

Chapitre 2

Convergence de mesures dans des espaces polonais

Nous allons maintenant étudier la convergence en loi lorsque nous avons à disposition une suite de variables aléatoires $(X_n)_{n \geq 1}$ à valeurs dans un espace polonais E . A ce propos, rappelons qu'un espace polonais désigne un espace métrique complet et séparable. Dans ce qui va suivre, E sera souvent un espace de Banach ou de Hilbert de dimension infinie. De manière alternative, nous allons nous focaliser sur la convergence des lois $(\mu_n)_{n \geq 1}$ des variables aléatoires $(X_n)_{n \geq 1}$. Pour plus de détails, nous renvoyons le lecteur vers [38].

2.1 Différences avec la dimension finie

Passons en revue certaines caractérisations de la convergence en loi que nous avons évoqué au chapitre précédent. En dimension infinie, le théorème d'Helly-Bray n'est, bien entendu, plus satisfait. Dans un cadre aussi général, la transformée de Fourier d'une mesure de probabilité μ peut se définir par

$$\phi_\mu(\xi) = \mathbb{E}[e^{i\langle \xi, X \rangle}] = \int_E e^{i\langle \xi, x \rangle} d\mu(x) \quad \text{avec } \xi \in E'$$

où $\langle \cdot, \cdot \rangle$ désigne le crochet de dualité entre E et E' . Malheureusement, la convergence des transformées de Fourier ne caractérise plus la convergence étroite. Ceci provient de l'absence, en dimension infinie, d'une inégalité similaire à celle de Lévy (cf. Lemme 1.4.1). Nous allons donc nous cantonner à la définition classique de la convergence en loi via les fonctions continues et bornées :

$$\mu_n \xrightarrow{n \rightarrow +\infty} \mu \quad \Longleftrightarrow \quad \lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)], \quad \forall f \in C_b(E) \quad \Longleftrightarrow \quad \lim_{n \rightarrow +\infty} X_n \stackrel{\mathcal{L}}{=} X.$$

2.2 Convergence étroite de mesures de probabilités

Dans tout ce qui va suivre (E, d) désignera un espace métrique, complet et séparable. Nous munirons E de sa tribu borélienne $\mathcal{B}(E)$ et si $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{B}(E))$ est une variables aléatoire,

nous noterons sa loi par μ_X . L'ensemble des mesures de probabilités sur $(E, \mathcal{B}(E))$ sera noté $\mathcal{P}(E)$.

2.2.1 Propriétés des éléments de $\mathcal{P}(E)$

Passons en revue quelques propriétés de cet ensemble de mesures de probabilités.

Proposition 2.2.1 (Régularité). *Soit $\mu \in \mathcal{P}(E)$ alors, pour tout boréliens $B \in \mathcal{B}(E)$, nous avons*

$$\begin{aligned}\mu(B) &= \inf\{\mu(O) \text{ avec } O \text{ un ouvert tel que } B \subset O\} \\ &= \sup\{\mu(F) \text{ avec } F \text{ un fermé tel que } F \subset B\}\end{aligned}$$

Démonstration. Voyons une esquisse de la démonstration. Tout d'abord, considérons l'ensemble

$$\mathcal{B}_1 = \{B \subset E ; \inf_{F \subset B \subset O} \mu(O \setminus F) = 0\}$$

avec F un ensemble fermé et O un ensemble ouvert de E . Il n'est pas difficile de montrer que $\mathcal{B}_1$ est une tribu sur E . Montrons à présent que $\mathcal{B}_1$ contient les ensembles fermés de E . A cet effet, soit F un fermé et pour tout $\delta > 0$ considérons son épaissi F_δ défini par

$$F_\delta = \{x \in E, d(x, F) < \delta\}$$

Observons au passage que F_δ est un ouvert de E . De plus, $\cap_{n \geq 1} F_{\frac{1}{n}} = F$. En particulier,

$$\lim_{n \rightarrow +\infty} \mu(F_{\frac{1}{n}}) = \mu(\cap_{n \geq 1} F_{\frac{1}{n}}) = \mu(F)$$

Donc $F \in \mathcal{B}_1$ puisque $F \subset F \subset F_{\frac{1}{n}}$ et $\lim_{n \rightarrow +\infty} \mu(F_{\frac{1}{n}} \setminus F) = 0$ par ce qui précède. En conséquence, $\mathcal{B}(E) \subset \mathcal{B}_1$.

Soit $B \in \mathcal{B}(E)$, selon supra nous avons $B \in \mathcal{B}_1$. C'est pourquoi, pour tout $\epsilon > 0$, il existe un fermé F et un ouvert O tels que $F \subset B \subset O$ et $\mu(O \setminus F) \leq \epsilon$. Ainsi, par inclusion, nous avons

$$\mu(O) \geq \mu(B) \geq \mu(F) \geq \mu(O) - \epsilon \quad \text{et} \quad \mu(F) \leq \mu(B) \leq \mu(O) \leq \mu(F) + \epsilon$$

ce qui montre bien que $\mu(B)$ peut être obtenu comme un infimum sur des ensembles ouverts et un supremum sur des ensembles fermés. $\square$

En dimension finie, nous avons observé qu'une mesure de probabilité était une partie tendue de $\mathcal{P}(E)$. C'est encore le cas dans ce contexte plus général.

Proposition 2.2.2. *Soit (E, d) un espace polonais et $\mu \in \mathcal{P}(E)$; alors μ est tendue au sens suivant : pour tout $\epsilon > 0$ il existe un compact K de E tel que*

$$\mu(K) \geq 1 - \epsilon$$

Remarque. La difficulté sera donc de construire le compact K . En effet, en dimension infinie, ce genre d'ensemble n'est plus caractérisé par les ensembles fermés et bornés. Nous allons donc profiter de la structure d'espace métrique pour laquelle les ensembles compacts sont caractérisés comme étant des ensembles complets et précompacts.

Démonstration. Fixons $\epsilon > 0$ pour le reste de la démonstration. Par séparabilité de (E, d) il existe $(x_n)_{n \in \mathbb{N}}$ une suite de points de E dense dans E . Autrement dit, pour tout $\delta > 0$,

$$\bigcup_{n \in \mathbb{N}} \overline{B(x_n, \delta)} = E$$

Choisissons à présent $\delta = \frac{1}{k}$ avec $k \geq 1$. Puisque $\mu(\bigcup_{n \in \mathbb{N}} \overline{B(x_n, \delta)}) = \mu(E) = 1$, par convergence monotone, la propriété suivante est satisfaite

$$\forall k \geq 1, \exists N_k \in \mathbb{N} \text{ tel que } \mu(\bigcup_{n \leq N_k} \overline{B(x_n, \delta)}) \geq 1 - \frac{\epsilon}{2^k}.$$

Posons alors $K = \bigcap_{k \geq 1} \bigcup_{n \leq N_k} \overline{B(x_n, \delta)}$ et montrons que K est un ensemble compact. Tout d'abord, K est fermé dans (E, d) puisqu'il est défini comme l'intersection d'ensembles fermés. Or (E, d) est complet donc (K, d) l'est également. Pour conclure, il ne reste plus qu'à montrer que K est précompact. Dans ce but, considérons $\eta > 0$ et observons qu'il existe $k_0 \in \mathbb{N}$ tel que $\frac{1}{k} < \eta$ pour tout $k \geq k_0$. Ainsi,

$$K = \bigcap_{k \geq 1} \bigcup_{n \leq N_k} \overline{B(x_n, \delta)} \subset \bigcup_{k=1}^{N_{k_0}} \overline{B(x_n, \frac{1}{k})} \subset \bigcup_{k=1}^{N_{k_0}} \overline{B(x_n, \eta)}$$

En conclusion K est précompact et complet, il s'agit donc d'un sous-ensemble compact de E . Vérifions qu'il satisfait la propriété voulue :

$$\begin{aligned} \mu(K^c) &= \mu\left(\bigcup_{k \geq 1} \left(\bigcup_{n \leq N_k} \overline{B(x_n, \frac{1}{k})}\right)^c\right) \\ &\leq \sum_{k \geq 1} \mu\left(\left(\bigcup_{n \leq N_k} \overline{B(x_n, \frac{1}{k})}\right)^c\right) \\ &\leq \sum_{k \geq 1} \frac{\epsilon}{2^k} = \epsilon \end{aligned}$$

□

Le résultat précédent permet de montrer qu'une mesure $\mu \in \mathcal{P}(E)$ vérifie aussi une propriété de régularité vis-à-vis des ensembles compacts de (E, d) .

Corollaire 2.2.1. *Soit (E, d) un espace polonais et $\mu \in \mathcal{P}(E)$ alors μ est régulière par rapport aux compacts. Autrement dit, pour tout boréliens $B \in \mathcal{B}(E)$, nous avons*

$$\mu(B) = \sup\{\mu(K) \text{ avec } K \text{ un compact tel que } K \subset B\}$$

Démonstration. En utilisant la régularité de la mesure μ par rapport aux ensembles fermés, nous savons que pour tout $\epsilon > 0$ il existe un fermé $F \subset B$ tel que

$$\mu(F) \leq \mu(B) \leq \mu(F) + \epsilon$$

La mesure μ étant également tendue, il existe un compact K de E tel que

$$\mu(K) \geq 1 - \epsilon$$

De ceci, nous en déduisons les inégalités suivantes. D'une part, puisque $F \subset B$,

$$\mu(F \cap K) \leq \mu(F) \leq \mu(B);$$

d'autre part,

$$\begin{aligned} \mu(B) \leq \mu(F) + \epsilon &\leq \mu(F \cap K) + \mu(F \cap K^c) + \epsilon \\ &\leq \mu(F \cap K) + \mu(K^c) + \epsilon \\ &\leq \mu(F \cap K) + 2\epsilon \end{aligned}$$

En résumé, nous avons montré que

$$\mu(F \cap K) \leq \mu(B) \leq \mu(F \cap K) + 2\epsilon.$$

Il suffit d'observer que $F \cap K$ est un compact pour conclure. $\square$

2.3 Rappels sur la hiérarchie des modes de convergence

Le lecteur a déjà du rencontrer des modes de convergence plus fortes que celle de la convergence étroite (i.e. convergence en loi). Par exemple la convergence *presque sûr* (en abrégé *p.s.*) de variables aléatoires entraîne la convergence en *probabilité* de ces variables aléatoires, qui elle-même entraîne la convergence en loi de celles-ci.

Rappelons la définition de ces différents modes de convergence. Dans un espace métrique (E, d) la convergence en probabilité d'une suite de variables aléatoires $(X_n)_{n \geq 0}$ vers une variable aléatoire X s'exprime de la manière suivante.

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(d(X_n, X) > \epsilon) = 0$$

Nous noterons ceci par $\lim_{n \rightarrow +\infty} X_n \stackrel{\mathbb{P}}{=} X$. Dans un même cadre, la convergence presque sûre, s'exprime comme suit : X_n converge *p.s.* vers X s'il existe un ensemble $\mathcal{N} \in \mathcal{B}(E)$ de mesure nulle tel que

$$\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega), \quad \text{pour tout } \omega \in \mathcal{N}^c$$

Nous noterons ceci par $\lim_{n \rightarrow +\infty} X_n = X$ *p.s.*. Par la suite, il pourra être utile d'avoir en tête les résultats suivants. Le premier explique qu'il est possible d'obtenir de la convergence presque sûr à partir d'une convergence en probabilité (à condition de se restreindre à une sous-suite) ; le deuxième montre que la notion de convergence en loi et de convergence en probabilité sont équivalentes lorsque la limite est une constante.

