat annoncé, et achève la preuve de la proposition. □

▷ **Processus de Poisson et processus de Poisson composés.** On peut utiliser les mesures aléatoires à paramètre temporel pour construire une grande classe de processus càdlàg, par intégration de fonctions $f \in \mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$. Dans ce qui suit, on fixe une mesure μ qui est s-finie sur un espace polonais $\mathfrak{X}$, et on note N la mesure poissonnienne à paramètre temporel d'intensité $\mu \otimes dt$ sur $\mathfrak{X} \times \mathbb{R}_+$.

Théorème 4.27. *Soit $f \in \mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$, et*

$$N_t(f) = \int_{\mathfrak{X}\times[0,t]} f(x)\, N(dx\, dt) = \sum_{0 \leq s \leq t,\, X_s \neq \dagger} f(X_s).$$

Le processus $(N_t(f))_{t\in\mathbb{R}_+}$ est bien défini et presque sûrement càdlàg, donc peut être considéré comme une variable à valeurs dans $\mathscr{D}(\mathbb{R}_+)$. De plus, pour tout temps $s \geq 0$, le processus $(N_{s+t}(f) - N_s(f))_{t\in\mathbb{R}_+}$ est indépendant de $(N_u(f))_{0\leq u\leq s}$, et de même loi que $(N_t(f))_{t\in\mathbb{R}_+}$.

Preuve. Pour tout temps T, la fonction $(x,t) \mapsto f(x)$ est mesurable et appartient à l'espace $\mathscr{L}^1(\mathfrak{X} \times [0,T], \mathscr{B}(\mathfrak{X} \times [0,T]), \mu \otimes dt)$, donc $N_T(f) = N_{|\mathfrak{X}\times[0,T]}(f)$ est bien définie et d'espérance $\int_{\mathfrak{X}\times[0,T]} f(x)\, \mu(dx)\, dt = T\, \mu(f)$. D'autre part, si $\omega \in \Omega$ est fixé, alors par convergence dominée, l'application $t \mapsto N_t(f)(\omega)$ est continue à droite et avec limite à gauche. On en déduit que $(N_t(f))_{t\in\mathbb{R}_+}$ est bien une variable aléatoire à valeurs dans $\mathscr{D}(\mathbb{R}_+)$.

Fixons un temps $s \geq 0$, et considérons $N_{s+t}(f) - N_s(f) = \int_{\mathfrak{X}\times[s,s+t]} f(x)\, N(dx\, dt)$. Ce processus est mesurable par rapport à la tribu engendrée par la mesure aléatoire $N_{|\mathfrak{X}\times(s,+\infty)}$, qui est indépendante de celle engendrée par $N_{|\mathfrak{X}\times[0,s]}$. Par conséquent, $(N_{s+t}(f) - N_s(f))_{t\in\mathbb{R}_+}$ est indépendant de $(N_u(f))_{0\leq u\leq s}$. Finalement, un argument d'invariance par translation de la mesure de Lebesgue montre que ce processus càdlàg a les mêmes lois fini-dimensionnelles que le processus initial (en considérant par exemple les transformées de Fourier, calculables par la formule de Campbell), donc est de même loi que $(N_t(f))_{t\in\mathbb{R}_+}$. □

Exemple 4.28. Soit ν une mesure sur $\mathbb{R}$, qui intègre $x \mapsto |x|$. On appelle processus de Poisson composé d'intensité ν le processus $(P_t(\nu))_{t\in\mathbb{R}_+} = (N_t(x))_{t\in\mathbb{R}_+}$ associé à l'intensité $\nu \otimes dt$ et la fonction identité. On a :

$$P_t(\nu) = N_t(x) = \int_{\mathbb{R}\times[0,t]} x\, N(dx\, dt) = \sum_{0\leq s\leq t,\, X_s \neq \dagger} X_s.$$

Le processus de Poisson homogène d'intensité λ correspond à la mesure $\nu = \lambda\delta_1$. Pour tout temps t,

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi P_t(\nu)}] = \mathbb{E}[\mathrm{e}^{N_t(\mathrm{i}\xi x)}] = \mathbb{E}\left[\mathrm{e}^{\int_{\mathbb{R}\times[0,t]} \mathrm{i}\xi x\, N(dx\, dt)}\right] = \exp\left(t \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}\xi x} - 1)\, \nu(dx)\right).$$

On verra dans le chapitre suivant que cette formule est un cas particulier d'une formule générale pour les lois de processus à accroissements stationnaires et indépendants (le cas d'un processus de Lévy purement discontinu de mesure de Lévy intégrant $|x|$). Notons que si ν n'intègre pas $|x|$ mais est de masse finie, alors on peut toujours considérer la variable aléatoire

$$P_t(\nu) = \sum_{0\leq s\leq t,\, X_s \neq \dagger} X_s,$$

qui est une somme finie presque sûrement puisque $\nu \otimes dt$ est de masse finie sur $\mathbb{R} \times [0,t]$. La transformée de Fourier de cette variable aléatoire est toujours donnée par $\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi P_t(\nu)}] = \exp(t \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}\xi x} - 1)\, \nu(dx))$; en revanche, $P_t(\nu)$ n'est plus forcément intégrable. Ces deux cas d'existence du processus de Poisson composé (mesure intégrant $|x|$ ou mesure de masse finie) joueront un rôle important dans la construction générale des processus de Lévy.

Références

L'essentiel de ce chapitre est issu de [J. F. C. Kingman, *Poisson processes*, Oxford Studies in Probability vol. 3, Oxford University Press, 1993]. On renvoie également à [K.-I. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge studies in advanced mathematics vol. 68, Cambridge University Press, 1999, Chapter 19]. Il y a néanmoins une différence entre notre exposé et celui des ouvrages précités, au niveau du cadre choisi pour les mesures aléatoires de Poisson. Dans notre cadre, l'intensité peut être n'importe quelle mesure s-finie sur un espace polonais, alors que le traitement classique utilise des intensités σ-finies, ou σ-finies sans atome sur un espace topologique vérifiant la condition de Hausdorff (ou au moins, mesurable et tel que la diagonale $\{(x,x)\,|\,x \in \mathfrak{X}\}$ soit mesurable dans $\mathfrak{X}^2$). L'avantage de notre cadre est que la stabilité par somme est une propriété immédiate.

Exercices

1. Soit N un nuage poissonnien d'intensité $dx\,dy$ sur $\mathbb{R}^2$. On note N_θ et N_R les nuages de Poisson sur $\mathbb{R}/2\pi\mathbb{Z}$ et sur $\mathbb{R}_+$ obtenus à partir de N en considérant les angles et les distances à l'origine des points de N. Donner les intensités de ces deux nuages. Sont-ils indépendants ?

2. Soit $N = \sum_{i\in I} \delta_{X_i}$ un nuage poissonnien d'intensité μ sur un espace polonais. Si $f : \mathfrak{X} \to \mathbb{R}$ est dans $\mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$, on sait que $N(f)$ est intégrable, donc fini presque sûrement. Montrer que plus généralement, $N(|f|) = \sum_{i\in I} |f(X_i)|$ est fini presque sûrement si et seulement si $\min(1, |f|)$ est dans $\mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$. On pourra remarquer que si $f \geq 0$, alors

$$N(f) \text{ fini p.s.} \iff \lim_{\lambda\to 0} \mathrm{e}^{-\lambda N(f)} = 1 \text{ p.s.} \iff \lim_{\lambda\to 0} \mathbb{E}[\mathrm{e}^{-\lambda N(f)}] = 1.$$

   Application : soit $\mu = dx$ la mesure uniforme sur $[0,1]$, $f(x) = \frac{1}{x}$. Montrer que le nuage de Poisson N d'intensité μ vérifie $N(f) < +\infty$ p.s., mais $\mathbb{E}[N(f)] = +\infty$.

3. Soit $N = \sum_{n\in\mathbb{Z}} \delta_{T_n}$ un nuage poissonnien d'intensité dx sur $\mathbb{R}$. Montrer qu'on peut effectivement indexer les atomes de N comme suit :

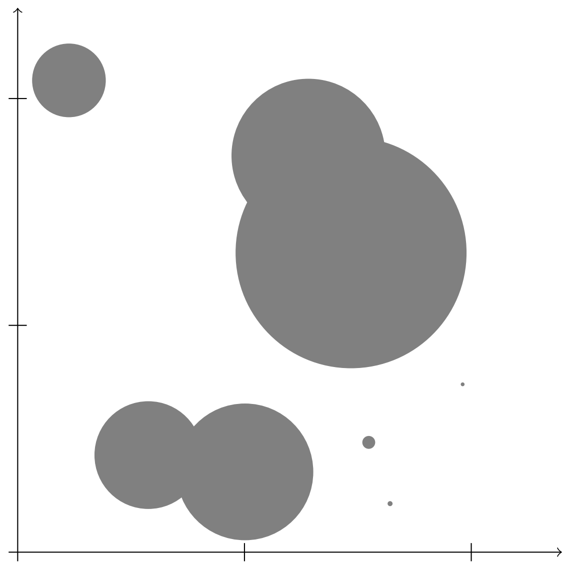
$$\cdots < T_{-n} < \cdots < T_{-1} < T_0 < 0 < T_1 < \cdots < T_n < \cdots$$

   avec $\lim_{n\to\infty} T_n = \lim_{n\to\infty} -T_{-n} = +\infty$ presque sûrement. Soit $(X_n)_{n\in\mathbb{Z}}$ une famille de variables aléatoires sur $\mathbb{R}$ indépendantes de N et toutes de même loi μ. On pose

$$\widetilde{N} = \sum_{n\in\mathbb{Z}} \delta_{T_n+X_n}.$$

   Montrer que $\widetilde{N}$ est encore un nuage poissonnien d'intensité dx sur $\mathbb{R}$.

4. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$, telle que $N = \sum_{n=1}^{\infty} \delta_{X_n}$ soit une mesure poissonnienne sur $\mathbb{R}^d$ d'intensité la mesure de Lebesgue. On considère également une suite $(R_n)_{n\in\mathbb{N}}$ de variables aléatoires à valeurs dans $\mathbb{R}_+$, indépendantes et de même loi.



On note $A = \bigcup_{n=1}^{\infty} B_{(X_n,R_n)}$, qui est un ensemble aléatoire de $\mathbb{R}^d$; et μ la loi des variables R_n.

(a) Pour tout $x \in \mathbb{R}^d$, montrer que le nombre de boules $B_n = B_{(X_n,R_n)}$ qui recouvrent x suit une loi de Poisson de paramètre

$$\lambda = \left(\int_0^{\infty} r^d \mu(dr)\right) \mathrm{vol}(B_{(0,1)}) = \left(\int_0^{\infty} r^d \mu(dr)\right) \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}.$$

On pourra utiliser la propriété de stabilité par marquage.

(b) On suppose que $\int_0^{\infty} r^d \mu(dr) = +\infty$. Montrer que $x \in \mathbb{R}^d$ est presque sûrement dans A. Montrer ensuite que A est presque sûrement dense dans $\mathbb{R}^d$.

(c) On suppose toujours $\int_0^{\infty} r^d \mu(dr) = +\infty$. Soit

$$\widetilde{A} = \bigcup_{n \,|\, R_n \geq 1} B_{(X_n,(R_n-1))}.$$

Que peut-on dire de $\widetilde{A}$? En déduire que $A = \mathbb{R}^d$ presque sûrement.

5. Soit μ une loi sur $\mathbb{R}$, et N une mesure aléatoire poissonnienne à paramètre temporel d'intensité $dx \otimes \mu$ sur $\mathbb{R} \times \mathbb{R}$. On note $N = \sum_{n\in\mathbb{Z}} \delta_{(X_n,V_n)}$ avec

$$\cdots < X_{-k} < \cdots < X_0 < 0 < X_1 < \cdots < X_l < \cdots,$$

et on considère N comme une distribution de particules sur la droite réelle, chaque particule n étant initialement en position X_n avec vitesse V_n.

(a) On pose $N_t = \sum_{n\in\mathbb{Z}} \delta_{(X_n+tV_n,V_n)}$. Montrer que N_t est encore une mesure aléatoire poissonnienne à paramètre temporel, d'intensité $dx \otimes \mu$.

(b) On suppose $\int_{\mathbb{R}} |v|\,\mu(dv) < \infty$. Soit T_n le temps (éventuellement négatif) auquel la particule n atteint 0 : $T_n = \frac{X_n}{V_n}$. Montrer que $\sum_{n\in\mathbb{Z}} \delta_{(T_n,V_n)}$ est une mesure aléatoire de Poisson sur $\mathbb{R} \times \mathbb{R}$, dont on donnera la loi.

(c) Calculer

$$\mathbb{P}[\{\forall s \in [0,t],\ \forall n \in \mathbb{N},\ X_n + sV_n \notin [a,b]\}],$$

où $[a,b]$ est un intervalle fixé.

6. Soit μ une mesure s-finie sur un espace polonais $\mathfrak{X}$, et N le processus ponctuel de Poisson d'intensité $\mu \otimes dt$ sur $\mathfrak{X} \times \mathbb{R}_+$. On équipe l'espace de probabilité $(\Omega, \mathscr{F}, \mathbb{P})$ sous-jacent à ce processus de la filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ définie par :

$$\mathscr{F}_t = \sigma(N_{|\mathfrak{X}\times[0,t]}).$$

(a) Pour tout $f \in \mathscr{L}^1(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$, montrer que $N_t(f) = \int_{\mathfrak{X}\times[0,t]} f(x) N(dx\,dt)$ est un processus adapté à $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$, et que

$$N_t(f) - \mathbb{E}[N_t(f)] = \int_{\mathfrak{X}\times[0,t]} f(x) N(dx\,dt) - t \int_{\mathfrak{X}} f(x)\,\mu(dx)$$

est une martingale pour cette filtration.

(b) On suppose $f \in \mathscr{L}^2(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \mu)$. Montrer que

$$(N_t(f))^2 - t \int_{\mathfrak{X}} (f(x))^2 \, \mu(dx)$$

est une martingale pour la filtration $(\mathscr{F}_t)_{t \in \mathbb{R}_+}$.

Chapitre 5

Processus de Lévy

Dans ce dernier chapitre, on utilise la théorie des chapitres précédents pour construire une famille assez large de lois et de processus continus à droite et avec limites à gauche, qui apparaissent à de nombreux endroits en théorie des probabilités, en particulier en tant qu'objets limites : les *lois infiniment divisibles* et les *processus de Lévy*.

Définition 5.1 (Loi infiniment divisible). *Une loi infiniment divisible sur $\mathbb{R}$ est une mesure de probabilités borélienne μ telle que, pour tout $n \geq 1$, il existe une loi ρ_n avec $\mu = (\rho_n)^{*n}$, où $(\rho_n)^{*n}$ désigne la convolée n-ième de ρ_n, c'est-à-dire la loi de la somme de n variables indépendantes de loi ρ_n :*

$$(\rho_n)^{*n}(A) = \int_{\mathbb{R}^n} 1_A(x_1 + x_2 + \cdots + x_n)\,\rho_n(dx_1)\cdots\rho_n(dx_n).$$

Définition 5.2 (Processus de Lévy). *On appelle processus de Lévy un processus $(X_t)_{t\geq 0}$ dans $\mathscr{D}(\mathbb{R}_+)$, tel que pour tout temps s, le processus $(X_{t+s} - X_s)_{t\geq 0}$ est indépendant de la tribu $\mathscr{F}_s = \sigma(X_{u\leq s})$, et de même loi que $(X_t)_{t\geq 0}$. Autrement dit, $(X_t)_{t\geq 0}$ est à accroissements indépendants et stationnaires.*

Si $(X_t)_{t\geq 0}$ est un processus de Lévy, il n'est pas difficile de voir que pour tout temps t, la loi μ_t de $X_t - X_0$ est infiniment divisible. Réciproquement, on verra au cours du chapitre que toute loi infiniment divisible est la marginale au temps $t = 1$ (ou à un temps arbitraire $t > 0$) d'un processus de Lévy. Par conséquent, la classification de ces deux types d'objets est essentiellement la même. Dans la première section 5.1, on montre que les lois infini-divisibles ont une transformée de Fourier qui prend une forme bien précise, dépendant de seulement trois paramètres

$$a \in \mathbb{R},\ \sigma^2 \geq 0,\ \nu,$$

où ν est une mesure positive sur $\mathbb{R}\setminus\{0\}$ qui intègre $x \mapsto \min(1, x^2)$. C'est la formule de Lévy–Khintchine (Théorème 5.12), qui peut être comprise intuitivement comme suit : toute loi infiniment divisible est la combinaison d'une loi gaussienne et de lois de Poisson composées et compensées. Au niveau des processus de Lévy, une décomposition équivalente existe et est démontrée dans la section 5.2 : tout processus de Lévy est la combinaison d'un mouvement brownien avec drift, et de processus de Poisson composés et compensés. Cette décomposition dite de Lévy–Itô prend une forme particulière lorsqu'on se restreint aux processus de

Lévy croissants, appelés subordinateurs ; cette classe de processus de Lévy est étudiée en détail dans la dernière section 5.3.

Remarque 5.3. La théorie développée dans ce chapitre admet une généralisation immédiate avec des lois et des processus càdlàg à valeurs dans $\mathbb{R}^d$ avec $d \geq 2$; voir les références en fin de chapitre.

5.1 Lois infiniment divisibles

▷ **Exemples et propriétés générales.** Pour commencer notre étude des lois infiniment divisibles, donnons une liste de lois classiques qui ont cette propriété :

Proposition 5.4. *Les lois suivantes sont toutes infiniment divisibles :*

— *les lois gaussiennes* $\mathscr{N}_{\mathbb{R}}(m,\sigma^2)$ *avec* $m \in \mathbb{R}$ *et* $\sigma^2 \geq 0$.

— *les lois de Poisson* $\mathscr{P}(\lambda)$ *avec* $\lambda \geq 0$.

— *les lois géométriques* $\mathscr{G}(p)$ *avec* $p \in (0,1)$.

— *les lois exponentielles* $\mathscr{E}(\alpha)$ *avec* $\alpha > 0$*, et plus généralement les lois* $\Gamma(\alpha,c)$ *de densité*

$$1_{x\geq 0}\,\frac{\alpha^c\,x^{c-1}\,\mathrm{e}^{-\alpha x}}{\Gamma(c)}\,dx$$

avec $\alpha > 0$ *et* $c > 0$.

— *les lois de Cauchy* $\mathscr{C}(m,\gamma)$ *de densité*

$$\frac{1}{\pi\gamma\left(1+\left(\frac{x-m}{\gamma}\right)^2\right)}\,dx$$

avec $m \in \mathbb{R}$ *et* $\gamma > 0$.

Preuve. En termes de transformées de Fourier, si $\mu = (\rho_n)^{*n}$, alors $\widehat{\mu}(\xi) = (\widehat{\rho_n}(\xi))^n$ et réciproquement. Par conséquent, il suffit de calculer les transformées de Fourier des lois listées dans la proposition et de montrer qu'elles ont des racines n-ièmes pour tout n.

— Si $\mu = \mathscr{N}_{\mathbb{R}}(m,\sigma^2)$, alors

$$\widehat{\mu}(\xi) = \mathrm{e}^{\mathrm{i}m\xi-\frac{\sigma^2\xi^2}{2}} = \left(\mathrm{e}^{\frac{\mathrm{i}m\xi}{n}-\frac{\sigma^2\xi^2}{2n}}\right)^n = (\widehat{\rho_n}(\xi))^n$$

avec $\rho_n = \mathscr{N}_{\mathbb{R}}(\frac{m}{n},\frac{\sigma^2}{n})$.

— Si $\mu = \mathscr{P}(\lambda)$, alors

$$\widehat{\mu}(\xi) = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\xi}-1)} = \left(\mathrm{e}^{\frac{\lambda}{n}(\mathrm{e}^{\mathrm{i}\xi}-1)}\right)^n = (\widehat{\rho_n}(\xi))^n$$

avec $\rho_n = \mathscr{P}(\frac{\lambda}{n})$.

— Si $\mu = \mathscr{G}(p)$ est la loi géométrique sur $\mathbb{N}$, donnée par $\mu(k) = p\,(1-p)^k$, alors

$$\widehat{\mu}(\xi) = \frac{p}{1-(1-p)\mathrm{e}^{\mathrm{i}\xi}} = (\widehat{\rho_n}(\xi))^n,$$

où ρ_n est la loi binomiale négative, donnée sur $\mathbb{N}$ par la formule

$$\rho_n(k) = \frac{(k-1+\frac{1}{n})(k-2+\frac{1}{n})\cdots\frac{1}{n}}{k!}\,p^{\frac{1}{n}}\,(1-p)^k.$$

— Si $\mu = \Gamma(\alpha, c)$ (le cas $c = 1$ correspondant aux lois exponentielles), alors

$$\widehat{\mu}(\xi) = \frac{1}{\left(1-\frac{\mathrm{i}\xi}{\alpha}\right)^c} = (\widehat{\rho_n}(\xi))^n$$

avec $\rho_n = \Gamma(\alpha, \frac{c}{n})$.