Proposition 2.3.1. *Dans le cadre précédent, les assertions suivantes sont vérifiées.*

1. Si $X_n \xrightarrow{\mathbb{P}} X$ lorsque $n \rightarrow +\infty$ alors il existe une sous-suite $(n_k)_{k \in \mathbb{N}}$ telle que $X_{n_k} \rightarrow X$ *p.s.* lorsque $k \rightarrow +\infty$.
2. Soit $p \in E$. Si $X_n \xrightarrow{\mathcal{L}} X$ lorsque $n \rightarrow +\infty$ avec $\mathcal{L}(X) = \delta_p$ alors $X_n \xrightarrow{\mathbb{P}} X$ lorsque $n \rightarrow +\infty$.

La démonstration de cette proposition est laissée en exercice. Utilisons la première assertion de ce résultat pour démontrer que la convergence en probabilité entraîne la convergence en loi.

Proposition 2.3.2. *Dans le cadre précédent, si $X_n \xrightarrow{\mathbb{P}} X$ alors $X_n \xrightarrow{\mathcal{L}} X$ lorsque $n \rightarrow +\infty$.*

Démonstration. Soit $(n_k)_{k \in \mathbb{N}}$ une sous-suite extraite. Puisque, par hypothèse, $X_n \xrightarrow{\mathbb{P}} X$ il existe alors une sous-suite $(n_{k_l})_{l \in \mathbb{N}}$ telle que

$$\lim_{l \rightarrow +\infty} X_{n_{k_l}} = X \quad p.s.$$

En particulier, si $f \in C_b(E)$ alors $\lim_{l \rightarrow +\infty} f(X_{n_{k_l}}) = f(X) \quad p.s.$ et en conséquence $\lim_{l \rightarrow +\infty} \mathbb{E}[f(X_{n_{k_l}})] = \mathbb{E}[f(X)]$. Autrement dit, $(\mu_{n_{k_l}})_{l \in \mathbb{N}}$ converge étroitement vers μ la loi de X .

Ceci entraîne que la suite initiale $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ . En effet, supposons que ce ne soit pas le cas. Il existe donc $f \in C_b(E)$ telle que

$$\lim_{n \rightarrow +\infty} \int_E f d\mu_n \neq \int_E f d\mu$$

En particulier, il existe $\epsilon > 0$ ainsi qu'une sous suite $(n_k)_{k \in \mathbb{N}}$ tels que pour tout $k \in \mathbb{N}$

$$\left| \int_E f d\mu_{n_k} - \int_E f d\mu \right| > \epsilon$$

Cependant, en reprenant l'argumentaire développé en début de preuve, nous pourrions trouver une sous suite $(n_{k_l})_{l \in \mathbb{N}}$ telle que

$$\left| \int_E f d\mu_{n_{k_l}} - \int_E f d\mu \right| \leq \epsilon$$

ce qui est absurde. □

Voyons à présent quelques critères permettant d'établir de la convergence étroite.

2.4 Quelques critères pour établir de la convergence étroite

Le théorème suivant fournit plusieurs manières équivalentes d'obtenir de la convergence étroite. Nous adopterons les abréviations suivantes $\liminf_{n \rightarrow +\infty}$ (resp. $\limsup_{n \rightarrow +\infty}$) pour désigner la limite inférieure (resp. la limite supérieure) lorsque n tend vers l'infini.

Théorème 2.4.1 (Portemanteau). *Soient (E, d) un espace polonais, $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$ et $\mu \in \mathcal{P}(E)$. Les assertions suivantes sont alors équivalentes :*

1. $\mu_n \Rightarrow \mu$ lorsque $n \rightarrow +\infty$.
2. Pour tout ouvert $O \subset E$, $\liminf_{n \rightarrow +\infty} \mu_n(O) \geq \mu(O)$.

3. Pour tout fermé $F \subset E$, $\limsup_{n \rightarrow +\infty} \mu_n(F) \leq \mu(F)$.

4. Pour tout $A \in \mathcal{B}(E)$ tel que $\mu(\partial A) = \mu(\bar{A} \setminus \mathring{A}) = 0$ alors

$$\limsup_{n \rightarrow +\infty} \mu_n(A) = \liminf_{n \rightarrow +\infty} \mu_n(A) = \mu(A)$$

Remarque. Les boréliens de E vérifiant la condition $\mu(\partial A) = 0$ sont appelés ensembles de continuité pour la mesure μ . Le point 4 est donc l'analogue au Théorème 1.5.1 de Helly-Bray en dimension infinie. D'ailleurs, lorsque $E = \mathbb{R}$, nous allons voir que le Théorème de Helly-Bray peut se démontrer à partir du Théorème du portemanteau.

Démonstration. Procédons de manière circulaire et commençons par montrer que (1) implique (2). Soit O un ouvert de E et notons par $F = O^c$ son complémentaire (qui est donc fermé). Pour tout $m \geq 1$, introduisons la fonction continue et bornée suivante :

$$f_m(x) = \min(1, md(x, F)) \quad \text{avec } x \in E.$$

Il n'est pas difficile de vérifier que cette fonction satisfait les propriétés suivantes, pour tout $m \geq 1$:

$$0 \leq f_m \leq 1_O \quad \text{et} \quad f_m \uparrow 1_O \quad \text{lorsque } m \rightarrow +\infty$$

Ainsi nous en déduisons que, pour tout $m \geq 1$,

$$\liminf_{n \rightarrow +\infty} \mu_n(O) = \liminf_{n \rightarrow +\infty} \int_E 1_O d\mu_n \geq \liminf_{n \rightarrow +\infty} \int_E f_m d\mu_n = \int_E f_m d\mu$$

par convergence étroite de $(\mu_n)_{n \in \mathbb{N}}$ vers μ . Il suffit ensuite de conclure par convergence monotone en faisant tendre m vers l'infini.

Par passage au complémentaire, il est évident que les points (2) et (3) sont équivalents.

Supposons que les points (2) et (3) soient satisfaits et démontrons que le point (4) est vérifié. A partir de ces hypothèses et à l'aide d'inclusions évidentes, nous en déduisons les inégalités suivantes

$$\begin{aligned} \mu(\mathring{A}) &\leq \liminf_{n \rightarrow +\infty} \mu_n(\mathring{A}) \leq \liminf_{n \rightarrow +\infty} \mu_n(A) \\ &\leq \limsup_{n \rightarrow +\infty} \mu_n(A) \leq \limsup_{n \rightarrow +\infty} \mu_n(\bar{A}) \leq \mu(\bar{A}) \end{aligned}$$

De plus, puisque $\mathring{A} \subset A \subset \bar{A}$, l'hypothèse $\mu(\partial A) = 0$ entraîne que $\mu(\mathring{A}) = \mu(\bar{A}) = \mu(A)$ ce qui permet de conclure.

Enfin, entamons la partie la plus délicate : démontrons que (4) implique (1). A cet effet, soit $f \in C_b(E)$. Sans perdre en généralité, nous pouvons supposer que $0 \leq f(x) \leq 1$ pour tout $x \in E$. Pour tout $y \in [0, 1]$ posons

$$F_y = \{x \in E; f(x) = y\}$$

et observons que ces ensembles F_y sont disjoints. Pour tout $k \geq 1$, introduisons l'ensemble I_k défini par

$$I_k = \left\{ y \in [0, 1]; \mu(F_y) \geq \frac{1}{k} \right\}$$

et observons que $\text{Card}(I_k) \leq k$. Si jamais ce n'était pas le cas, nous aboutirions à une contradiction. Ainsi,

$$\{y \in [0, 1]; \mu(F_y) > 0\} = \cup_{k \geq 1} I_k$$

Nous désignerons cette réunion dénombrable d'ensembles par I .

Soit $\epsilon = \frac{1}{N}$, avec $N \geq 1$ et découpons l'intervalle $[0, 1]$ en segment de longueur $1/N$. Dans chaque intervalle, $]\frac{k}{N}, \frac{k+1}{N}[$, avec $k = -1, 0, 1, \dots, N$ choisissons alors $z_k \notin I$. Ceci est bien envisageable car I est au plus dénombrable. Posons ensuite

$$B_k = \{x \in E; z_k \leq f(x) < z_{k+1}\}$$

et notons que, par construction, $\partial B_k \subset F_{z_k} \cup F_{z_{k+1}}$ pour tout $k \geq 1$ et que $\mu(F_{z_k}) = 0$ puisque $z_k \notin I$. En conséquence, $\mu(\partial B_k) = 0$ pour tout $k \geq 1$. Il est maintenant possible d'appliquer le point (4), nous avons donc

$$\begin{aligned} \int_E f d\mu &= \sum_{k=-1}^N \int_{B_k} f d\mu \leq \sum_{k=-1}^N z_{k+1} \mu(B_k) = \lim_{n \rightarrow +\infty} \sum_{k=-1}^n z_{k+1} \mu_n(B_k) \\ &\leq \liminf_{n \rightarrow +\infty} \int_E f d\mu_n + \frac{2}{N} \\ &\leq \limsup_{n \rightarrow +\infty} \int_E f d\mu_n + \frac{2}{N} \end{aligned}$$

car, d'une part, $z_{k+1} \leq z_k + \frac{2}{N}$ (par construction) et, d'autre part, $\sum_{k=-1}^N z_k \mu_n(B_k) \leq \int_E f d\mu_n$ puisque les ensembles $(B_k)_{k \geq 1}$ forment une partition de E .

Finalement, en utilisant à nouveau la définition de l'ensemble B_k ainsi que le point (4), nous en déduisons que

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \int_E f d\mu_n + \frac{2}{N} &= \lim_{n \rightarrow +\infty} \sum_{k=-1}^n z_{k+1} \mu_n(B_k) + \frac{2}{N} = \sum_{k=-1}^N z_{k+1} \mu(B_k) + \frac{2}{N} \\ &\leq \int_E f d\mu + \frac{4}{N}. \end{aligned}$$

En conclusion, pour $\epsilon = \frac{1}{N}$, nous avons montré que

$$\int_E f d\mu \leq \liminf_{n \rightarrow +\infty} \int_E f d\mu_n + 2\epsilon \leq \limsup_{n \rightarrow +\infty} \int_E f d\mu_n + 2\epsilon \leq \int_E f d\mu + 4\epsilon.$$

C'est pourquoi, lorsque $\epsilon \rightarrow 0$, $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$. □

Comme annoncé plus tôt, il est possible de démontrer le Théorème 1.5.1 d'Helly-Bray à partir du point (4) du Théorème précédent. Rappelons d'ailleurs que la fonction de répartition d'une mesure de probabilité $\mu \in \mathcal{P}(\mathbb{R})$ est définie par $F_\mu(t) = \mu(]-\infty, t])$ pour tout $t \in \mathbb{R}$.

Corollaire 2.4.1. *Si $E = \mathbb{R}$ les assertions suivantes sont équivalentes :*

1. $\mu_n \Rightarrow \mu$ lorsque $n \rightarrow +\infty$
2. $\lim_{n \rightarrow +\infty} F_{\mu_n}(t) = F_\mu(t)$ en tout point de continuité de F_μ .

Démonstration. Démontrons le sens direct. Si t est un point de continuité de F_μ alors $A =]-\infty, t]$ est un ensemble de continuité pour μ (i.e. $\mu(\partial A) = 0$). Si $\mu_n \Rightarrow \mu$ lorsque $n \rightarrow +\infty$ alors, d'après le Théorème du portemanteau, $F_{\mu_n}(t) \rightarrow F_\mu(t)$.

Traisons à présent le sens réciproque qui est plus délicat. Soit O un ouvert de $\mathbb{R}$. L'objectif est de montrer que la convergence des fonctions de répartition entraîne l'inégalité suivante :

$$\liminf_{n \rightarrow +\infty} \mu_n(O) \geq \mu(O)$$

afin de conclure grâce au Théorème du portemanteau. Notons qu'un ouvert O de $\mathbb{R}$ peut s'écrire comme réunion dénombrable d'intervalles ouverts $O = \cup_{j \geq 1}]a_j, b_j[$ avec $a_j, b_j \in O$ pour tout $j \geq 1$. Soit $\epsilon > 0$, par convergence monotone, il est possible de choisir $m \in \mathbb{N}$ suffisamment grand de sorte que

$$\mu(O_m) \geq \mu(O) - \epsilon \quad \text{où} \quad O_m = \cup_{j=1}^m]a_j, b_j[$$

De plus, comme a_j et b_j sont des éléments de O pour tout $j \geq 1$, il est également possible de trouver $a'_j < a_j$ et $b'_j > b_j$ tels que $]a'_j, b'_j[\subset O$ tout en imposant que a'_j et b'_j ne soient pas des points de discontinuités de F_μ et de F_{μ_n} et ceci pour tout $n \in \mathbb{N}$ (ceci est possible car, étant croissantes, les applications $t \mapsto F_\mu(t)$ et $t \mapsto F_{\mu_n}(t)$ n'admettent qu'un nombre au plus dénombrable de points de discontinuité).

Ainsi, si $O'_m = \cup_{j=1}^m]a'_j, b'_j[$, l'hypothèse de convergence des fonctions de répartition entraîne que

$$\lim_{n \rightarrow +\infty} \mu_n(O'_m) = \mu(O'_m).$$

En conclusion, puisque par construction $O_m \subset O'_m \subset O$, nous avons

$$\liminf_{n \rightarrow +\infty} \mu_n(O) \geq \liminf_{n \rightarrow +\infty} \mu_n(O'_m) = \mu(O'_m) \geq \mu(O_m) \geq \mu(O) - \epsilon$$

Enfin, ϵ étant arbitraire, la conclusion s'ensuit. $\square$

2.5 Distances pour la convergence étroite

Dans cette section nous allons voir qu'il est possible de munir $\mathcal{P}(E)$, l'ensemble des mesures de probabilités sur (E, d) , de différentes distances. Nous verrons par la suite que ces distances permettent d'étudier la convergence étroite d'un point de vue métrique. Avant cela, introduisons la notion d'élargi d'un ensemble.

Définition 2.5.1. Soit $A \subset E$, pour tout $\epsilon > 0$ l'ouvert A_ϵ désigne l'élargi de A . Celui-ci est défini par

$$A_\epsilon = \{x \in E; d(x, A) < \epsilon\}$$

La première des distances que nous allons étudier utilise cette notion d'élargissement.

2.5.1 Distance de Lévy-Prokhorov

Définition 2.5.2. Soient $\mu, \nu \in \mathcal{P}(E)$, la distance de Lévy-Prokhorov entre μ et ν est définie par

$$\rho(\mu, \nu) = \inf\{\epsilon > 0, \mu(B) \leq \nu(B_\epsilon) + \epsilon, \forall B \in \mathcal{B}(E)\}$$

Remarque. 1. Bien entendu, la terminologie est trompeuse et il faudra démontrer qu'il s'agit bien d'une distance.

2. Puisque, pour tout $\epsilon > 0$, $B_\epsilon = (\bar{B})_\epsilon$ il est possible de remplacer les boréliens de la définition par des fermés.
3. Parfois, certains auteurs ajoutent la condition supplémentaire : $\nu(B) \leq \mu(B_\epsilon) + \epsilon$ dans la définition de ρ .

A titre d'illustration.

Exemple 2.5.1. Soient $x, y \in E$, si $\mu = \delta_x$ et $\nu = \delta_y$ alors $\rho(\delta_x, \delta_y) = \min(1, d(x, y))$.

Proposition 2.5.1. ρ est une distance sur $\mathcal{P}(E)$.

Démonstration. Tout d'abord observons que $0 \leq \rho \leq 1$ et $\rho(\mu, \mu) = 0$. Montrons ensuite que ρ est symétrique, soit $\epsilon > 0$ et supposons que $\rho(\mu, \nu) > \epsilon$. Cela signifie qu'il est possible de trouver un borélien B tel que

$$\mu(B) > \nu(B_\epsilon) + \epsilon \iff \nu((B_\epsilon)^c) > \mu((B_\epsilon)^c) + \epsilon$$

Observons de plus que $((B_\epsilon)^c)_\epsilon \subset B^c$, d'où

$$\mu\left[\left((B_\epsilon)^c\right)_\epsilon\right] + \epsilon \leq \mu(B^c) + \epsilon < \nu((B_\epsilon)^c)$$

En conséquence, $\rho(\nu, \mu) \geq \epsilon$ donc $\rho(\nu, \mu) \geq \rho(\mu, \nu)$. Il suffit ensuite d'échanger les rôles de μ et ν pour conclure que $\rho(\mu, \nu) = \rho(\nu, \mu)$.

Montrons que $\rho(\mu, \nu) = 0$ implique alors $\mu = \nu$. Si $\rho(\nu, \mu) = 0$ alors, pour tout fermé $F \subset E$ et pour $\epsilon = \frac{1}{k}$ avec $k \geq 1$, nous avons

$$\mu(F) \leq \nu(F_{\frac{1}{k}}) + \frac{1}{k}$$

Or, puisque F est fermé, nous avons $F = \bigcap_{k \geq 1} F_{\frac{1}{k}}$. Ainsi, en passant à la limite nous obtenons que $\mu(F) \leq \nu(F)$ et, par symétrie, $\nu(F) \leq \mu(F)$. En d'autres termes, $\mu = \nu$.

Il ne reste plus qu'à établir l'inégalité triangulaire. A cet effet, soient $\mu, \nu, \gamma \in \mathcal{P}(E)$ et considérons $\epsilon > 0$ ainsi que $\delta > 0$ tels que

$$\rho(\gamma, \mu) < \epsilon \quad \text{et} \quad \rho(\mu, \nu) < \delta$$

Si $B \in \mathcal{B}(E)$ alors

$$\gamma(B) \leq \mu(B_\epsilon) + \epsilon \leq \mu((B_\epsilon)_\delta) + \delta + \epsilon \leq \mu((B)_{\epsilon+\delta}) + \epsilon + \delta$$

donc $\rho(\gamma, \mu) \leq \epsilon + \delta$. En conséquence, $\rho(\gamma, \mu) \leq \rho(\gamma, \nu) + \rho(\mu, \nu)$. $\square$

2.5.2 Distance de Kantorovich-Rubinstein

Dans un espace métrique, il n'est pas vraiment possible de parler de la régularité de fonctions en terme de différentiabilité. Toutefois, il est possible de trouver un substitut naturel : il s'agit de la notion de fonctions lipschitziennes. Nous allons voir que de telles fonctions jouent un rôle important pour mesurer, en terme de distance, la convergence étroite. Avant cela, rappelons quelques définitions.

Définition 2.5.3. Soit $f : E \rightarrow \mathbb{R}$ une fonction définie sur l'espace métrique (E, d) . Cette fonction est dite lipschizienne si le rapport suivant est fini :

$$\|f\|_{Lip} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}$$

L'ensemble des fonctions bornées et lipschitziennes sur E sera noté $BL(E)$ (pour « bounded-Lipschitz »).

Remarque. Par la suite, nous adopterons la notation suivante $\|f\|_{BL} = \|f\|_\infty + \|f\|_{Lip}$.

Bien que nous ne le démontrerons pas, il est possible de définir une distance sur $\mathcal{P}(E)$ à partir des fonctions de l'espace $BL(E)$. Nous retrouverons cette distance dans le cadre du transport optimal dans un chapitre ultérieur.

Proposition 2.5.2 (Kantorovich-Rubinstein). Soient $\mu, \nu \in \mathcal{P}(E)$, la quantité suivante est une distance sur $\mathcal{P}(E)$.

$$\beta(\mu, \nu) = \sup_{\|f\|_{BL} \leq 1} \left\{ \left| \int_E f d\mu - \int_E f d\nu \right| \right\}$$

Il s'agit de la distance de Kantorovich-Rubinstein.

2.5.3 Topologie de $\mathcal{P}(E)$

A partir des distances introduites précédemment, nous allons voir qu'elles permettent de quantifier la convergence étroite de mesures de probabilités.

Théorème 2.5.1. Soient (E, d) un espace polonais ainsi que $\mu_n, \mu \in \mathcal{P}(E)$. Les assertions suivantes sont alors équivalentes.

$$1. \mu_n \xrightarrow{n \rightarrow +\infty} \mu.$$

2. Pour toutes fonctions $f \in BL(E)$, nous avons $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$.

3. $\lim_{n \rightarrow +\infty} \beta(\mu_n, \mu) = 0$.

4. $\lim_{n \rightarrow +\infty} \rho(\mu_n, \mu) = 0$.

Remarque. Certaines implications sont évidentes. Notamment $(1) \Rightarrow (2)$ et $(3) \Rightarrow (2)$. En revanche, l'application $(2) \Rightarrow (3)$ semble plus délicate. D'ailleurs, pour démontrer ceci, nous aurons besoin du Théorème d'Arzélà-Ascoli qui caractérise les parties relativement compactes dans l'espace des fonctions continues $C(K)$ sur un ensemble compact K . Rappelons son énoncé.

Théorème 2.5.2 (Arzélà-Ascoli). *Soit (K, d) un espace métrique compact et considérons $C(K)$ muni de la distance uniforme $\|\cdot\|_\infty$. Les assertions suivantes sont alors équivalentes :*

1. Une partie $A \subset C(K)$ est relativement compacte ;

2. A est bornée et équicontinue : pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que

$$i.e. \quad \sup_{f \in A} \sup_{\substack{s, t \in K \\ d(s, t) < \eta}} |f(s) - f(t)| \leq \epsilon$$

Démonstration. Démontrons que $(2) \Rightarrow (3)$, la preuve s'effectue en deux étapes.

1ère étape : soit $\mu \in \mathcal{P}(E)$, montrons que la tension de cette mesure à un impact sur μ_n . Soit $0 < \epsilon < \frac{1}{2}$, puisque μ est tendue il existe un compact K de E tel que $\mu(K) \geq 1 - \epsilon$. Posons alors

$$g(x) = \max \left(0, 1 - \frac{1}{\epsilon} d(x, K) \right) \quad \text{avec } x \in E.$$

Il est alors possible de montrer que $g \in BL(E)$ et que l'inégalité suivante est vérifiée :

$$1_K \leq g \leq 1_{K_\epsilon}$$

Alors, d'après l'inégalité précédente et l'hypothèse (2), nous avons, pour $n \in \mathbb{N}$ suffisamment grand,

$$\mu_n(K_\epsilon) \geq \int_E g d\mu_n \geq \int_E g d\mu - \epsilon \geq \mu(K) - \epsilon \geq 1 - 2\epsilon$$

où la dernière minoration s'obtient grâce à la propriété de tension de μ . En conclusion, nous venons de montrer que pour tout $\epsilon \in]0, \frac{1}{2}[$ et n suffisamment grand, nous avons

$$\mu_n(K_\epsilon) \geq 1 - 2\epsilon$$

Autrement dit, quitte à élargir l'ensemble K , la tension de μ se « transmet » à μ_n (attention K_ϵ n'a aucune raison d'être compact d'où la présence des précédents guillemets).