— Enfin, si $\mu = \mathscr{C}(m, \gamma)$, alors

$$\widehat{\mu}(\xi) = \mathrm{e}^{\mathrm{i}m\xi-\gamma|\xi|} = \left(\mathrm{e}^{\frac{\mathrm{i}m\xi}{n}-\frac{\gamma|\xi|}{n}}\right)^n = (\widehat{\rho_n}(\xi))^n$$

avec $\rho_n = \mathscr{C}(\frac{m}{n}, \frac{\gamma}{n})$ (pour le calcul de la transformée de Fourier de la loi de Cauchy, on peut utiliser la formule d'inversion de Fourier comme dans la preuve du lemme 2.11). □

A contrario, les lois binomiales et les lois uniformes ne sont pas infiniment divisibles, compte tenu de :

Proposition 5.5. *Soit μ une loi infiniment divisible. Si μ a un support borné, alors $\mu = \delta_c$ pour une certaine constante c.*

Preuve. Supposons μ de support borné, inclus dans $[-C, C]$. Alors, si $\mu = (\rho_n)^{*n}$, le support de ρ_n est inclus dans $[-\frac{C}{n}, \frac{C}{n}]$, donc $\mathrm{var}(\rho_n) \leq \frac{C^2}{n^2}$. Par conséquent,

$$\mathrm{var}(\mu) = n\,\mathrm{var}(\rho_n) \leq \frac{C^2}{n}$$

pour tout $n \geq 1$, donc $\mathrm{var}(\mu) = 0$ et μ est la loi d'une variable constante. □

Une autre propriété très importante des lois infiniment divisibles est :

Proposition 5.6. *Soit μ une loi infiniment divisible. La transformée de Fourier $\widehat{\mu}$ ne s'annule pas sur $\mathbb{R}$.*

Preuve. Soient X et X' deux variables indépendantes de loi μ. La différence $Z = X - X'$ a transformée de Fourier

$$\phi_Z(\xi) = \phi_X(\xi)\,\phi_{-X'}(\xi) = \widehat{\mu}(\xi)\,\widehat{\mu}(-\xi) = |\widehat{\mu}(\xi)|^2.$$

Notons ρ_n une loi telle que $\mu = (\rho_n)^{*n}$, et $Z_n = X_n - X'_n$, où X_n et X'_n sont deux variables indépendantes de loi ρ_n. On a de même $\phi_{Z_n}(\xi) = |\widehat{\rho_n}(\xi)|^2$, et comme ce nombre est positif,

$$\phi_{Z_n}(\xi) = (\phi_Z(\xi))^{1/n}.$$

La limite ponctuelle lorsque n tend vers l'infini de cette racine n-ième est $1_{\phi_Z(\xi)\neq 0} = 1_{\phi_X(\xi)\neq 0}$. De plus, la suite $(Z_n)_{n\in\mathbb{N}}$ est tendue, car

$$\begin{aligned}\mathbb{P}\left[|Z_n| > \frac{2}{\delta}\right] &\leq \frac{1}{\delta}\int_{-\delta}^{\delta}(1-\phi_{Z_n}(\xi))\,d\xi =\leq \frac{1}{\delta}\int_{-\delta}^{\delta}\left(1-|\phi_X(\xi)|^{\frac{2}{n}}\right)d\xi \\ &\leq \frac{2}{n}\left(\frac{1}{\delta}\int_{-\delta}^{\delta}\log\frac{1}{|\phi_X(\xi)|}\,d\xi\right) \leq 2\left(\frac{1}{\delta}\int_{-\delta}^{\delta}\log\frac{1}{|\phi_X(\xi)|}\,d\xi\right),\end{aligned}$$

qui pour δ petit peut être bornée par une quantité arbitrairement petite, car $\xi \mapsto |\phi_X(\xi)|$ est continue en 0. Par conséquent, il existe une extraction $(n_k)_{k\in\mathbb{N}}$ telle que $(Z_{n_k})_{k\in\mathbb{N}}$ admet une limite en loi lorsque k tend vers l'infini. Or, la transformée de Fourier de cette limite $1_{\phi_X(\xi)\neq 0}$ doit être une fonction continue, ce qui n'est possible que si ϕ_X ne s'annule pas (car $\phi_X(0) = 1$). □

Lemme 5.7 (Relèvement des fonctions continues ne s'annulant pas). *Soit $f : \mathbb{R} \to \mathbb{C}$ une fonction continue qui ne s'annule pas. Il existe une fonction continue $g : \mathbb{R} \to \mathbb{C}$ telle que $f(\xi) = \exp(g(\xi))$.*

Preuve. Comme $|f|$ est une fonction continue de $\mathbb{R}$ dans $\mathbb{R}_+$ qui ne s'annule pas, on peut considérer $r = \log|f|$, qui est une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$. Alors, $\widetilde{f} = \frac{f}{\exp r} = \frac{f}{|f|}$ est une fonction continue de $\mathbb{R}$ dans $\mathbb{T} = \{z \in \mathbb{C} \,|\, |z| = 1\}$, et on s'est donc ramené au cas d'une fonction à valeurs dans le cercle unité. Sans perte de généralité, on peut aussi supposer $f(0) = 1$. Notons maintenant que, si $h_1 : I_1 \to \mathbb{R}$ et $h_2 : I_2 \to \mathbb{R}$ sont deux fonctions continues définies sur des intervalles contenant 0 et telles que $\exp(\mathrm{i}h_1(\xi)) = f(\xi) = \exp(\mathrm{i}h_2(\xi))$, alors $(h_1 - h_2)(\xi) \in 2\pi\mathbb{Z}$ pour tout $\xi \in I_1 \cap I_2$, et par continuité, $h_1 - h_2 = 2\pi k$ pour un certain entier k constant. Si l'on impose $h_1(0) = h_2(0) = 0$, on a donc $h_1 = h_2$ sur $I_1 \cap I_2$. On a donc unicité du relèvement h si l'on impose $h(0) = 0$. Notons

$$T = \sup\left\{t > 0 \mid \exists h : [-t,t] \to \mathbb{R} \text{ continue, avec } \begin{cases} h(0) = 0, \\ \exp(\mathrm{i}h(\xi)) = f(\xi) \text{ pour tout } |\xi| \leq t \end{cases}\right\}.$$

Si $T = +\infty$, alors le lemme est démontré. Supposons par l'absurde T fini. Comme f est continue en T et $-T$, il existe des intervalles $[T-\varepsilon, T+\varepsilon]$ et $[-T-\varepsilon, -T+\varepsilon]$ avec $\varepsilon > 0$ tels que

$$\forall \xi \in [T-\varepsilon, T+\varepsilon],\ \ \frac{f(\xi)}{f(T)} \in B_{(1,\frac{1}{2})},$$

et de même pour le rapport $\frac{f(\xi)}{f(-T)}$ sur l'intervalle $[-T-\varepsilon, -T+\varepsilon]$. Or, le logarithme complexe $\log(1+z)$ est de rayon de convergence 1, donc on peut considérer les séries absolument convergentes

$$h_1(\xi) = \frac{1}{\mathrm{i}}\log\left(\frac{f(\xi)}{f(T)}\right) \qquad ; \qquad h_2(s) = \frac{1}{\mathrm{i}}\log\left(\frac{f(\xi)}{f(-T)}\right)$$

définies respectivement sur $[T-\varepsilon, T+\varepsilon]$ et $[-T-\varepsilon, -T+\varepsilon]$, et qui sont des fonctions continues. Par définition de T, il existe h continue sur $[-T+\varepsilon, T-\varepsilon]$ telle que $h(0)=0$ et $f(\xi)=\exp(\mathrm{i}h(\xi))$ sur l'intervalle $[-T+\varepsilon, T-\varepsilon]$. Notons θ_T et θ_{-T} deux arguments tels que $f(T)=\exp(\mathrm{i}\theta_T)$ et $f(-T)=\exp(\mathrm{i}\theta_{-T})$. Quitte à modifier ces exemples par des multiples de 2π, on a :

$$h(T-\varepsilon)=h_1(T-\varepsilon)+\theta_T;$$
$$h(-T-\varepsilon)=h_2(-T-\varepsilon)+\theta_{-T}.$$

On prolonge alors h sur $[-T-\varepsilon, T+\varepsilon]$ en posant :

$$h(\xi)=\begin{cases} h_1(\xi)+\theta_T & \text{si } \xi\in[T-\varepsilon, T+\varepsilon]; \\ h_2(\xi)+\theta_{-T} & \text{si } \xi\in[-T-\varepsilon, -T+\varepsilon]. \end{cases}$$

On obtient alors une fonction continue sur $[-T-\varepsilon, T+\varepsilon]$, telle que $f(\xi)=\exp(\mathrm{i}h(\xi))$ sur tout cet intervalle ; ceci contredit la définition de T. $\square$

Définition 5.8 (Exposant de Lévy–Khintchine)**.** *On appelle exposant de Lévy–Khintchine d'une loi infiniment divisible μ l'unique fonction continue $\psi:\mathbb{R}\to\mathbb{C}$ telle que $\psi(0)=0$ et $\widehat{\mu}(\xi)=\exp(\psi(\xi))$.*

La formule de Lévy–Khintchine 5.12 permettra de classifier tous les exposants possibles pour une loi infiniment divisible.

Remarque 5.9. L'existence et l'unicité de l'exposant de Lévy–Khintchine implique le résultat non trivial suivant : si μ est infiniment divisible d'exposant ψ, et si $\mu=(\rho_n)^{*n}$ avec $n\geq 1$, alors ρ_n est aussi infiniment divisible, d'exposant $\frac{\psi}{n}$. Fixons $m\geq 1$, et ρ_{nm} telle que $(\rho_{nm})^{*nm}=\mu$. Comme $\widehat{\mu}$ ne s'annule pas, il en va de même pour les transformées de Fourier $\widehat{\rho_n}$ et $\widehat{\rho_{nm}}$, puisque des puissances de celles-ci donnent $\widehat{\mu}$. Il existe donc des fonctions continues ψ_n et ψ_{nm} qui s'annulent en 0 et telles que

$$\widehat{\rho_n}(\xi)=\exp(\psi_n(\xi)) \qquad ; \qquad \widehat{\rho_{nm}}(\xi)=\exp(\psi_{nm}(\xi)).$$

Comme $\widehat{\rho_n}^{\,n}=\widehat{\rho_{nm}}^{\,nm}=\widehat{\mu}$, ceci force

$$\psi_n(\xi)=\frac{\psi(\xi)+2\mathrm{i}\pi k(\xi)}{n} \qquad ; \qquad \psi_{nm}(\xi)=\frac{\psi(\xi)+2\mathrm{i}\pi\, l(\xi)}{nm}$$

et la continuité des exposants ainsi que l'annulation en 0 imposent $k(\xi)=l(\xi)=0$. Ainsi,

$$\psi_n(\xi)=\frac{\psi(\xi)}{n} \qquad ; \qquad \psi_{nm}(\xi)=\frac{\psi(\xi)}{nm}$$

et en particulier, $\widehat{\rho_{nm}}^{\,m}=\widehat{\rho_n}$, ce qui prouve que ρ_n est infiniment divisible, d'exposant de Lévy–Khintchine $\frac{\psi}{n}$.