2ième étape : Posons $\mathcal{F} = \{f \in BL(E); \|f\|_{BL(E)} \leq 1\}$ et observons que la famille des restrictions à l'ensemble K des fonctions $f \in \mathcal{F}$ forme, d'après le Théorème d'Arzélà-Ascoli, une

partie relativement compact de $(C(K), \|\cdot\|_\infty)$.

En particulier, cette partie de $C(K)$ est précompacte : il est donc possible de trouver $f_1, \dots, f_k \in \mathcal{F}|_K$ telles que pour toute fonction $f \in \mathcal{F}|_K$ il existe $j \in \{1, \dots, k\}$ tel que

$$\|f - f_j\|_\infty \leq \epsilon \iff \sup_{y \in K} |f(y) - f_j(y)| \leq \epsilon$$

De ceci, il découle aussi que $\sup_{y \in K_\epsilon} |f(y) - f_j(y)| \leq 3\epsilon$. En effet, si $y \in K_\epsilon$, il existe $z \in K$ tel que $d(y, z) < \epsilon$. Ainsi,

$$\begin{aligned} |(f(y) - f_j(y))| &\leq |f(y) - f(z)| + |f(z) - f_j(z)| + |f_j(z) - f_j(y)| \\ &\leq d(y, z) + \epsilon + d(z, y) \leq 3\epsilon \end{aligned}$$

puisque f et f_j sont 1-lipschitziennes et $\|f - f_j\|_\infty \leq \epsilon$.

Conclusion : Soit $f \in \mathcal{F}$, pour n assez grand nous avons

$$\begin{aligned} \left| \int_E f d\mu_n - \int_E f d\mu \right| &\leq \int_E |f - f_j| d\mu_n + \int_E |f - f_j| d\mu + \left| \int_E f_j d\mu_n - \int_E f_j d\mu \right| \\ &\leq \int_{K_\epsilon} |f - f_j| d\mu_n + \int_{(K_\epsilon)^c} |f - f_j| d\mu_n \\ &\quad + \int_K |f - f_j| d\mu_n + \int_{K^c} |f - f_j| d\mu_n \\ &\quad + \left| \int_E f_j d\mu_n - \int_E f_j d\mu \right| \end{aligned}$$

En utilisant ce que nous avons démontré précédemment, nous en déduisons la majoration suivante

$$\begin{aligned} \left| \int_E f d\mu_n - \int_E f d\mu \right| &\leq 3\epsilon \mu_n(K_\epsilon) + 2\mu_n((K_\epsilon)^c) + \epsilon \mu(K) + 2\mu(K^c) + \max_{j=1, \dots, n} \left| \int_E f_j d\mu_n - \int_E f_j d\mu \right| \\ &\leq 3\epsilon + 4\epsilon + \epsilon + 2\epsilon + \max_{j=1, \dots, n} \left| \int_E f_j d\mu_n - \int_E f_j d\mu \right| \end{aligned}$$

Ainsi,

$$\limsup_{n \rightarrow +\infty} \sup_{f \in \mathcal{F}} \left| \int_E f d\mu_n - \int_E f d\mu \right| \leq 10\epsilon + \limsup_{n \rightarrow +\infty} \max_{j=1, \dots, n} \left| \int_E f_j d\mu_n - \int_E f_j d\mu \right| = 10\epsilon$$

puisque, par hypothèse, $\lim_{n \rightarrow +\infty} \left| \int_E f_j d\mu_n - \int_E f_j d\mu \right| = 0$ pour tout $j = 1, \dots, k$. Enfin, il suffit de faire tendre ϵ vers 0 pour conclure.

Montrons à présent que (3) $\Rightarrow$ (4). Pour cela, nous allons montrer qu'il est possible de comparer les distances ρ et β . Soient $B \in \mathcal{B}(E)$ et $\mu, \nu \in \mathcal{P}(E)$, sans perdre en généralité, il est possible de

supposer que $\mu(B) \leq \nu(B)$. Soit $0 < \epsilon < \frac{1}{2}$ et considérons la fonction g définie par

$$g(x) = \max \left(0, 1 - \frac{1}{\epsilon} d(x, B) \right), \quad x \in E$$

Rappelons que $g \in BL(E)$ avec $\|g\|_{BL} \leq 1 + \frac{1}{\epsilon}$ et $1_B \leq g \leq 1_{B_\epsilon}$. C'est pourquoi nous avons les inégalités suivantes,

$$\begin{aligned} \mu(B) \leq \nu(B) &\leq \int_E g d\nu \leq \int_E g d\nu + \left(1 + \frac{1}{\epsilon}\right) \beta(\mu, \nu) \\ &\leq \nu(B_\epsilon) + \left(1 + \frac{1}{\epsilon}\right) \beta(\mu, \nu) \\ &\leq \nu(B_\delta) + \delta \end{aligned}$$

où $\delta = \max \left(\epsilon, \left(1 + \frac{1}{\epsilon}\right) \beta(\mu, \nu) \right)$. D'où, par définition de ρ , nous avons

$$\rho(\mu, \nu) \leq \delta$$

Il suffit ensuite de choisir $\epsilon = 2\sqrt{\beta(\mu, \nu)}$ pour en déduire que $\rho(\mu, \nu) \leq 2\sqrt{\beta(\mu, \nu)}$ afin de conclure que (3) $\Rightarrow$ (4).

Enfin, démontrons que (4) $\Rightarrow$ (1). Pour cela, nous allons utiliser le théorème du portemanteau. A cet effet, soient A un ensemble de continuité de μ et $\epsilon > 0$. Considérons également $\delta \in]0, \epsilon[$ suffisamment petit de sorte que

$$\mu(A_\delta \setminus A) < \epsilon \quad \text{et} \quad \mu((A^c)_\delta \setminus A^c) < \epsilon \quad (2.5.1)$$

Ceci est possible car A est un ensemble de continuité pour μ (i.e. $\mu(\partial A) = \mu(\bar{A} \setminus \mathring{A})$), il suffit alors de faire décroître les ensembles A_δ vers A .

Pour n assez grand, puisque $\lim_{n \rightarrow +\infty} \rho(\mu_n, \mu) = 0$ et d'après les relations (2.5.1), nous avons

$$\mu_n(A) \leq \mu(A_\delta) + \delta \leq \mu(A) + \epsilon + \delta \leq \mu(A) + 2\epsilon$$

et, par passage au complémentaire,

$$\mu_n(A^c) \leq \mu((A_\delta)^c) + \delta \leq \mu(A^c) + \delta + \epsilon.$$

Autrement dit $|\mu_n(A) - \mu(A)| \leq 2\epsilon$, ce qui permet de conclure. $\square$

Le théorème qui va suivre est essentiel car il permet de caractériser les parties relativement compactes de $\mathcal{P}(E)$. De manière assez remarquable, ces ensembles sont obtenus à partir des compacts de l'espace (E, d) et de la notion de tension. Dans ce but, nous proposons ci après la définition d'une partie tendue de $\mathcal{P}(E)$.

Définition 2.5.4. Une partie A de $\mathcal{P}(E)$ est dite uniformément tendue si pour tout $\epsilon > 0$ il existe K un compact de E tel que

$$\forall \mu \in \mathcal{P}(E) \quad \mu(K) \geq 1 - \epsilon$$

Remarque. Par simplicité, la plupart du temps nous parlerons de partie tendue pour désigner une partie uniformément tendue. Les observations valables en dimension finies le sont encore :

1. une partie finie de $\mathcal{P}(E)$ est toujours tendue (puisque une mesure seule l'est) ;
2. si $\mu_n \xrightarrow{n \rightarrow +\infty} \mu$ alors $(\mu_n)_{n \in \mathbb{N}}$ est tendue.

Voici à présent la caractérisation des parties relativement compactes de $\mathcal{P}(E)$.

Théorème 2.5.3 (Prokhorov). *Soit (E, d) un espace métrique polonais et $A \subset \mathcal{P}(E)$. Les assertions suivantes sont équivalentes :*

1. *A est une partie tendue.*
2. *De toute suite de A , il existe une sous-suite qui converge étroitement.*
3. *l'ensemble A est relativement compact pour la distance β (ou ρ).*
4. *l'ensemble A est pré-compact pour la distance β (ou ρ).*

Enonçons un corollaire immédiat avant de démontrer ce résultat.

Corollaire 2.5.1. *L'espace métrique $(\mathcal{P}(E), \beta)$ est un espace complet.*

Remarque. Bien entendu, ceci reste vrai si l'on remplace la distance β par ρ .

Démonstration. En effet, si $(\mu_n)_{n \in \mathbb{N}}$ est une suite de Cauchy de (E, β) elle définit un ensemble pré-compact. D'après le Théorème de Prokhorov 2.5.3, il est alors possible d'en extraire une sous-suite convergente. En conséquence, la suite initiale converge et l'espace est donc complet. $\square$

Passons à la démonstration du théorème de Prokhorov.

Démonstration. (du Théorème de Prokhorov 2.5.3) Notons que (3) $\iff$ (4) car $\mathcal{P}(E)$ est un espace métrique. (3) $\Rightarrow$ (4) est trivial et (4) $\Rightarrow$ (1) est laissé à titre d'exercice (*indication : il faut utiliser le fait qu'une partie finie est tendue*).

Concentrons nous sur l'implication principale (1) $\Rightarrow$ (2). Soit $(\mu_n)_{n \in \mathbb{N}} \subset A$ une partie tendue. Nous devons donc extraire une sous-suite convergente de la suite $(\mu_n)_{n \in \mathbb{N}}$.

Utilisons la propriété de tension avec $\epsilon = \frac{1}{j}$, pour $j \geq 1$. Ceci permet de construire une suite $(K_j)_{j \geq 1}$ de compacts emboîtés (i.e. $K_j \subset K_{j+1}$, $j \geq 1$) telle que

$$\forall n \in \mathbb{N}, \quad \mu_n(K_j) \geq 1 - \frac{1}{j}$$

Ensuite, pour tout $j \geq 1$ et $n \in \mathbb{N}$, désignons par $\nu_n^j = \mu_n|_{K_j}$ la restriction de μ_n à l'ensemble K_j . Notons ensuite que l'espace $C(K_j, \|\cdot\|_\infty)$ un espace de Banach séparable (puisque K_j est un ensemble compact) et $C(K_j)' = \mathcal{M}(K_j)$ l'espace des mesures signées bornées sur K_j (d'après le Théorème 1.2.2 de dualité de Riesz).

En outre, puisque nous avons à disposition une suite bornée dans $\mathcal{M}(K_j)$, le Théorème 1.1.1 de Banach-Alaoglu nous assure qu'il est possible d'extraire une sous-suite convergente de $(\nu_n^j)_{n \in \mathbb{N}}$ (rappelons que j est fixé). Plus précisément, il existe ν^j telle que la sous-suite $(\nu_n^j)_{n \in \mathbb{N}}$ (pour alléger les notations, la sous-suite extraite sera notée comme la suite initiale) converge faiblement (et donc étroitement puisque K_j est un ensemble compact) vers ν^j .

Par l'argument diagonal de Cantor, il existe alors une sous-suite (toujours notée avec un indice n pour alléger les notations) telle que

$$\nu_n^j \xrightarrow{n \rightarrow +\infty} \nu^j \quad \text{pour tout } j \geq 1.$$

Ceci permet de définir $\mu(B) = \lim_{j \rightarrow +\infty} \uparrow \nu^j(B \cap K_j)$ pour tout $B \in \mathcal{B}(E)$. Notons que μ est une mesure de probabilité sur E car elle est obtenue comme limite croissante de mesures de probabilités.

Pour conclure, nous devons donc montrer qu'il est possible d'extraire une sous-suite de $(\mu_n)_{n \in \mathbb{N}}$ telle que cette sous-suite converge étroitement vers la mesure μ . Soit $f \in C_b(E)$ et montrons que

$$\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$$

Sans perdre en généralité, il est possible de supposer que $0 \leq f \leq 1$. Pour tout $n \in \mathbb{N}$ et pour tout $j \geq 1$, nous avons les inégalités suivantes

$$\begin{aligned} \left| \int_E f d\mu_n - \int_E f d\mu \right| &\leq \left| \int_E f d\mu_n - \int_{K_j} f d\mu^j \right| + \left| \int_{K_j} f d\nu^j - \int_E f d\mu \right| \\ &\leq \left| \int_{K_j} f d\mu_n - \int_{K_j} f d\nu^j \right| + \mu_n(K_j^c) + \left| \int_{K_j} f d\nu^j - \int_E f d\mu \right| \\ &\leq \left| \int_{K_j} f d\nu_n^j - \int_{K_j} f d\nu^j \right| + \frac{1}{j} + \left| \int_{K_j} f d\nu^j - \int_E f d\mu \right| \end{aligned}$$

Puisque $\mu_n|_{K_j} = \nu_n^j$ et $\mu_n(K_j^c) \leq \frac{1}{j}$ par tension. Ensuite, puisque $(\nu_n^j)_{n \in \mathbb{N}}$ converge étroitement vers ν^j , nous en déduisons que

$$\limsup_{n \rightarrow +\infty} \left| \int_E f d\mu_n - \int_E f d\mu \right| \leq \frac{1}{j} + \left| \int_{K_j} f d\nu^j - \int_E f d\mu \right|$$

Il suffit ensuite de faire tendre j vers l'infini pour conclure. En effet, par définition de μ (pour le choix du borélien $B = E$) nous avons

$$\int_{K_j} f d\nu^j \xrightarrow{j \rightarrow +\infty} \int_E f d\mu$$

Donc $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$ ce qui démontre (2). □

2.6 Applications

Nous avons déjà rappelé plus tôt quelques résultats de convergence de mesures lorsque $E = \mathbb{R}^d$, voyons ce qui se produit lorsque $E = C([0, 1])$ muni de $\|\cdot\|_\infty$. Notons que dans un tel cadre, nous aurions (à ω fixé) affaire à des courbes aléatoires (continues) de la forme $t \mapsto X_t(\omega)$.

Supposons que nous cherchions à démontrer un résultat de convergence en loi ($X_n \xrightarrow{\mathcal{L}} X$ lorsque $n \rightarrow +\infty$) dans un tel espace polonais. Il est tentant d'essayer d'utiliser à nouveau la transformée de Fourier. Celle-ci est définie par dualité

$$\phi_{X_n}(\xi) = \mathbb{E}[e^{i\langle \xi, X_n \rangle}] \quad \text{avec} \quad \xi \in E'$$

Malheureusement, comme mentionné plus tôt, l'outil de la fonction caractéristique ne suffit plus à décrire la convergence en loi (en dimension infinie, l'inégalité de Lévy fournissant de la tension à partir de la convergence des fonctions caractéristiques n'est plus disponible). Voici alors le schéma de démonstration qu'il faut suivre :

1. Etablir que $(\mu_n)_{n \in \mathbb{N}}$ (la loi de X_n) est tendue.
2. Utiliser le théorème de Prokhorov pour extraire une sous-suite qui converge étroitement.
3. Identifier la limite en montrant qu'il n'y a qu'une seule valeur d'adhérence. A cet étape, l'injectivité de la transformée de Fourier peut s'avérer utile.

Remarque. Dans la pratique, l'étape délicate revient à établir cette propriété de tension.

Comme auparavant, nous passerons régulièrement des notations probabilistes avec les variables aléatoires $(X_n)_{n \in \mathbb{N}}$ aux notations du calcul intégral impliquant les lois de probabilités $(\mu_n)_{n \in \mathbb{N}}$ associées.

2.6.1 Critère de convergence étroite dans $C[0, 1]$

Voyons ce qu'il est possible de dire lorsque $E = C[0, 1]$. Comme nous avons besoin de produire des ensembles compacts dans cet espace, il est certain que le théorème d'Arzélà-Ascoli va jouer un rôle important. Voici une proposition utile allant dans ce sens.

Proposition 2.6.1. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variable aléatoire à valeurs dans $E = C[0, 1]$. La suite $(X_n)_{n \in \mathbb{N}}$ est alors tendue si et seulement si*

1. $\lim_{M \rightarrow +\infty} \sup_{n \in \mathbb{N}} \mathbb{P}(|X_n(t_0)| \geq M) = 0$ pour un certain $t_0 \in [0, 1]$
2. $\forall \epsilon > 0, \quad \lim_{\delta \rightarrow 0} \sup_{n \in \mathbb{N}} \mathbb{P}\left(\sup_{|s-t| \leq \delta} |X_n(s) - X_n(t)| \geq \epsilon\right) = 0$

Remarque. Pour démontrer ce résultat nous allons devoir utiliser le Théorème d'Arzélà-Ascoli 2.5.2. A ce propos, étant donnée une suite de fonctions continues $(x_n)_{n \in \mathbb{N}}$, observons que l'hypothèse : il existe $M > 0$ telle que

$$\sup_{n \in \mathbb{N}} \|x_n\|_\infty \leq M$$

peut être affaiblie en : il existe $M > 0$ et $t_0 \in [0, 1]$ tels que

$$\sup_{n \in \mathbb{N}} |x_n(t_0)| \leq M$$

En effet, cette nouvelle hypothèse combinée à la propriété d'équicontinuité entraîne que $\sup_{n \in \mathbb{N}} \|x_n\|_\infty \leq M$.

Démonstration. Supposons que les deux hypothèses soient satisfaites et montrons que la suite $(X_n)_{n \in \mathbb{N}}$ est tendue. Soit $\eta > 0$, en utilisant la première hypothèse choisissons M suffisamment grand de sorte que

$$\sup_{n \in \mathbb{N}} \mu_n(x \in E, |x(t_0)| > M) \leq \eta$$

Ensuite, grâce à la deuxième hypothèse, pour $\epsilon = \frac{1}{k}$, avec $k \geq 1$, choisissons $\delta_k > 0$ tel que

$$\sup_{n \in \mathbb{N}} \mu_n \left(x \in E; \sup_{|s-t| \leq \delta_k} |x(s) - (t)| > \frac{1}{k} \right) \leq \frac{\eta}{2^k}$$

Définissons ensuite l'ensemble suivant :

$$K = \left\{ x \in E; |x(t_0)| \leq M; \sup_{|s-t| \leq \delta_k} |x(s) - (t)| \leq \frac{1}{k} \right\}$$

et observons qu'il s'agit d'un ensemble fermé dans $(E, \|\cdot\|_\infty)$. De plus, d'après le Théorème 2.5.2 d'Arzela-Ascoli, l'ensemble K est donc compact dans $(E, \|\cdot\|_\infty)$. Vérifions à présent que K vérifie la propriété de tension voulue. Pour tout $n \in \mathbb{N}$, nous avons, par construction,

$$\begin{aligned} \mu_n(K^c) &\leq \mu_n(x \in E; |x(t_0)| > M) + \sum_{k \geq 1} \mu_n \left(x \in E; \sup_{|s-t| \leq \delta_k} |x(s) - (t)| > \frac{1}{k} \right) \\ &\leq \eta + \sum_{k \geq 1} \frac{\eta}{2^k} = 2\eta \end{aligned}$$

En conclusion, nous avons construit un compact de $(E, \|\cdot\|_\infty)$ tel que

$$\forall n \in \mathbb{N} \quad \mu_n(K) \geq 1 - 2\eta$$

Autrement dit, $(\mu_n)_{n \in \mathbb{N}}$ est tendue. L'implication réciproque est laissée en exercice. $\square$

Démonstration. En pratique, nous avons besoin de la tension qu'à partir d'un certain rang N . Il est donc possible de remplacer les limites de la proposition par des $\limsup_{n \rightarrow +\infty}$. $\square$

2.6.2 Principe d'invariance de Donsker

Dans cette section, nous allons démontrer une version analogue du théorème de la limite centrale en dimension infinie (lorsque $E = C([0, 1])$). L'objet limite sera un processus gaussien que nous allons décrire plus en détails ultérieurement.

Considérons $(V_j)_{j \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées (*i.i.d.* en abrégé) sur un espace probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Nous supposons par la suite que $\mathbb{E}[V_1] = 0$ et $\mathbb{E}[V_1^2] = 1$. Définissons également la quantité suivante :

$$S_n = \sum_{j=1}^n V_j \quad \text{avec} \quad \forall n \geq 1 \quad \text{et} \quad S_0 = 0$$

A partir de ceci, nous pouvons construire l'interpolation linéaire, entre les points $S_0, S_1, \dots, S_n, \dots$, notée Y_u , pour $u \geq 0$ par la formule suivante

$$Y_u = S_{[u]} + (u - [u])V_{[u]+1}, \quad u \geq 0$$

où $[\cdot]$ correspond à la partie entière d'un nombre réel. Le principal objet d'intérêt de cette section est le suivant :

$$X_t^n = \frac{1}{\sqrt{n}} Y_{nt}, \quad t \in [0, 1] \quad \text{et} \quad n \geq 1$$

Observons que, pour tout $n \geq 1$, $X^n = (X_t^n)_{t \in [0,1]}$ est une variable aléatoire à valeurs dans $E = C([0, 1])$. Nous allons montrer que X^n converge en loi vers une variable aléatoire X à déterminer.

Identification de la limite

Pour identifier la limite éventuelle, nous allons utiliser les lois (en dimension finie) induites par X^n .

Proposition 2.6.2. *Soient $0 \leq t_1 < t_2 < \dots < t_d \leq 1$ alors*

$$(X_{t_1}^n, X_{t_2}^n, \dots, X_{t_d}^n) \xrightarrow{\mathcal{L}} G \quad \text{dans} \quad \mathbb{R}^d \quad \text{lorsque} \quad n \rightarrow \infty$$

où G est un vecteur gaussien centré de matrice covariance $\Gamma = (\Gamma_{ij})_{1 \leq i, j \leq d}$ avec $\Gamma_{ij} = \min(t_i, t_j)$ pour tout $1 \leq i, j \leq d$.

Démonstration. Pour simplifier, nous ferons la démonstration lorsque $d = 2$. Le cas général est similaire mais les notations alourdissent inutilement la preuve.