▷ Convergence de lois infiniment divisibles.

Proposition 5.10 (Stabilité par passage à la limite de l'infini divisibilité)**.** *Soit $(\mu^{(k)})_{k\in\mathbb{N}}$ une suite de lois infiniment divisibles, qui converge en loi vers une mesure de probabilité μ. La loi μ est encore infiniment divisible.*

Preuve. Pour tout entiers $n,k\geq 1$, notons $\rho_n^{(k)}$ une loi telle que $(\rho_n^{(k)})^{*n}=\mu^{(k)}$. Si n est fixé, alors la suite de lois $(\rho_n^{(k)})_{k\in\mathbb{N}}$ est tendue. En effet, notant $X^{(k)}=X_1^{(k)}+\cdots+X_n^{(k)}$ une somme de variables indépendantes de loi $\rho_n^{(k)}$, on a

$$\rho_n^{(k)}([-C,C]^{\mathrm{c}})=\mathbb{P}[|X_1^{(k)}|\geq C]\leq\left(\mathbb{P}[|X^{(k)}|\geq nC]\right)^{\frac{1}{n}}=\left(\rho^{(k)}([-nC,nC]^{\mathrm{c}})\right)^{\frac{1}{n}},$$

donc la tension de $(\rho^{(k)})_{k\in\mathbb{N}}$ implique celle de $(\rho_n^{(k)})_{k\in\mathbb{N}}$. Soit ρ_n une limite d'une sous-suite convergente $(\rho_n^{(k_l)})_{l\in\mathbb{N}}$. On a

$$\begin{aligned}\rho_n^{(k_l)}) &\rightharpoonup \rho_n;\\ \left(\rho_n^{(k_l)})\right)^{*n} &\rightharpoonup (\rho_n)^{*n}\end{aligned}$$

et d'autre part, $\left(\rho_n^{(k_l)})\right)^{*n}=\mu^{(k_l)}$, donc $(\rho_n)^{*n}=\mu$. □

Soit ν une mesure sur $\mathbb{R}$ de masse finie. Une loi de Poisson composée $\mu=P(\nu)$ d'intensité ν est la loi d'un processus de Poisson composé $P_t(\nu)$ au temps $t=1$. Sa transformée de Fourier est

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi P(\nu)}]=\exp\left(\left(\int_{\mathbb{R}}(\mathrm{e}^{\mathrm{i}x\xi}-1)\,\nu(dx)\right)\right),$$

voir l'exemple à la toute fin du chapitre 4. Cette loi est toujours infiniment divisible : $\mu=(\rho_n)^{*n}$ et $\rho_n=P(\frac{\nu}{n})$.

Proposition 5.11. *Une loi μ sur $\mathbb{R}$ est infiniment divisible si et seulement si c'est une limite de lois de Poisson composées.*

Preuve. La proposition précédente montre que les limites de lois de Poisson composées sont effectivement infiniment divisibles. Réciproquement, soit μ une loi infiniment divisible, et ψ son exposant de Lévy–Khintchine. On définit la loi de Poisson composée μ_n par la formule $\mu_n=P(n\rho_n)$, où ρ_n est telle que $(\rho_n)^{*n}=\mu$. Par une remarque précédente, l'exposant de Lévy–Khintchine de ρ_n est $\frac{\psi}{n}$. Par conséquent.

$$\widehat{\mu_n}(\xi)=\exp\left(n\int_{\mathbb{R}}(\mathrm{e}^{\mathrm{i}\xi x}-1)\,\rho_n(dx)\right)=\exp\left(n(\widehat{\rho_n}(\xi)-1)\right)=\exp\left(n(\mathrm{e}^{\frac{\psi(\xi)}{n}}-1)\right).$$

La limite ponctuelle de $\widehat{\mu_n}(\xi)$ est $\exp(\psi(\xi))=\widehat{\mu}(\xi)$, ce qui achève la démonstration. □

▷ Formule de Lévy–Khintchine. D'après la proposition précédente, toute loi infiniment divisible peut être approchée par une loi de Poisson composée d'exposant de Lévy–Khintchine

$$\psi(\xi)=\int_{\mathbb{R}}(\mathrm{e}^{\mathrm{i}\xi x}-1)\,\nu(dx),$$

où ν est une mesure de masse finie sur $\mathbb{R}$. La formule de Lévy–Khintchine précise ce résultat, et décrit ce qui peut apparaître dans l'exposant par passage à limite d'une suite de lois de Poisson composées.

Théorème 5.12 (Formule de Lévy–Khintchine)**.** *Soit μ une loi infiniment divisible sur $\mathbb{R}$, d'exposant ψ. Il existe $a \in \mathbb{R}$, $\sigma^2 \geq 0$ et une mesure borélienne positive ν sur $\mathbb{R}$ telle que*

$$\nu(\{0\}) = 0 \qquad ; \qquad \int_{\mathbb{R}} \min(1, x^2)\, \nu(dx) < +\infty$$

et

$$\psi(\xi) = \mathrm{i}a\xi - \frac{\sigma^2\xi^2}{2} + \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\, \mathrm{i}\xi x)\, \nu(dx).$$

De plus, le triplet (a, σ^2, ν) dans cette représentation est unique, et réciproquement, tout triplet correspond à une loi infiniment divisible. La mesure ν est appelée mesure de Lévy associée à μ.

Remarque 5.13. Les hypothèses sur la mesure de Lévy impose que sa restriction au complémentaire d'un voisinage arbitraire de 0 est de masse finie. Par conséquent, une mesure de Lévy ne peut être de masse infinie que si de la masse s'accumule autour de 0, tout en laissant $\int_{|x|<1} x^2\, \nu(dx)$ finie.

Lemme 5.14. *Tout triplet (a, σ^2, ν) correspond à une loi infiniment divisible.*

Preuve. Il est évident que la convolée de deux lois infiniment divisibles est encore infiniment divisible. Comme la gaussienne $\mathscr{N}(a, \sigma^2)$ a pour exposant $\mathrm{i}a\xi - \frac{\sigma^2\xi^2}{2}$, pour démontrer l'existence d'une loi infiniment divisible de triplet (a, σ^2, ν), on peut donc se ramener au cas où $a = \sigma^2 = 0$. Notons $\nu_n = \nu_{|S_n}$, avec

$$S_n = \begin{cases} (-\infty, -1] \sqcup [1, \infty) & \text{si } n = 0, \\ (-\frac{1}{n}, -\frac{1}{n+1}] \sqcup [\frac{1}{n+1}, \frac{1}{n}) & \text{si } n \geq 1. \end{cases}$$

On a $\nu = \sum_{n\in\mathbb{N}} \nu_n$, et chaque ν_n est une mesure de masse finie. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes, avec X_0 de loi de Poisson composée $P(\nu_0)$, et

$$X_n + \left(\int_{\mathbb{R}} x\, \nu_n(dx)\right) \sim P(\nu_n)$$

si $n \geq 1$. Les variables X_n sont toutes infiniment divisibles, et

$$\mathbb{E}\Big[\mathrm{e}^{\mathrm{i}\xi \sum_{n=0}^N X_n}\Big] = \exp\left(\int_{|x|\geq \frac{1}{N+1}} (\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\, \mathrm{i}\xi x) \nu(dx)\right).$$

Notons que pour tout $\xi \in \mathbb{R}$, $\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\, \mathrm{i}\xi x = O(x^2)$. Comme ν intègre x^2 au voisinage de 0, le terme de droite de l'égalité ci-dessus a une limite lorsque N tend vers l'infini. Par conséquent, la suite $(\mu^{(N)})_{N\in\mathbb{N}}$ des lois infiniment divisibles des variables $S^{(N)} = \sum_{n=0}^N X_n$ converge vers une loi μ, qui est infiniment divisible par la proposition 5.10, et d'exposant de Lévy–Khintchine

$$\psi(\xi) = \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\, \mathrm{i}\xi x) \nu(dx). \qquad \square$$

Remarque 5.15. La preuve fait apparaître une loi infiniment divisible de triplet $(0,0,\nu)$ comme limite des lois de Poisson composées et compensées

$$P\left(\nu_{|\mathbb{R}\setminus(-\frac{1}{N},\frac{1}{N})}\right) - \int_{1>|x|\geq \frac{1}{N}} x\,\nu(dx).$$

Ainsi, toute loi infiniment divisible s'écrit comme convolée d'une loi gaussienne et d'une limite de lois de Poisson composées compensées.

Lemme 5.16. *Si μ loi infiniment divisible a son exposant ψ représenté par un triplet (a,σ^2,ν), alors ce triplet est unique.*

Preuve. Si ψ est représenté par (a,σ^2,μ), alors

$$\frac{\psi(\xi)}{\xi^2} = \frac{\mathrm{i}a}{\xi} - \frac{\sigma^2}{2} + \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\mathrm{i}\xi x}{\xi^2}\,\nu(dx).$$

Dans l'intégrale,

$$\left|\frac{\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\mathrm{i}\xi x}{\xi^2}\right| \leq 1_{|x|<1}\frac{|x|^2}{2} + \frac{2}{|\xi|^2}\,1_{|x|\geq 1}$$

et tend ponctuellement vers 0 lorsque ξ tend vers l'infini, donc par le théorème de convergence dominée,

$$\lim_{\xi\to+\infty} \frac{\psi(\xi)}{\xi^2} = -\frac{\sigma^2}{2}$$

et σ^2 est entièrement déterminé par l'exposant ψ. Posons maintenant $\pi(\xi) = \psi(\xi) + \frac{\sigma^2\xi^2}{2}$, et considérons

$$\int_{-1}^{1} (\pi(\xi) - \pi(\xi+t))\,dt = \int_{\mathbb{R}}\int_{-1}^{1} \mathrm{e}^{\mathrm{i}\xi x}(1-\mathrm{e}^{\mathrm{i}tx})\,\nu(dx)\,dt = 2\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi x}\left(1 - \frac{\sin x}{x}\right)\nu(dx).$$

C'est la transformée de Fourier de la mesure finie positive $\rho(dx) = (1-\frac{\sin x}{x})\,\nu(dx)$, et elle est entièrement déterminée par ψ. On peut donc récupérer ρ, puis ν à partir de ψ. Finalement, on récupère a en soustrayant à ψ les deux autres termes précédemment calculés. □

Dans ce qui suit, $c(x)$ désigne une fonction continue positive sur $\mathbb{R}$, à valeurs dans $[0,1]$, égale à 1 dans un voisinage de 0, et qui s'annule pour $|x| \geq 1$. Étant donné un triplet (a,σ^2,ν), on note $\widetilde{\sigma}^2 = \sigma^2 + \int_{\mathbb{R}} x^2\,c(x)\,\nu(dx)$.

Lemme 5.17. *Soit $(\mu^{(k)})_{k\in\mathbb{N}}$ une suite de lois infiniment divisibles associées à des triplets $(a^{(k)},(\sigma^{(k)})^2,\nu^{(k)})$. Les assertions suivantes sont équivalentes :*

1. *Il existe un triplet (a,σ^2,ν) avec $a^{(k)}\to a$, $(\widetilde{\sigma}^{(k)})^2 \to \widetilde{\sigma}^2$, et pour toute fonction mesurable f qui s'écrit sous la forme $f(x) = \min(1,x^2)\,g(x)$ avec g continue bornée et $g(0)=0$, $\nu^{(k)}(f)\to\nu(f)$.*
2. *La suite $(\mu^{(k)})_{k\in\mathbb{N}}$ converge vers une loi (infiniment divisible) μ.*

De plus, dans ce cas, μ est associée au triplet (a,σ^2,ν).

Preuve. Supposons d'abord la convergence des suites $(a^{(k)})_{k\in\mathbb{N}}$, $((\widetilde{\sigma}^{(k)})^2)_{k\in\mathbb{N}}$ et $(\nu_k)_{k\in\mathbb{N}}$. Pour tout k, on a l'écriture suivante pour l'exposant de Lévy–Khintchine $\psi^{(k)}$ de $\mu^{(k)}$:

$$\begin{aligned}\psi^{(k)}(\xi) &= \mathrm{i}a^{(k)}\xi - \frac{(\sigma^{(k)})^2\xi^2}{2} + \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\mathrm{i}\xi x)\,\nu^{(k)}(dx)\\ &= \mathrm{i}a^{(k)}\xi - \frac{(\widetilde{\sigma}^{(k)})^2\xi^2}{2} + \int_{\mathbb{R}} \left(\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\mathrm{i}\xi x + c(x)\,\frac{\xi^2x^2}{2}\right)\nu^{(k)}(dx).\end{aligned}$$

Comme la fonction intégrée est un $o(x^2)$ au voisinage de 0, elle s'écrit bien sous la forme $\min(1,x^2)\,g(x)$ avec g continue bornée et $g(0)=0$. Par hypothèse,

$$\begin{aligned}\lim_{k\to\infty} \psi^{(k)}(\xi) &= \mathrm{i}a\xi - \frac{(\widetilde{\sigma})^2\xi^2}{2} + \int_{\mathbb{R}} \left(\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\mathrm{i}\xi x + c(x)\frac{\xi^2x^2}{2}\right)\nu(dx)\\ &= \mathrm{i}a\xi - \left(\widetilde{\sigma}^2 - \int_{\mathbb{R}} x^2\,c(x)\,\nu(dx)\right)\frac{\xi^2}{2} + \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}\xi x} - 1 - 1_{|x|<1}\mathrm{i}\xi x)\,\nu(dx).\end{aligned}$$

Si l'on montre que $\sigma^2 = \widetilde{\sigma}^2 - \int_{\mathbb{R}} x^2\,c(x)\,\nu(dx) \geq 0$, alors l'expression ci-dessus sera l'exposant de Lévy–Khintchine d'une certaine loi infiniment divisible μ, et on aura $\mu^{(k)} \rightharpoonup \mu$ puisque les transformées de Fourier convergent. Pour $\varepsilon > 0$, notons c_ε une fonction continue positive, à valeurs dans $[0,1]$, qui vaut 0 en 0, et 1 en dehors d'un intervalle $(-\varepsilon,\varepsilon)$. Notons $\rho^{(k,\varepsilon)}(dx) = c_\varepsilon(x)\,\min(1,x^2)\,\nu^{(k)}(dx)$; ce sont des mesures finies, et par hypothèse, pour toute fonction g continue bornée, $\rho^{(k,\varepsilon)}(g) \to \rho^{(\varepsilon)}(g)$, où $\rho^{(\varepsilon)}(dx) = c_\varepsilon(x)\,\min(1,x^2)\,\nu(dx)$. Autrement dit, $\rho^{(k,\varepsilon)} \rightharpoonup \rho^{(\varepsilon)}$. On peut alors écrire :

$$\begin{aligned}\widetilde{\sigma}^2 = \lim_{k\to\infty} (\widetilde{\sigma}^{(k)})^2 &= \lim_{k\to\infty} \left((\sigma^{(k)})^2 + \int_{\mathbb{R}} c(x)\,\min(1,x^2)\,\nu^{(k)}(dx)\right)\\ &\geq \limsup_{k\to\infty}\left((\sigma^{(k)})^2 + \int_{\mathbb{R}} c(x)\,\rho^{(k,\varepsilon)}(dx)\right) = \left(\limsup_{k\to\infty}(\sigma^{(k)})^2\right) + \int_{\mathbb{R}} c(x)\,\rho^{(\varepsilon)}(dx).\end{aligned}$$

Comme c'est vrai pour tout $\varepsilon > 0$, on obtient :

$$\widetilde{\sigma}^2 \geq \left(\limsup_{k\to\infty}(\sigma^{(k)})^2\right) + \int_{\mathbb{R}} c(x)\,\min(1,x^2)\,\nu(dx) = \left(\limsup_{k\to\infty}(\sigma^{(k)})^2\right) + \int_{\mathbb{R}} c(x)\,x^2\,\nu(dx).$$

On en déduit que $0 \leq \limsup_{k\to\infty}(\sigma_k)^2 \leq \sigma^2$, ce que l'on voulait démontrer. Ceci achève la preuve du sens direct.

Supposons maintenant $\mu^{(k)} \rightharpoonup \mu$; par la proposition 5.10, μ est infiniment divisible, et son exposant de Lévy–Khintchine ψ est la limite (localement uniforme) des exposants de Lévy–Khintchine $\psi^{(k)}$ des lois $\mu^{(k)}$. Notons $\rho^{(k)}(dx) = \min(1,x^2)\,\nu^{(k)}(dx)$; c'est une mesure borélienne finie. On va montrer que la suite $(\rho^{(k)})_{k\in\mathbb{N}}$ est tendue, au sens suivant :

$$\sup_{k\in\mathbb{N}} \rho^{(k)}(\mathbb{R}) < \infty \qquad ; \qquad \lim_{M\to\infty}\sup_{k\in\mathbb{N}} \rho^{(k)}([-M,M]^{\mathrm{c}}) = 0.$$

Par une extension facile du théorème de Prohorov aux suites de mesures finies de masse bornée, ceci impliquera la convergence faible d'une sous-suite $(\rho^{(k_l)})_{l\in\mathbb{N}}$ vers une mesure finie ρ. Remarquons qu'on a l'identité :

$$-\frac{1}{2t}\int_{-t}^{t}\psi^{(k)}(\xi)\,d\xi = \frac{(\sigma^{(k)})^2}{2}\frac{t^2}{3}+\int_{\mathbb{R}}\left(1-\frac{\sin tx}{tx}\right)\nu^{(k)}(dx)$$
$$\geq \int_{\mathbb{R}}\left(1-\frac{\sin tx}{tx}\right)\nu^{(k)}(dx).$$

Avec $t=1$, comme $(1-\frac{\sin x}{x})\geq \frac{1}{6}\min(1,x^2)$ pour tout $x\in\mathbb{R}$, et comme la limite du terme de droite est $-\frac{1}{2}\int_{-1}^{1}\psi(\xi)\,d\xi$, ceci prouve que les lois $\rho^{(k)}$ ont bien leurs masses bornées par une constante. Fixons maintenant $\varepsilon>0$, et $t\in(0,1)$ assez petit tel que $-\frac{1}{2t}\int_{-t}^{t}\psi(\xi)\,d\xi\leq\varepsilon$; un tel t existe car $\psi(0)=0$ et ψ est continue. Alors, pour $k\geq k_0$ assez grand, la même intégrale avec $\psi^{(k)}$ est plus petite que 2ε, donc

$$2\varepsilon\geq\frac{1}{6}\int_{\mathbb{R}}\min(1,t^2x^2)\,\nu^{(k)}(dx)\geq\frac{1}{6}\int_{|x|\geq\frac{1}{t}}\nu^{(k)}(dx).$$

Avec $M=\frac{1}{t}$, ceci implique $\lim_{M\to 0}\sup_{k\in\mathbb{N}}\nu^{(k)}([-M,M]^{\mathrm{c}})=0$, et *a fortiori* pour les mesures $\rho^{(k)}$. On dispose donc d'une sous-suite $\rho^{(k_l)}\rightharpoonup\rho$, et posant $\nu(dx)=\frac{1_{x\neq 0}}{\min(1,x^2)}\,\rho(dx)$, on a alors

$$\nu^{(k_l)}\to\nu$$

au sens de la convergence contre les fonctions de la forme $f(x)=\min(1,x^2)\,g(x)$ avec g continue bornée et $g(0)=0$. La borne

$$-\frac{1}{2}\int_{-1}^{1}\psi^{(k)}(\xi)\,d\xi\geq\frac{(\sigma^{(k)})^2}{6}$$

montre pareillement que la suite $((\sigma^{(k)})^2)_{k\in\mathbb{R}}$ est bornée. On a alors

$$\left(\widetilde{\sigma}^{(k_l)}\right)^2=\left(\sigma^{(k_l)}\right)^2+\int_{\mathbb{R}}c(x)\,x^2\,\nu^{(k_l)}(dx)$$
$$=\left(\sigma^{(k_l)}\right)^2+\int_{\mathbb{R}}c(x)\,\rho^{(k_l)}(dx)\leq\left(\sigma^{(k_l)}\right)^2+\int_{\mathbb{R}}\rho^{(k_l)}(dx)$$

qui reste bornée, donc à extraction près, on peut aussi supposer $(\widetilde{\sigma}^{(k_l)})^2\to\widetilde{\sigma}^2$ pour une certaine constante $\widetilde{\sigma}^2$. La convergence ponctuelle des exposants de Lévy–Khintchine impose alors également la convergence des sous-suites $(a^{(k_l)})_{l\in\mathbb{N}}$ vers une limite a. Par le sens direct, on en déduit que μ est associée à un triplet (a,σ^2,ν). De plus, comme les triplets déterminent entièrement les lois infiniment divisibles correspondantes, le triplet limite possible est unique, ce qui démontre la convergence des triplets. □

Preuve du théorème 5.12. Soit μ une loi infiniment divisible ; il reste seulement à montrer l'existence d'un triplet (a,σ^2,ν) correspondant à μ. D'après la proposition 5.11, il existe une suite de lois de Poisson composées $P(\nu^{(k)})$, $k\geq 0$ telle que $P(\nu^{(k)})\rightharpoonup\mu$. Or, les lois de Poisson composées ont pour triplets $(a^{(k)},0,\nu^{(k)})$, avec

$$a^{(k)}=\int_{|x|\leq 1}x\,\nu^{(k)}(dx).$$

D'après le lemme précédent, μ est donc associée à un certain triplet (a,σ^2,ν). □