Tout d'abord, observons que pour tout $t \in [0, 1]$

$$\left| X_t^n - \frac{1}{\sqrt{n}} S_{[nt]} \right| \leq \frac{1}{\sqrt{n}} |V_{[nt]+1}|$$

De plus, pour tout $\epsilon > 0$, puisque les variables ont même loi,

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} |V_{[nt]+1}| > \epsilon\right) = \mathbb{P}(|V_1| > \sqrt{n}\epsilon).$$

D'où, d'après l'inégalité de Tchebyshev, $\frac{1}{\sqrt{n}} |V_{[nt]+1}| \xrightarrow{\mathbb{P}} 0$ lorsque $n \rightarrow +\infty$. C'est donc également le cas de $|X_t^n - \frac{1}{\sqrt{n}} S_{[nt]}|$.

En conséquence, la norme euclidienne suivante $\left\| (X_{t_1}^n, X_{t_2}^n) - \left(\frac{1}{\sqrt{n}} S_{\lfloor nt_1 \rfloor}, \frac{1}{\sqrt{n}} S_{\lfloor nt_2 \rfloor} \right) \right\|_2$ tend vers 0 en probabilité.

Pour poursuivre notre étude, nous utiliserons le lemme ci-dessous.

Lemme 2.6.1. *Soient Y_n, X_n et X des variables aléatoires. Supposons que $Y_n \xrightarrow{\mathcal{L}} X$ et $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$ lorsque $n \rightarrow +\infty$. Alors $X_n \xrightarrow{\mathcal{L}} X$ lorsque $n \rightarrow +\infty$.*

Au vu de ce qui précède, le précédent lemme nous assure que nous pouvons travailler avec $\frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}$ plutôt que X_t^n .

Puisque nous sommes en dimension finie, nous pouvons utiliser la transformée de Fourier pour établir un résultat de convergence en loi. Ici, pour $(u_1, u_2) \in \mathbb{R}^2$, nous avons

$$\begin{aligned} \mathbb{E} \left[e^{iu_1 \frac{1}{\sqrt{n}} S_{\lfloor nt_1 \rfloor} + iu_2 \frac{1}{\sqrt{n}} S_{\lfloor nt_2 \rfloor}} \right] &= \mathbb{E} \left[e^{i(u_1 + u_2) \frac{1}{\sqrt{n}} S_{\lfloor nt_1 \rfloor} + iu_2 \frac{1}{\sqrt{n}} (S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor})} \right] \\ &= \mathbb{E} \left[e^{i(u_1 + u_2) \frac{\sqrt{t_1}}{\sqrt{nt_1}} S_{\lfloor nt_1 \rfloor}} \right] \mathbb{E} \left[e^{iu_2 \frac{1}{\sqrt{n}} (S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor})} \right] \end{aligned}$$

Puisque $S_{\lfloor nt_1 \rfloor}$ et $(S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor})$ sont des variables aléatoires indépendantes (car $t_1 < t_2$). Nous pouvons alors utiliser le Théorème 1.4.3 de la limite centrale dans $\mathbb{R}^2$ qui nous assure que

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[e^{i(u_1 + u_2) \frac{\sqrt{t_1}}{\sqrt{nt_1}} S_{\lfloor nt_1 \rfloor}} \right] \mathbb{E} \left[e^{iu_2 \frac{1}{\sqrt{n}} (S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor})} \right] = e^{-(u_1 + u_2) \frac{t_1}{2}} \times e^{-u_2^2 \frac{t_2 - t_1}{2}}$$

En d'autres termes,

$$(X_{t_1}^n, X_{t_2}^n) \xrightarrow{\mathcal{L}} G \quad \text{lorsque } n \rightarrow +\infty$$

où G est un vecteur gaussien centré de matrice covariance $\Gamma = \begin{pmatrix} t_1 & t_1 \\ t_1 & t_2 \end{pmatrix}$. □

2ième étape : la suite $(\mu_n)_{n \geq 1}$ est tendue

Nous devons établir une propriété de tension pour la suite $\mu_n = \mathcal{L}(X^n)$ dans $E = C([0, 1])$. Pour cela nous allons nous reposer sur la Proposition 2.6.1.

Nous devons donc montrer que les assertions suivantes sont satisfaites :

1. il existe $t_0 \in [0, 1]$ tel que $\lim_{M \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(|X_{t_0}^n| > M) = 0$.
2. pour tout $\epsilon > 0$, $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{|s-t| \leq \delta} |X_s^n - X_t^n| \geq \epsilon \right) = 0$.

Comme nous allons le voir, la deuxième assertion est plus délicate tandis que la première est clairement vérifiée en $t_0 = 0$ puisque $X_0^n = 0$.

Pour démontrer que la deuxième assertion est vérifiée, nous allons devoir étudier des quantités de la forme

$$\sup_{(s,t) \in [0,1]^2; |s-t| \leq \delta} |X_s^n - X_t^n| = \frac{1}{\sqrt{n}} \sup_{(u,v) \in [0,n]^2; |u-v| \leq n\delta} |Y_u - Y_v| \quad (2.6.1)$$

où, rappelons-le, $Y_u = S_{[u]} + (u - [u])V_{[u]+1}$ et $X_t^n = \frac{1}{\sqrt{n}}Y_{nt}$, $t \in [0, 1]$. Observons à présent que les majorations suivantes sont satisfaites :

$$\begin{aligned} \frac{1}{\sqrt{n}} \sup_{(u,v) \in [0,n]^2; |u-v| \leq n\delta} |Y_u - Y_v| &\leq \frac{1}{\sqrt{n}} \sup_{(u,v) \in [0,n]^2; |u-v| \leq n\delta} |S_{[u]} - Y_{[v]}| + \frac{2}{\sqrt{n}} \max_{j=1, \dots, n+1} |V_j| \\ &\leq \frac{1}{\sqrt{n}} \max_{l=0, \dots, n} |S_{k+l} - S_l| + \frac{2}{\sqrt{n}} \max_{j=1, \dots, n+1} |V_j| \\ &\quad \quad \quad k=1, \dots, [n\delta]+1 \end{aligned}$$

Le fait que l'on puisse remplacer le premier supremum par un maximum découle du fait que nous avons affaire à une fonction linéaire par morceaux. En outre, nous avons également, pour tout $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\frac{2}{\sqrt{n}} \max_{j=1, \dots, n} |V_j| \leq \epsilon\right) &\leq \sum_{j=1}^{n+1} \mathbb{P}\left(|V_j| \geq \frac{\epsilon\sqrt{n}}{2}\right) \\ &= (n+1) \mathbb{P}\left(|V_1| \geq \frac{\epsilon\sqrt{n}}{2}\right) \\ &\leq \frac{4(n+1)}{\epsilon^2 n} \int_{\{|V_1| \geq \frac{\epsilon\sqrt{n}}{2}\}} |V_1|^2 d\mathbb{P} \xrightarrow{n \rightarrow +\infty} 0 \end{aligned}$$

où, nous avons utilisé l'inégalité de Markov pour obtenir la dernière majoration. En effet, rappelons que si Z est une variable aléatoire positive, alors, pour tout $t > 0$,

$$\mathbb{P}(Z \geq t) \leq \frac{1}{t} \int_{\{Z \geq t\}} Z d\mathbb{P}.$$

Il ne reste plus qu'à étudier $\frac{1}{\sqrt{n}} \max_{l=0, \dots, n} |S_{k+l} - S_l|$. Dans un premier temps, nous allons nous concentrer sur la quantité suivante :

$$\frac{1}{\sqrt{n}} \max_{k=1, \dots, [n\delta]+1} |S_k|$$

Soient $\epsilon, \eta > 0$ et posons $T_k = \sum_{j=1}^k V_j 1_{\{|V_j| \leq \eta\sqrt{n}\}}$ pour $k = 1, \dots, [n\delta]+1$. Notons que ces variables aléatoires ne sont pas centrées (contrairement aux variables aléatoires V_j pour $j = 1, \dots, n$). Mettons en place un argument de troncature.

$$\mathbb{P}\left(\max_{k=1, \dots, [n\delta]+1} |S_k| \geq \epsilon\sqrt{n}\right) \leq \mathbb{P}\left(\max_{j=1, \dots, [n\delta]+1} |V_j| \geq \eta\sqrt{n}\right) + \mathbb{P}\left(\max_{k=1, \dots, [n\delta]+1} |T_k| \geq \epsilon\sqrt{n}\right) \quad (2.6.2)$$

Auparavant nous avons déjà montré que le premier terme du membre de droite tend vers 0 lorsque $n \rightarrow +\infty$. Traitons à présent le second terme. Pour cela, il va être nécessaire d'estimer $\mathbb{E}[T_k]$ pour $k = 1, \dots, \lfloor n\delta \rfloor + 1$. Ici, nous avons l'égalité suivante

$$\mathbb{E}[T_k] = \sum_{j=1}^l \mathbb{E}[V_j 1_{\{|V_j| \leq \eta\sqrt{n}\}}] = - \sum_{j=1}^k \mathbb{E}[V_j 1_{\{|V_j| \geq \eta\sqrt{n}\}}]$$

puisque les variables aléatoires $(V_j)_{j=1, \dots, n}$ sont centrées. D'où,

$$\begin{aligned} \mathbb{E}[T_k] &\leq \sum_{j=1}^k \mathbb{E}[|V_j| 1_{\{|V_j| \geq \eta\sqrt{n}\}}] = k \mathbb{E}[|V_1| 1_{\{|V_1| \geq \eta\sqrt{n}\}}] \\ &\leq (\lfloor n\delta \rfloor + 1) \mathbb{P}(|V_1| \geq \eta\sqrt{n})^{1/2} \mathbb{E}[|V_1|^2]^{1/2} \\ &\leq (\lfloor n\delta \rfloor + 1) \frac{1}{\eta^2 n} \end{aligned}$$

d'après l'inégalité de Cauchy-Schwarz, puis celle de Tchebyshev. En conséquence, si δ est suffisamment petit (i.e. $\delta \leq \delta_0(\eta, \epsilon)$), nous avons

$$\mathbb{E}[T_k] \leq \frac{\epsilon}{2} \sqrt{n} \quad \text{pour } k = 1, \dots, \lfloor n\delta \rfloor + 1$$

Ceci entraîne, par l'inégalité triangulaire, que

$$\mathbb{P}\left(\max_{k=1, \dots, \lfloor n\delta \rfloor + 1} |T_k| \geq \epsilon \sqrt{n}\right) \leq \mathbb{P}\left(\max_{k=1, \dots, \lfloor n\delta \rfloor + 1} |T_k - \mathbb{E}[T_k]| \geq \frac{\epsilon}{2} \sqrt{n}\right)$$

Posons ensuite $Z_k = T_k - \mathbb{E}[T_k]$ pour $k = 1, \dots, \lfloor n\delta \rfloor + 1$ et observons qu'il s'agit d'une somme de variables aléatoires centrées, de même loi. En particulier, il s'agit d'une martingale (pour la filtration induite par les variables aléatoires sous-jacentes). Il est donc possible d'utiliser l'inégalité maximale de Doob avec $p = 4$ (cf. [14]) : il existe donc une constante $C > 0$ telle que

$$\mathbb{P}\left(\max_{k=1, \dots, \lfloor n\delta \rfloor + 1} |T_k - \mathbb{E}[T_k]| \geq \frac{\epsilon}{2} \sqrt{n}\right) \leq \frac{C}{\epsilon^4 n^4} \mathbb{E}[|T_{\lfloor n\delta \rfloor + 1} - \mathbb{E}[T_{\lfloor n\delta \rfloor + 1}]|^4]$$