Remarque 5.18. Dans la preuve du théorème, remarquons que si $\mu^{(k)} \rightharpoonup \mu$ avec des lois infiniment divisibles, alors $\limsup_{k\to\infty}(\sigma^{(k)})^2 \leq \sigma^2$, c'est-à-dire que la partie gaussienne des lois augmente toujours par passage à la limite. Par exemple, si $\mu^{(k)} = P(k\,\mathcal{N}(0,\frac{1}{k}))$, on sait que $\mu^{(k)} \rightharpoonup \mathcal{N}(0,1)$, et on a par ailleurs

$$(\sigma^{(k)})^2 = 0 \qquad ; \qquad \lim_{k\to\infty} (\widetilde{\sigma}^{(k)})^2 = 1 = \widetilde{\sigma}^2 = \sigma^2.$$

Remarque 5.19. La formule 5.12 met en jeu une fonction de troncation $t(x,\xi) = 1_{|x|<1}\mathrm{i}\xi x$, qui vient compenser la différence $\mathrm{e}^{\mathrm{i}\xi x} - 1$ pour en faire un $O(x^2)$ au voisinage de 0. On peut choisir d'autres fonctions de troncation avec les mêmes propriétés, et qui donnent des formules de Lévy–Khintchine légèrement différentes. Si la mesure de Lévy ν intègre $\min(1,|x|) \geq \min(1,x^2)$, on peut par ailleurs réécrire la formule comme suit :

$$\psi(\xi) = \mathrm{i}m\xi - \frac{\sigma^2\xi^2}{2} + \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}\xi x} - 1)\,\nu(dx),$$

où $m = a - \int_{|x|<1} x\,\nu(dx)$. Cette formule signifie que ψ est l'exposant de la loi μ d'une somme :

— d'une gaussienne $\mathcal{N}(m,\sigma^2)$;

— et d'une loi de Poisson composée indépendante $P(\nu)$.

Si ν n'intègre pas $|x|$ au voisinage de 0, alors la loi de Poisson composée $P(\nu)$ est mal définie, et une telle interprétation n'est pas possible. Ceci explique pourquoi on doit travailler avec des lois compensées.

5.2 Processus de Lévy et décomposition de Lévy–Itô

Dans cette section, on considère un processus de Lévy $(X_t)_{t\in\mathbb{R}_+}$, c'est-à-dire un processus càdlàg à accroissements indépendants et stationnaires. On notera μ_t la loi de X_t. Notons que si $(X_t)_{t\in\mathbb{R}_+}$ est un processus de Lévy, alors $(X_t - X_0)_{t\in\mathbb{R}_+}$ est encore un processus de Lévy, cette fois-ci issu de 0. Sans perte de généralité, on pourra donc supposer $X_0 = 0$ presque sûrement.

▷ **Semi-groupe et exposant d'un processus de Lévy.** Si $t \in \mathbb{R}_+$ et $n \geq 1$, alors

$$X_t = X_{\frac{t}{n}} + \left(X_{\frac{2t}{n}} - X_{\frac{t}{n}}\right) + \cdots + \left(X_t - X_{\frac{(n-1)t}{n}}\right)$$

est la somme de n variables aléatoires indépendantes de même loi $\mu_{\frac{t}{n}}$. Les lois marginales μ_t sont donc toutes infiniment divisibles, et elles forment un semi-groupe de mesures de probabilité sur $\mathbb{R}$, c'est-à-dire une famille $(\mu_t)_{t\in\mathbb{R}_+}$ telle que $\mu_s * \mu_t = \mu_{s+t}$. Soit ψ l'exposant de Lévy–Khintchine de la marginale μ_1.

Proposition 5.20. *Pour tout temps $t \in \mathbb{R}_+$, l'exposant de Lévy–Khintchine de μ_t est $\psi_t(\xi) = t\,\psi(\xi)$.*

Preuve. Si $t = \frac{p}{q}$ est rationnel, alors $(\mu_{\frac{p}{q}})^{*q} = \mu_q$, donc l'exposant $\psi_{\frac{p}{q}}$ vérifie $q\,\psi_{\frac{p}{q}} = \psi_p = p\,\psi$. On a donc $\psi_t = t\,\psi$ pour tout t rationnel. Pour démontrer l'identité pour tout t réel, il suffit maintenant de montrer que

$$\lim_{\varepsilon\to 0} \mu_\varepsilon = \delta_0$$

au sens de la convergence en loi. En effet, choisissant alors une suite $(\frac{p_n}{q_n})_{n\in\mathbb{N}}$ de temps rationnels qui tendent vers t par valeurs supérieures, on aura alors $\lim_{n\to\infty} \mu_{\frac{p_n}{q_n}-t} = \delta_0$, et donc

$$\lim_{n\to\infty} \mu_{\frac{p_n}{q_n}} = \mu_t * \left(\lim_{n\to\infty} \mu_{\frac{p_n}{q_n}-t}\right) = \mu_t * \delta_0 = \mu_t.$$

Ceci est équivalent à la convergence ponctuelle et même localement uniforme des transformées de Fourier, et donc, $\psi_t(\xi) = \lim_{n\to\infty} \psi_{\frac{p_n}{q_n}}(\xi) = t\,\psi(\xi)$.

Soit V un voisinage de 0. La convergence en loi $\mu_\varepsilon \rightharpoonup \delta_0$ est équivalente à $\lim_{\varepsilon\to 0} \mu_\varepsilon(V) = 1$. Or, comme le processus $(X_t)_{t\in\mathbb{R}_+}$ est continu à droite en 0,

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty} \bigcap_{t\leq \frac{1}{n}} \{X_t \in V\}\right] = 1.$$

Soit $\eta > 0$, et n tel que $\varepsilon = \frac{1}{n}$ vérifie

$$\mathbb{P}\left[\bigcap_{t\leq\varepsilon} \{X_t \in V\}\right] \geq 1 - \eta.$$

Alors, pour tout $t \leq \varepsilon$, $\mu_t(V) \geq 1 - \eta$, ce qui achève la preuve. □

Proposition 5.21. *L'exposant de Lévy–Khintchine ψ de la loi μ_1 détermine entièrement la loi du processus $(X_t)_{t\in\mathbb{R}_+}$.*

Preuve. L'exposant ψ détermine toutes les lois fini-dimensionnelles $\mu_{t_1<\cdots<t_d}$, et d'après la discussion du chapitre 3, ceci détermine la loi μ de $(X_t)_{t\in\mathbb{R}_+}$. □

Ainsi, à tout processus de Lévy, on peut associer un triplet (a, σ^2, ν) qui le détermine entièrement (en loi). Le paragraphe suivant montre que réciproquement, tout triplet correspond à un processus de Lévy.

▷ **Décomposition de Lévy–Itô.** Si $(B_t)_{t\in\mathbb{R}_+}$ est un mouvement brownien standard (limite en loi dans $\mathscr{D}(\mathbb{R}_+)$ de marches aléatoires renormalisées et construites à partir de variables i.i.d. comme dans le théorème de Donsker 3.38), alors $(at + \sigma B_t)_{t\in\mathbb{R}_+}$ est un processus de Lévy continu (à valeurs dans $\mathscr{C}(\mathbb{R}_+)$), d'exposant

$$\psi(\xi) = \mathrm{i}a\xi - \frac{\sigma^2\xi^2}{2}.$$

Par conséquent, pour construire un processus de Lévy associé à un triplet (a, σ^2, ν), on peut se ramener au cas où $a = \sigma^2 = 0$, puisque la somme de deux processus de Lévy indépendants d'exposants (a_1, σ_1^2, ν_1) et (a_2, σ_2^2, ν_2) est un processus de Lévy d'exposant

$(a_1+a_2, \sigma_1^2+\sigma_2^2, \nu_1+\nu_2)$. Fixons donc une mesure ν sans atome en 0 et qui intègre $\min(1,x^2)$. On note comme précédemment

$$S_n = \begin{cases} (-\infty,-1] \sqcup [1,\infty) & \text{si } n=0, \\ (-\frac{1}{n},-\frac{1}{n+1}] \sqcup [\frac{1}{n+1},\frac{1}{n}) & \text{si } n \geq 1. \end{cases}$$

et $\nu_n = \nu_{|S_n}$. Chaque ν_n est une mesure finie, et $\nu = \sum_{n=0}^{\infty} \nu_n$. On introduit alors pour chaque n une mesure aléatoire de Poisson $N^{(n)}$ d'intensité $\nu \otimes dt$ sur $\mathbb{R} \times \mathbb{R}_+$, ces mesures étant prises indépendamment. Le support de $N^{(n)}$ est presque sûrement inclus dans $S_n \times \mathbb{R}_+$. On définit les processus de Poisson composés et compensés

$$P_t^{(n)} = \begin{cases} N_t^{(0)}(x) & \text{si } n=0, \\ N_t^{(n)}(x) - t \int_{\mathbb{R}} x\, \nu_n(dx) & \text{si } n \geq 1. \end{cases}$$

Si $n=0$, alors ν_0 n'intègre pas forcément x, et il faut comprendre la définition comme

$$P_t^{(0)} = \sum_{0 \leq s \leq t,\, X_s \neq \dagger} A_s,$$

avec $(N^{(0)})_{|\mathbb{R}\times[0,t]} = \sum_{0\leq s\leq t, X_s \neq \dagger} \delta_{A_s}$, cette somme étant presque sûrement finie. D'autre part, si $n \geq 1$, alors la mesure ν_n intègre x, et on peut bien considérer $N_t^{(n)}(x)$ et compenser par son espérance. Les processus $(P_t^{(n)})_{t\in\mathbb{R}_+}$ sont càdlàg, et même localement constants.

Proposition 5.22. *Soit $M_t^{(n)} = \sum_{m=1}^{n} P_t^{(m)}$. La suite de processus càdlàg $(M^{(n)})_{n\in\mathbb{N}}$ converge en loi dans $\mathscr{D}(\mathbb{R}_+)$ vers un processus de Lévy d'exposant $(0,0,\nu_{|(-1,1)})$.*

Preuve. Notons que chaque processus compensé $P^{(m)}$ avec $m \geq 1$ est une martingale, avec $\mathbb{E}[P_t^{(m)}] = 0$ pour tout temps t. Pour tous temps $s < t$, on peut alors calculer

$$\begin{aligned} \mathbb{E}\big[(M_t^{(n)} - M_s^{(n)})^2\big] &= \sum_{m=1}^{n} \mathbb{E}\big[(P_t^{(m)} - P_s^{(m)})^2\big] = (t-s) \sum_{m=1}^{n} \int_{\mathbb{R}} x^2\, \nu_m(dx) \\ &= (t-s) \int_{\frac{1}{n+1} \leq |x| < 1} x^2\, \nu(dx) \leq (t-s) \int_{(-1,1)} x^2\, \nu(dx). \end{aligned}$$

En effet, la transformée de Fourier de $P_t^{(m)} - P_s^{(m)}$ est $\mathrm{e}^{(t-s)\int_{\mathbb{R}}(\mathrm{e}^{\mathrm{i}\xi x}-1-\mathrm{i}\xi x)\,\nu_m(dx)}$, et son développement de Taylor en $\xi = 0$ permet de calculer le second moment $(t-s)\int_{\mathbb{R}} x^2\, \nu_m(dx)$. Notons $C = \int_{(-1,1)} x^2\, \nu(dx)$. On a donc, pour tout $r \leq s \leq t$,

$$\sup_{n\in\mathbb{N}} \mathbb{E}\big[(M_t^{(n)} - M_s^{(n)})^2 (M_s^{(n)} - M_r^{(n)})^2\big] \leq C^2\,(t-s)(s-r) \leq C^2\,(t-r)^2,$$

ce qui permet d'utiliser le critère de tension 3.34. En effet, on a par ailleurs convergence des lois fini-dimensionnelles des processus $(M_t^{(n)})_{t\in\mathbb{R}_+}$ vers les lois fini-dimensionnelles d'un processus dans $\mathbb{R}^{\mathbb{R}_+}$ à accroissements indépendants et stationnaires, d'exposant de Lévy–Khintchine $\psi(\xi) = \int_{(-1,1)} (\mathrm{e}^{\mathrm{i}\xi x} - 1 - \mathrm{i}\xi x)\,\nu(dx)$. Le critère de tension assure que cette limite est bien donnée par une loi sur $\mathscr{D}(\mathbb{R}_+)$, et qu'on a convergence en loi dans cet espace. $\square$

Remarque 5.23. En utilisant l'inégalité maximale de Doob pour les martingales, on peut en fait montrer une convergence beaucoup plus forte, à savoir presque sûrement uniformément sur les compacts. En particulier, les séries $\sum_{m=1}^{\infty} P_t^{(m)}$ sont bien définies presque sûrement, et elles correspondent à une martingale càdlàg de carré intégrable. Pour tout $p \geq 2$, on a aussi convergence dans $\mathscr{L}^p(\Omega, \mathscr{F}, \mathbb{P})$ des séries aléatoires $\sum_{m=1}^{\infty} P_t^{(m)}$.

La proposition précédente et la remarque qui suit et dont on admettra le résultat mène directement à :

Théorème 5.24 (Lévy–Itô)**.** *Soit (a, σ^2, ν) un triplet comme dans la formule de Khintchine. Il existe un processus de Lévy $(X_t)_{t\in\mathbb{R}_+}$ associé à ce triplet, et de plus, la loi de $(X_t)_{t\in\mathbb{R}_+}$ est celle d'une somme*

$$at + \sigma B_t + P_t^{(0)} + \sum_{m=1}^{\infty} P_t^{(m)},$$

avec :

- — *$(B_t)_{t\in\mathbb{R}_+}$ est un mouvement brownien standard ;*
- — *$(P_t^{(0)})_{t\in\mathbb{R}_+}$ est un processus de Poisson composé d'intensité $\nu_0 = \nu_{|S_0}$;*
- — *$(P_t^{(m)})_{t\in\mathbb{R}_+}$ est un processus de Poisson composé d'intensité $\nu_m = \nu_{|S_m}$, compensé par sa moyenne $t \int_{\mathbb{R}} x\, \nu_m(dx)$;*

tous ces termes étant indépendants, et la série étant uniformément convergente sur les compacts.

Corollaire 5.25. *Soit $(X_t)_{t\in\mathbb{R}_+}$ un processus de Lévy d'exposant (a, σ^2, ν). On considère la mesure atomique aléatoire*

$$N_X = \sum_{t \,|\, \Delta X_t \neq 0} \delta_{(\Delta X_t, t)},$$

où $\Delta X_t = X_t - X_{t\_}$, qui est non nul si et seulement si X a une discontinuité au temps t. Cet élément aléatoire de $\mathscr{M}^{\mathrm{atom}}(\mathbb{R} \times \mathbb{R}_+)$ est une mesure aléatoire de Poisson d'intensité $\nu \otimes dt$.

Dans l'énoncé du corollaire, on utilise le fait que l'application

$$\begin{aligned} \mathscr{D}(\mathbb{R}_+) &\to \mathscr{M}^{\mathrm{atom}}(\mathbb{R} \times \mathbb{R}_+) \\ X &\mapsto N_X \end{aligned}$$

est une application mesurable pour la topologie de Skorohod.

Une application de la décomposition de Lévy–Itô est l'étude de la régularité des processus de Lévy. Pour commencer, on voit qu'un processus de Lévy est continu presque sûrement si et seulement si c'est un mouvement brownien avec drift $at + \sigma B_t$; sinon, $\nu \neq 0$ et l'ensemble des sauts de X est presque sûrement dénombrable infini. Supposons maintenant $\nu \neq 0$. Si ν n'intègre pas $|x|$ au voisinage de 0, alors la somme des sauts $\sum_{s\leq u\leq t} |\Delta X_u|$ est infinie presque sûrement sur tout intervalle $[s, t]$. Ceci implique que X n'est de variation bornée sur aucun intervalle. Si ν intègre $|x|$ au voisinage de 0, alors au contraire la somme des sauts est convergente presque sûrement, et dans ce cas X est localement à variation bornée si et seulement si sa partie brownienne est nulle. Ainsi :

Proposition 5.26. *Un processus de Lévy est localement à variation bornée si et seulement si* $\sigma = 0$ *et* $\int_{\mathbb{R}} \min(1, |x|)\, \nu(dx) < +\infty$.

Remarque 5.27. La discussion précédente montre que si ν n'intègre pas $|x|$ au voisinage de 0, alors ses sauts ne sont pas sommables. Il est donc inexact de considérer les processus de Lévy comme "mouvements browniens auxquels on ajoute des sauts" : le cas général est plus compliqué que cette intuition.

5.3 Subordinateurs

Pour conclure notre étude des processus de Lévy, on s'intéresse aux processus de Lévy croissants.

Définition 5.28 (Subordinateur)**.** *On appelle subordinateur un processus de Lévy* $(X_t)_{t\geq 0}$ *qui est croissant presque sûrement.*

La correspondance entre processus de Lévy et lois infiniment divisibles fait correspondre les subordinateurs et les lois infiniment divisibles à support positif, inclus dans $\mathbb{R}_+$. Un processus croissant est nécessairement localement à variation bornée, puisque pour tout intervalle $[s,t]$

$$\sup_{s=t_0<t_1<\cdots<t_n=t} \sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}| = X_t - X_s < +\infty.$$

D'après ce qui précède, l'exposant de Lévy–Khintchine de $(X_t)_{t\in\mathbb{R}_+}$ est donc forcément de la forme $(a, 0, \nu)$ avec la mesure de Lévy ν qui intègre $\min(1, |x|)$. Sous ces hypothèses, la compensation $t \int_{|x|<1} x\, \nu(dx)$ est finie, et la somme de processus de Poisson composés $\sum_{m=1}^{\infty} N_t^{(m)}(x)$ est absolument convergente presque sûrement. On peut donc réécrire la décomposition de Lévy–Itô sous la forme :

$$X_t = mt + P_t^{(0)} + Q_t$$

avec :

- $m = a - \int_{|x|<1} x\, \nu(dx)$;
- $P_t^{(0)} = \sum_{0\leq s\leq t, A_s\neq \dagger} A_s$ processus de Poisson composé d'intensité $\nu_0 = \nu_{|S_0}$, cette somme ayant presque sûrement un nombre fini de termes ;
- $Q_t = N_t(x)$ processus de Poisson composé d'intensité $\nu_{|(-1,1)}$, la somme correspondante étant cette fois-ci une série absolument convergente.

Lemme 5.29. *Le coefficient* m *d'un subordinateur est positif ou nul.*

Preuve. Supposons par l'absurde $m < 0$, et fixons $\varepsilon > 0$ suffisamment petit, tel que

$$\int_{|x|<\varepsilon} |x|\, \nu(dx) \leq -\frac{m}{2}.$$

On peut alors réécrire la décomposition de Lévy–Itô sous la forme suivante :

$$X_t = mt + D_t + E_t$$

avec $D_t = \sum_{0 \leq s \leq t, A_s \neq \dagger} A_s$ processus de Poisson composé d'intensité $\nu_{|\mathbb{R}\setminus(-\varepsilon,\varepsilon)}$, cette somme ayant presque sûrement un nombre fini de termes ; et $(E_t)_{t\in\mathbb{R}_+}$ processus de Poisson composé d'intensité $\nu_{|(-\varepsilon,\varepsilon)}$, indépendant de $(D_t)_{t\in\mathbb{R}_+}$. Dans cette décomposition, la variable D_1 n'est pas forcément intégrable, mais son nombre de termes suit une loi de Poisson de paramètre $\lambda = \nu(\mathbb{R} \setminus (-\varepsilon, \varepsilon))$, qui est fini. Par conséquent, $\mathbb{P}[D_1 = 0] \geq \mathrm{e}^{-\lambda} = p > 0$. D'autre part, la variable E_1 est elle intégrable, d'espérance

$$\mathbb{E}[E_1] = \int_{|x|<\varepsilon} x\, \nu(dx) \leq -\frac{m}{2}.$$

Ceci implique qu'avec probabilité $q > 0$, $E_1 \leq -\frac{m}{2}$. Alors, avec probabilité au moins $pq > 0$, on a

$$X_1 \leq m + 0 - \frac{m}{2} < 0,$$

ce qui est impossible car $X_1 \geq X_0 = 0$ avec probabilité 1. □

Lemme 5.30. *La mesure de Lévy ν d'un subordinateur est supportée par $\mathbb{R}_+$.*

Preuve. Raisonnons de nouveau par l'absurde, et supposons $\nu((-\infty, -\varepsilon]) = \eta > 0$ pour un certain $\varepsilon > 0$. On décompose $X_t = mt + D_t + E_t$ comme dans la preuve précédente. Soit C une constante positive telle que $\mathbb{P}[m + E_1 \leq C] = p > 0$. On a d'autre part

$$D_1 = \sum_{i=1}^{N} A_i$$

où les A_i sont indépendantes et suivent une loi $\frac{\nu_{|\mathbb{R}\setminus(-\varepsilon,\varepsilon)}(dx)}{\nu(\mathbb{R}\setminus(-\varepsilon,\varepsilon))}$, et où N suit une loi de Poisson de paramètre $\nu(\mathbb{R} \setminus (-\varepsilon, \varepsilon))$. Avec probabilité q non nulle, N prend une valeur $n \geq \frac{2C}{\varepsilon}$, et avec probabilité η^n, toutes les variables $A_{i\leq n}$ sont inférieures à ε, donc

$$\mathbb{P}[D_1 \leq -2C] \geq q\eta^n > 0.$$

On conclut qu'avec probabilité plus grande que $pq\eta^n > 0$, $X_1 = m + D_1 + E_1 \leq -C < 0$, ce qui est absurde comme dans le preuve du lemme précédent. □

Réciproquement, si $m \geq 0$ et ν est supportée par $\mathbb{R}_+$, la décomposition de Lévy–Itô prouve immédiatement que $(X_t)_{t\in\mathbb{R}_+}$ est croissant presque sûrement, donc :

Théorème 5.31. *Les subordinateurs sont les processus de Lévy dont l'exposant de Lévy–Khintchine s'écrit*

$$\psi(\xi) = \mathrm{i}m\xi + \int_{\mathbb{R}_+} (\mathrm{e}^{\mathrm{i}\xi x} - 1)\, \nu(dx)$$

avec $m \geq 0$ et ν supportée par $\mathbb{R}_+^*$ et intégrant $\min(1, x)$. En termes de processus, un subordinateur $(X_t)_{t\in\mathbb{R}_+}$ avec cet exposant s'écrit $X_t = mt + P_t(\nu)$ avec $(P_t(\nu))_{t\in\mathbb{R}_+}$ processus de Poisson composé.