Estimons à présent l'espérance du membre de droite. Pour cela, introduisons les variables aléatoires $\tilde{V}_j$ pour $j = 1, \dots, \lfloor n\delta \rfloor + 1$ définies par

$$T_{\lfloor n\delta \rfloor + 1} - \mathbb{E}[T_{\lfloor n\delta \rfloor + 1}] = \sum_{j=1}^{\lfloor n\delta \rfloor + 1} \left(V_j 1_{\{|V_j| \leq \eta\sqrt{n}\}} - \mathbb{E}[V_j 1_{\{|V_j| \leq \eta\sqrt{n}\}}] \right) = \sum_{j=1}^{\lfloor n\delta \rfloor + 1} \tilde{V}_j$$

Par indépendance des variables $(V_j)_{j=1, \dots, \lfloor n\delta \rfloor + 1}$, il existe des constantes $C_1, C_2 > 0$ telles que

$$\mathbb{E}[|T_{\lfloor n\delta \rfloor + 1} - \mathbb{E}[T_{\lfloor n\delta \rfloor + 1}]|^4] = \mathbb{E}\left[\left|\sum_{j=1}^{\lfloor n\delta \rfloor + 1} \tilde{V}_j\right|^4\right] \leq C_1 \sum_{j=1}^{\lfloor n\delta \rfloor + 1} \mathbb{E}[\tilde{V}_j^4] + C_2 \sum_{i \neq j} \mathbb{E}[\tilde{V}_i^2] \mathbb{E}[\tilde{V}_j^2]$$

Pour conclure la démonstration, nous allons devoir revenir aux variables $(V_j)_{j \geq 1}$. Dans ce but, nous allons utiliser l'inégalité suivante : pour tout $p \in \mathbb{N}_*$, il existe une constante $c_p > 0$ telle que

$$\forall a, b \in \mathbb{R} \quad |a - b|^p \leq c_p (|a|^p + |b|^p)$$

D'où, en utilisant ce qui précède (pour $p = 4$) puis l'inégalité de Jensen, nous obtenons

$$\mathbb{E}[|\tilde{V}_j|^4] = \mathbb{E}[|T_{\lfloor n\delta \rfloor + 1} - \mathbb{E}[T_{\lfloor n\delta \rfloor + 1}]|^4] \leq C\mathbb{E}[V_j^4 1_{\{|V_j| \leq \eta\sqrt{n}\}}], \quad \forall j = 1, \dots, \lfloor n\delta \rfloor + 1$$

Ainsi, nous en déduisons que

$$\begin{aligned} \mathbb{E}[|T_{\lfloor n\delta \rfloor + 1} - \mathbb{E}[T_{\lfloor n\delta \rfloor + 1}]|^4] &\leq C_1 \sum_{j=1}^{\lfloor n\delta \rfloor + 1} \mathbb{E}[V_j^4 1_{\{|V_j| \leq \eta\sqrt{n}\}}] + C_2 \sum_{i \neq j} \mathbb{E}[V_i^2 1_{\{|V_i| \leq \eta\sqrt{n}\}}] \mathbb{E}[V_j^2 1_{\{|V_j| \leq \eta\sqrt{n}\}}] \\ &\leq C_1(\eta^2 n) \sum_{j=1}^{\lfloor n\delta \rfloor + 1} \mathbb{E}[V_j^2] + C_2 \sum_{i \neq j} \mathbb{E}[V_i^2] \mathbb{E}[V_j^2] \\ &\leq C_1(\lfloor n\delta \rfloor + 1)(\eta^2 n) + C_2(\lfloor n\delta \rfloor + 1)^2 \end{aligned}$$

puisque $\mathbb{E}[V_j^2] = 1$ pour tout $j = 1, \dots, \lfloor n\delta \rfloor + 1$. En conclusion, nous avons montré que

$$\begin{aligned} \mathbb{P}\left(\max_{j=1, \dots, \lfloor n\delta \rfloor + 1} |T_k - \mathbb{E}[T_k]| \geq \frac{\epsilon}{2}\sqrt{n}\right) &\leq \frac{C}{\epsilon^4 n^2} \left(C_1(\lfloor n\delta \rfloor + 1)(\eta^2 n) + C_2(\lfloor n\delta \rfloor + 1)^2\right) \\ &\leq \frac{C}{\epsilon^4}(\delta\eta^2 + \delta^2) \end{aligned}$$

En résumé, nous venons de montrer le résultat suivant

Proposition 2.6.3. *Pour tout $\epsilon > 0$, $\lim_{\delta \rightarrow 0} \frac{1}{\delta} \left[\limsup_{n \rightarrow +\infty} \mathbb{P}\left(\max_{k=1, \dots, \lfloor n\delta \rfloor + 1} |S_k| \geq \epsilon\sqrt{n}\right) \right] = 0$*

Il est possible d'en déduire le résultat suivant (cf. [53]) dont la démonstration, reposant sur le même genre de technique, sera admise.

Corollaire 2.6.1. *Avec les notations précédentes, nous avons*

$$\lim_{\delta \rightarrow 0} \left[\limsup_{n \rightarrow +\infty} \mathbb{P}\left(\max_{\substack{l=0, \dots, n \\ k=1, \dots, \lfloor n\delta \rfloor + 1}} |S_{k+l} - S_l| \geq \epsilon\sqrt{n}\right) \right] = 0$$

Ce résultat permet enfin de justifier que la suite de mesures de probabilités $(\mu_n)_{n \geq 1}$ est tendue.

Voici l'énoncé du résultat principal de cette section : le principe d'invariance de Donsker.

Théorème 2.6.1 (Donsker). *Soit $(X^n)_{n \geq 1}$ la suite d'interpolation linéaire introduite plus tôt. Cette suite converge en loi dans $E = C([0, 1])$ vers une variable aléatoire $B = (B_t)_{t \in [0, 1]}$ de E . De plus, l'objet limite est gaussien au sens suivant : pour tout $0 \leq t_1 < t_2 < \dots < t_d \leq 1$, le vecteur aléatoire $(B_{t_1}, \dots, B_{t_d})$ est un vecteur gaussien centré de $\mathbb{R}^d$ de matrice de covariance*

$$\Gamma_{ij} = \mathbb{E}[B_{t_i} B_{t_j}] = \min(t_i, t_j), \quad \forall 1 \leq i, j \leq d$$

et $B_0 = 0$ p.s. La loi μ de B sur $(E, \mathcal{B}(E))$ est appelée : la mesure de Wiener et le processus $(B_t)_{t \in [0, 1]}$ est appelé mouvement brownien.

Démonstration. Notons par $\mu_n = \mathcal{L}(X^n)$, $n \in \mathbb{N}$. Nous avons montré plus tôt que $(\mu_n)_{n \in \mathbb{N}}$ est une suite tendue. Alors, d'après le Théorème de Prokhorov, pour toutes sous-suites de $(\mu_n)_{n \in \mathbb{N}}$ il est possible d'extraire une nouvelle sous-suite qui converge étroitement. Si μ désigne l'une des potentielles lois limites, nous savons (d'après les conditions induites par les lois finies dimensionnelles 2.6.2) qu'il s'agit de la loi d'un processus gaussien $(B)_{t \in [0,1]}$ de E tel que

$$\text{Cov}(B_{t_1}, \dots, B_{t_d}) = \min(t_1, \dots, t_d), \quad \forall 0 \leq t_1 < t_2 < \dots < t_d \leq 1$$

□

Remarque. Implicitement, nous avons utilisé le fait que la connaissance des lois finies dimensionnelles de $(B_{t_1}, \dots, B_{t_d})$ pour tout $0 \leq t_1 < \dots < t_d \leq 1$ suffit à caractériser la loi μ du processus B . En effet, si $x \in E = C([0, 1])$ et si x^k désigne l'interpolation linéaire de x au points $0, \frac{1}{k}, \frac{2}{k}, \dots, 1$ alors $\lim_{k \rightarrow +\infty} \|x^k - x\|_\infty = 0$. En particulier, la variable aléatoire (servant d'interpolation linéaire) X^k est entièrement déterminée par $B_{\frac{1}{k}}, B_{\frac{2}{k}}, \dots$ et donc $\lim_{k \rightarrow +\infty} \|X^k - B\|_\infty = 0$. D'où,

$$\mathbb{E}[f(X^k)] \xrightarrow{k \rightarrow +\infty} \mathbb{E}[f(B)] = \int_E f d\mu, \quad \forall f \in C_b(E, \mathbb{R})$$

2.7 Quelques propriétés du mouvement brownien et de la mesure de Wiener

Passons rapidement en revue quelques propriétés de ce nouvel objet, les démonstrations sont laissées en exercices. Nous étudierons plus en détail le mouvement brownien au chapitre suivant. Dans ce qui suit, $B = (B_t)_{t \in [0,1]} \in E = C([0,1])$ désignera un mouvement brownien et nous noterons par $\mu = \mathcal{L}(B)$ la mesure de Wiener.

Proposition 2.7.1. 1. *Tout d'abord, les trajectoires $t \mapsto B_t$ du mouvement brownien sont continues et $B(0) = 0$ p.s..*

2. *Ce processus est à accroissement indépendants. Autrement dit, pour tout $0 \leq t_1 < \dots < t_d \leq 1$ les variables aléatoires*

$$B_{t_1}, B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \dots, B_{t_d} - B_{t_{d-1}} \quad \text{sont mutuellement indépendantes.}$$

3. *Pour tout $s < t$, $B_t - B_s$ est une variable aléatoire gaussienne centrée de variance $s - t$. Nous noterons ceci par*

$$B_t - B_s \sim \mathcal{N}(0, s - t)$$

4. *$(B_t)_{t \in [0,1]}$ est un processus gaussien : pour tout d -uplet $0 \leq t_1 < \dots < t_d \leq 1$,*

$$(B_{t_1}, B_{t_2}, \dots, B_{t_d}) \sim \mathcal{N}_d(0, \Gamma)$$

avec $\Gamma_{ij} = \min(i, j)$ pour tout $1 \leq i, j \leq d$.

5. *$(X_t)_{t \geq 0}$ est une martingale pour la filtration $(\mathcal{F}_t)_{t \geq 0}$ où $\mathcal{F}_t = \sigma(B_s, s \leq t)$ est la tribu engendré par le processus jusqu'à l'instant t . Autrement dit,*

- (a) $\mathbb{E}[B_t|B_s] = B_s$ pour tout $s < t$;
- (b) pour tout $t \geq 0$, B_t est $\mathcal{F}_t$ -mesurable ;
- (c) pour tout $t \geq 0$, $\mathbb{E}[|B_t|] < \infty$.

Bien sûr, la liste ne s'arrête pas là (cf. [75, 83, 38]) et il est possible d'exhiber de nombreuses autres propriétés.

Présentons rapidement, une autre construction du mouvement brownien. Pour cela, considérons l'espace de Hilbert $H = L^2([0, 1], d\lambda)$ où λ désigne la mesure de Lebesgue et considérons $(h_i)_{i \in \mathbb{N}}$ une base orthonormée de H . Il est possible de choisir, par exemple, la base définie à l'aide des ondelettes de Haar (cf. [83, 38]). Celles-ci s'obtiennent par translation et dilatation de la fonction suivante

$$H(t) = \begin{cases} -1 & \text{si } 0 \leq t < \frac{1}{2} \\ 1 & \text{si } \frac{1}{2} \leq t \leq 1 \\ 0 & \text{sinon} \end{cases}$$

Notons alors, que pour tout entier $n \in \mathbb{N}$, il existe un unique couple $(j, k) \in \mathbb{N}^2$ telle que

$$n = 2^j + k \quad \text{avec} \quad 0 \leq k < 2^j$$

Pour de tels entiers, nous posons alors, pour tout $t \in [0, 1]$, $H_n(t) = 2^{j/2}H(2^j t - k)$ et imposons $H_0(t) = 1$. La famille ainsi définie $(H_n)_{n \geq 0}$ est une base orthonormée de $L^2[0, 1]$.