Remarque 5.32. Si $(X_t)_{t\geq 0}$ est un subordinateur, alors on peut considérer plus généralement la transformée de Laplace $\mathbb{E}[\mathrm{e}^{-\lambda X_t}]$, bien définie pour tout $\lambda \in \mathbb{R}_+$, ou plus généralement $\lambda \in \mathbb{C}$ avec $\mathrm{Re}(\lambda) \geq 0$. La représentation de $(X_t)_{t\geq 0}$ à l'aide d'un processus de Poisson composé démontre que

$$\mathbb{E}[\mathrm{e}^{-\lambda X_t}] = \mathrm{e}^{-\lambda m t + t \int_{\mathbb{R}+} (\mathrm{e}^{-\lambda x} - 1)\, \nu(dx)}.$$

L'utilisation des transformées de Laplace au lieu des transformées de Poisson est classique dans ce cadre.

Remarque 5.33. Supposons ν de masse infinie dans la représentation d'un subordinateur. Alors, comme les sauts de $(X_t)_{t\geq 0}$ forment un processus ponctuel de Poisson d'intensité ν, l'ensemble des instants de sauts de $(X_t)_{t\geq 0}$ est dense dans $\mathbb{R}$; néanmoins, la somme de ces sauts est finie sur chaque intervalle compact.

Références

On a essentiellement suivi [K.-I. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge studies in advanced mathematics vol. 68, Cambridge University Press, 1999], et on renvoie aussi à [J. Bertoin, *Lévy Processes*, Cambridge Tracts in Mathematics vol. 121, Cambridge University Press, 1998]. Pour la théorie des subordinateurs, un survol est proposé dans [J. Bertoin, *Subordinators : Examples and Applications* in *Lectures on Probability Theory and Statistics, École d'été de Probabilités de Saint-Flour XXVII, 1997*, Lecture Notes in Mathematics vol. 1717, Springer-Verlag, 1999]. On n'a pas du tout abordé la théorie importante des processus stables, ni les propriétés markoviennes des processus de Lévy ; on renvoie aux ouvrages précités pour ces sujets.

Exercices

1. Vérifier que la convolée de lois sur $\mathbb{R}$ est une opération continue de $\mathscr{M}^1(\mathbb{R}) \times \mathscr{M}^1(\mathbb{R})$ vers $\mathscr{M}^1(\mathbb{R})$ (c'est une propriété qu'on utilise implicitement plusieurs fois dans ce chapitre).

2. Soit μ une loi infiniment divisible de triplet (a, σ^2, ν).
   (a) À quelle condition μ admet-elle un premier moment ? Montrer que dans ce cas,
   $$\int_{\mathbb{R}} x\, \mu(dx) = a + \int_{|x|\geq 1} x\, \nu(dx).$$
   (b) À quelle condition μ admet-elle un second moment ? Montrer que dans ce cas,
   $$\int_{\mathbb{R}} x^2\, \mu(dx) - \left(\int_{\mathbb{R}} x\, \mu(dx)\right)^2 = \sigma^2 + \int_{\mathbb{R}} x^2\, \nu(dx).$$

3. La loi de Cauchy $\mathscr{C}(c)$ de paramètre $c > 0$ est définie par la densité
   $$\frac{1}{c\pi\,(1+\frac{x^2}{c^2})}\, dx.$$
   (a) Montrer que la transformée de Fourier de $\mu \sim \mathscr{C}(c)$ est $\widehat{\mu}(\xi) = \mathrm{e}^{-c|\xi|}$. En déduire que la loi de Cauchy est infiniment divisible.
   (b) Calculer le triplet de la loi de Cauchy standard $\mathscr{C}(1)$ (indication : calculer la limite de $n\rho_n$, avec $(\rho_n)^{*n} = \mu \sim \mathscr{C}(1)$).

4. Soit $(B_t)_{t\geq 0}$ un mouvement brownien standard, et $T_x = \inf\{t > 0 \,|\, B_t > x\}$.
   (a) Montrer que $(T_x)_{x\geq 0}$ est un subordinateur.
   (b) Montrer que T_1 suit une loi de densité $1_{s>0}\, \frac{1}{\sqrt{2\pi}\, s^{3/2}}\, \mathrm{e}^{-\frac{1}{2s}}\, ds$. C'est la distribution de Lévy standard.
   (c) Montrer que la transformée de Laplace de T_1 est $\mathbb{E}[\mathrm{e}^{-\lambda T_1}] = \mathrm{e}^{-\sqrt{2\lambda}}$. On pourra utiliser le fait que $(\mathrm{e}^{\lambda B_t - \frac{\lambda^2 t}{2}})_{t\geq 0}$ est une martingale, et que T_1 est un temps d'arrêt pour cette martingale.
   (d) Montrer que $(T_x)_{x\geq 0}$ est un processus de Poisson composé d'intensité la mesure $1_{v>0}\, \sqrt{\frac{1}{2\pi v^3}}\, dv$.

5. Soit $(B_t)_{t\geq 0}$ et $(C_t)_{t\geq 0}$ deux mouvements browniens standards indépendants, et $(T_x)_{x\geq 0}$ le subordinateur défini comme dans l'exercice précédent. On pose $D_x = C_{T_x}$.
   (a) Donner une interprétation géométrique de D_x en termes du mouvement brownien plan $(B_t, C_t)_{t\geq 0}$.
   (b) Montrer que $(D_x)_{x\geq 0}$ est un processus de Lévy.

(c) Montrer que $(D_x)_{x\geq 0}$ a pour exposant de Lévy–Khintchine $\psi(\xi) = -|\xi|$. On pourra utiliser la martingale complexe $(\exp(\mathrm{i}\xi C_t + |\xi| B_t))_{t\geq 0}$ et le temps d'arrêt T_x. Quelle est la loi de D_x ?

6. Soit $(X_t)_{t\in\mathbb{R}_+}$ un subordinateur. Donner des conditions simples sur son exposant de Lévy–Khintchine pour que le processus soit strictement croissant avec probabilité 1.

7. Soit $(X_t)_{t\in\mathbb{R}_+}$ un subordinateur d'exposant $\psi(\xi) = \int_{\mathbb{R}_+} (\mathrm{e}^{\mathrm{i}\xi} - 1)\, \nu(dx)$; autrement dit, $m = 0$ et $(X_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson composé d'intensité ν supportée par $\mathbb{R}_+$. On suppose également ν de masse finie. Montrer qu'on a l'identité en loi
$$(X_t)_{t\in\mathbb{R}_+} = (W_{P_t})_{t\in\mathbb{R}_+}$$
où $(P_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson d'intensité $\nu(\mathbb{R}_+)$, et où $(W_n)_{n\in\mathbb{N}}$ est une marche aléatoire dont les accroissements $W_n - W_{n-1}$ sont indépendants et suivent la loi $\frac{\nu(\cdot)}{\nu(\mathbb{R}_+)}$.

8. Donner la représentation de Lévy–Khintchine d'une loi $\Gamma(\alpha, c)$.

9. Soit $(X_t)_{t\in\mathbb{R}_+}$ un subordinateur d'exposant de Lévy–Khintchine $\mathrm{i}m\xi + \int_{\mathbb{R}_+} (\mathrm{e}^{\mathrm{i}\xi x} - 1)\, \nu(dx)$.
   (a) Montrer que lorsque t tend vers l'infini,
   $$\frac{X_t}{t} \to_{\mathbb{P}} m + \int_{\mathbb{R}_+} x\, \nu(dx),$$
   y compris si l'intégrale $\int_{\mathbb{R}_+} x\, \nu(dx)$ est infinie.
   (b) On suppose $\int_{\mathbb{R}_+} x\, \nu(dx) < \infty$ et $\int_{\mathbb{R}_+} x^2\, \nu(dx) < \infty$. Montrer que lorsque t tend vers l'infini,
   $$\sqrt{t}\left(\frac{X_t}{t} - m - \int_{\mathbb{R}_+} x\, \nu(dx)\right)$$
   converge en loi vers une gaussienne, dont on précisera la variance.
   (c) Sous les mêmes hypothèses, établir la convergence en loi des processus càdlàg
   $$X^{(n)}(t) = \sqrt{n}\left(\frac{X_{nt}}{n} - tm - t\int_{\mathbb{R}_+} x\, \nu(dx)\right)$$
   vers un certain processus continu qu'on précisera (indication : utiliser le critère de Kolmogorov discontinu).

10. Soit μ une loi infiniment divisible, de triplet $(a, 0, \nu)$ avec
$$\nu = \frac{c_+\, 1_{x>0}}{x^{1+\alpha}}\, dx + \frac{c_-\, 1_{x<0}}{x^{1+\alpha}}\, dx, \quad c_+ + c_- > 0.$$
   (a) À quelles conditions sur α une telle mesure donne-t-elle effectivement une loi infiniment divisible ?

(b) Donner une condition simple sur a pour que le processus de Lévy $(X_t)_{t\geq 0}$ avec $X_1 \sim \mu$ vérifie la propriété :

$$\forall b > 0,\ (X_{bt})_{t\geq 0} = (b^{\frac{1}{a}} X_t)_{t\geq 0}$$

au sens de l'égalité en loi dans $\mathscr{D}(\mathbb{R}_+)$. On parle alors de processus stable d'exposant α.

Table des matières