A partir d'une telle famille, il est possible de construire une base Schauder $(\Delta_n)_n$ de E (i.e. une famille dense d'un espace de Banach). Pour cela, posons

$$\Delta_n(t) = \int_0^t H_n(u) du \quad \text{pour tout } t \in [0, 1].$$

De plus, si $d : [0, 1] \rightarrow \mathbb{R}$ est définie par

$$d(t) = \begin{cases} 2t & \text{si } 0 \leq t < \frac{1}{2} \\ 2(1-t) & \text{si } \frac{1}{2} \leq t \leq 1 \\ 0 & \text{sinon} \end{cases}$$

et $d_n(t) = d(2^j t - k)$ pour $n = 2^j + k$ avec $0 \leq k < 2^j$ alors

$$\Delta_n(t) = \lambda_n d_n(t)$$

où $\lambda_0 = 1$ et, pour $n \geq 1$, $\lambda_n = \frac{1}{2}2^{-j/2}$ avec $n = 2^j + k$ et $0 \leq k < 2^j$. A partir de cette famille, il est possible de construire le mouvement brownien. En effet, soit $(g_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées telle que $g_1 \sim \mathcal{N}(0, 1)$ sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et posons

$$X_t^n = \sum_{i=1}^n g_i \Delta_i(t), \quad t \in [0, 1].$$

Notons qu'il s'agit d'éléments aléatoires de $C([0, 1])$. De plus, nous avons la convergence suivante :

2.7. QUELQUES PROPRIÉTÉS DU MOUVEMENT BROWNIEN ET DE LA MESURE DE WIENER 47

Théorème 2.7.1. $(X^n)_{n \in \mathbb{N}}$ converge presque sûrement et uniformément sur $[0, 1]$ vers une limite X qui est un mouvement brownien.

Démonstration. tout d'abord voyons ce qui se produit au niveau des fonctions de covariances. Vérifions rapidement que $\lim_{n \rightarrow +\infty} \mathbb{E}[X_s^n X_t^n] = \mathbb{E}[X_s X_t] = \min(s, t)$. Pour cela, nous utiliserons le fait que les variables $(g_i)_{i \in \mathbb{N}}$ sont centrées et indépendantes.

$$\begin{aligned} \mathbb{E}[X_s^n X_t^n] &= \mathbb{E}\left[\left(\sum_{i=1}^n g_i \Delta_i(s)\right)\left(\sum_{j=1}^n g_j \Delta_j(t)\right)\right] \\ &= \sum_{i=1}^n \Delta_i(s) \Delta_i(t) \\ &= \sum_{i=1}^n \left(\int_0^s H_i(s) d\lambda\right) \left(\int_0^t H_i(u) d\lambda\right) \\ &= \sum_{i=1}^n \langle H_i, 1_{[0,s]} \rangle_{L^2(d\mathbb{P})} \langle H_i, 1_{[0,t]} \rangle_{L^2(d\mathbb{P})} \end{aligned}$$

Pour conclure, il suffit d'utiliser le Théorème de Parseval qui nous assure que

$$\lim_{n \rightarrow +\infty} \sum_{i=1}^n \langle H_i, 1_{[0,s]} \rangle_{L^2} \langle H_i, 1_{[0,t]} \rangle_{L^2} = \langle 1_{[0,s]}, 1_{[0,t]} \rangle_{L^2} = \min(s, t)$$

Nous reverrons ultérieurement ce genre d'argument et de construction pour démontrer le Théorème de Schilder.

Démontrons à présent la convergence uniforme. Pour cela, nous aurons besoin du Lemme suivant.

Lemme 2.7.1. Dans le cadre précédent, il existe une variable aléatoire C telle que

$$|g_n| \leq C \sqrt{\log n} \quad \text{pour tout } n \geq 1 \quad \text{et} \quad \mathbb{P}(C < \infty) = 1$$

Démonstration. Pour $x \geq 1$, nous avons

$$\begin{aligned} \mathbb{P}(|g_n| \geq x) &= \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du \\ &\leq \sqrt{\frac{2}{\pi}} \int_x^\infty u e^{-u^2/2} du = e^{-x^2/2} \sqrt{\frac{2}{\pi}} \end{aligned}$$

D'où, pour tout $\alpha > 1$, nous obtenons

$$\mathbb{P}(|g_n| \geq \sqrt{2\alpha \log n}) \leq \exp(-\alpha \log n) \sqrt{\frac{2}{\pi}} = n^{-\alpha} \sqrt{\frac{2}{\pi}}$$

De plus, puisque $\alpha > 1$, cette dernière quantité est sommable et le Lemme de Borel-Cantelli entraîne que

$$\mathbb{P}(|g_n| \geq \sqrt{2\alpha \log n} \text{ infiniment souvent}) = 0$$

En conséquence, la variable aléatoire $C = \sup_{2 \leq n \leq \infty} \frac{|g_n|}{\sqrt{\log n}}$ est finie presque sûrement. $\square$

Fort de ce lemme nous pouvons établir la convergence uniforme de $X_n(t)$ vers X_t . Pour cela, observons que pour tout $n \in [2^j, 2^{j+1}[$ nous avons :

- $\log n < j + 1$,
- pour tout $t \in [0, 1]$, $d_n(t) \neq 0$ pour au plus un entier $n \in [2^j, 2^{j+1}[$ (cette localisation du support justifie pleinement l'utilisation des fonctions d_n).

C'est pourquoi, en utilisant le précédent lemme, nous avons, pour tout $M \geq 2^J$ avec $J \geq 1$,

$$\begin{aligned} \sum_{n=M}^{\infty} |g_n| \Delta_n(t) &= \sum_{n=M}^{\infty} |g_n| \lambda_n d_n(t) \leq C \sum_{n=M}^{\infty} \lambda_n \sqrt{\log n} d_n(t) \\ &\leq C \sum_{j=J}^{2^j-1} \sum_{k=0}^1 \frac{1}{2} 2^{-j/2} \sqrt{j+1} d_{2^j+k}(t) \\ &= C \sum_{j=J}^{\infty} \frac{1}{2} 2^{-j/2} \sqrt{j+1} \end{aligned}$$

puisque $0 \leq d_n(t) \leq 1$ pour tout $n \in \mathbb{N}$ et pour tout $t \in [0, 1]$. Le dernier terme des inégalités ci-dessus tend vers zéro lorsque $J \rightarrow +\infty$. Ceci entraîne que presque sûrement X_t^n converge uniformément ; en outre, le processus limite $(X_t)_{t \in [0, 1]}$ est continue presque sûrement.

La démonstration n'est pas entièrement terminée. Il reste à montrer qu'il s'agit d'un processus gaussien. Notons que la forme de la fonction de covariance permet facilement de montrer que les accroissements du processus sont indépendants. A l'aide de cette observation, pour tout $0 \leq t_1 < t_2 < \dots < t_d \leq 1$, il n'est pas difficile de montrer (en passant par la fonction caractéristique) que le vecteur

$$(X_{t_1}, \dots, X_{t_d})$$

est un vecteur gaussien. Ces détails sont laissés à l'attention du lecteur. $\square$

2.8 Remarques historiques

Il est difficile d'obtenir un résumé de l'évolution de certains travaux mathématiques lorsque ceux-ci sont relativement récent. Nous empruntons donc ce qui va suivre aux ouvrages [38, 53, 75, 35].

Tout d'abord expliquons la terminologie des espaces polonais. De l'aveu de Roger Godement, il s'agit initialement d'un trait d'esprit qui a ensuite été adopté dans le langage courant des mathématiciens. En effet, dans son ouvrage *Analyse Mathématique IV*, il dit la chose suivante :

« Abréviation que, pour rire, j'ai suggérée à N. Bourbaki en 1949 après avoir appris le sujet dans la Topologie de Casimir Kuratowski [...] et constaté la contribution des Polonais ; j'aurais du reste aussi bien pu les appeler « polono-russes ». La plaisanterie, qui n'en était pas entièrement une, fut prise au sérieux et tous les experts ont depuis adopté cette étrange terminologie, généralement sans mentionner N. Bourbaki à l'exception de de Kuratowski lui-même qui, dans un petit livre paru en 1974 sur l'histoire des mathématiques polonaises, y voit à juste titre un hommage à celles-ci. »

D'après [74], en 1905 Borel essaya de traduire le théorème des segments emboîtés en terme d'ensemble ; cela donna naissance à la notion de fermés. Dans des ouvrages philosophie mathématiques (traitant de la notion d'infini et de continuité), Russel introduit la notion de « compacité » pour désigner des suites de nombres rationnels vérifiant une certaine propriété. Les idées de Russel sont ensuite développées dans « Principles of mathematics » en 1903. L'introduction de ce nouveau terme se fit presque de manière simultanée avec la définition des ensembles compacts proposée par Fréchet. Cette proposition apparaît en 1906 dans la thèse de Fréchet et fait suite au mouvement (initié par Hilbert) cherchant trouver des espaces généraux permettant d'englober plusieurs théorie mathématiques à la fois (afin d'en saisir plus profondément les points essentiels).

En 1940, Alexandrov semble être le premier à vouloir traiter la convergence en loi dans des espaces généraux. Il fut également l'auteur du Théorème du portmanteau dont la terminologie provient du fait que cet objet est pratique pour rassembler des choses de même nature.

La notion de fonction lipschitzienne provient de travaux de Lipschitz (1864) sur la résolution d'équations différentielles. Il présente également la notion de fonctions hölderiennes qui apparaissent dans le chapitre suivant et qui doivent leur paternité à Hölder en 1852.

L'espace des fonctions lipschitzienne et bornées et de leurs propriétés furent utilisées en théorie des probabilités dès 1953 par Fortet et Mourier. En 1956, Prokhorov introduit la distance ρ qui porte son nom et démontre qu'elle permet de quantifier la convergence en loi.

La notion de mesure régulière semble apparaître dans les travaux de Halmos en 1950 tandis que la terminologie de « tension » semble due à Le Cam (1957).

Le mouvement brownien est un objet probabiliste qui porte le nom du botaniste Brown. Celui-ci a observé le comportement erratiques du mouvement de particules microscopiques suspendues dans un liquide. En 1900, Bachelier étudie également cet objet pour l'appliquer à des problèmes financiers. En 1905, Einstein, met en évidence le lien entre mouvement brownien et équation de la chaleur. Enfin, l'étude rigoureuse ainsi que l'existence de ce processus est du à Wiener en 1923. La plupart des propriétés délicates de cet objet ont été démontrées par Lévy entre 1939 et 1945.

La construction du mouvement brownien à partir des ondelettes provient des travaux de P. Lévy qui simplifia la construction de Wiener en introduisant les fonction de Schauder. La construction initiale de Wiener était plus complexe et utilisait la représentation suivante :

$$B_z = tZ_0 + \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} Z_k \sqrt{2} \frac{\sin \pi k t}{\pi k}$$

qui était alors plus naturelle pour lui, étant donné ses travaux en analyse harmonique.