

Théorèmes limites : du microscopique au macroscopique

Alexandre Richard¹

25 janvier 2025

1. Université Paris-Saclay, CentraleSupélec, alexandre.richard@centralesupelec.fr.

Table des matières

1	Exemples introductifs	3
1.1	Changement d'échelle pour les chaînes de Markov	3
1.1.1	Théorème de Donsker	4
1.1.2	Théorie de Stroock-Varadhan sur l'approximation des diffusions . . .	5
1.2	Limites champ moyen	6
1.2.1	Particules déterministes et équation de Vlasov	6
1.2.2	Particules stochastiques et équations de McKean-Vlasov	9
1.2.3	Propagation du chaos	10
1.3	Limites hydrodynamiques	11
2	Convergence des mesures de probabilité	12
2.1	Rappels sur la convergence des variables aléatoires	12
2.2	Convergence en loi : cas des variables aléatoires dans un espace euclidien . .	14
2.3	Convergence étroite des mesures de probabilité sur un espace métrique . . .	15
2.3.1	Propriétés des mesures	15
2.3.2	Convergence étroite des mesures	17
2.3.3	Espaces polonais et propriétés topologiques de $\mathcal{P}(E)$	20
2.4	Compacité dans $\mathcal{P}(E)$	23
3	Convergence des processus continus (et un peu càdlàg)	27
3.1	Convergence étroite dans l'espace des fonctions continues sur un compact .	27
3.1.1	L'espace $\mathcal{C}$	27
3.1.2	Cylindres et lois finies-dimensionnelles	29
3.1.3	Convergence étroite dans $\mathcal{P}(\mathcal{C})$	31
3.1.4	Tension dans $\mathcal{P}(\mathcal{C})$	32
3.2	Convergence étroite dans l'espace des fonctions continues sur $\mathbb{R}_+$	35
3.3	Théorème de Donsker	36
3.4	Bref aperçu de la tension dans l'espace des fonctions càdlàg	39
4	Théorie de Stroock-Varadhan de l'approximation des diffusions	42
4.1	Rappels sur les processus de Markov et les martingales	42
4.2	Compléments sur les équations différentielles stochastiques et les générateurs	44
4.3	Problèmes de martingales	46
4.4	Résultat principal	47
4.5	L'exemple de l'urne d'Ehrenfest	52

5	Particules diffusives en interaction et propagation du chaos	54
5.1	Couplage entre le système de particules et l'équation de McKean-Vlasov . .	54
5.2	Propagation du chaos	60
5.2.1	Propriétés générales	60
5.2.2	Applications aux systèmes de particules à coefficients Lipschitz . . .	64
5.3	Considérations diverses sur les interactions singulières	65
A	Résultats divers	66
A.1	Topologie et analyse fonctionnelle	66

Chapitre 1

Exemples introductifs

Les premiers théorèmes limites qu'on apprend généralement en probabilités sont la loi des grands nombres et le théorème central limit. L'objectif de ce cours est de présenter quelques résultats classiques qui étendent en quelque sorte ces théorèmes au cas où les variables aléatoires sont à valeurs dans des espaces de dimension infinie, typiquement des espaces fonctionnels.

1.1 Changement d'échelle pour les chaînes de Markov

Théorème 1.1 (Loi forte des grands nombres de Kolmogorov¹). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles i.i.d.. Alors on a équivalence entre :*

$$(i) \quad \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{} \mu \quad p.s. ;$$

$$(ii) \quad X_n \in L^1(\mathbb{P}).$$

Dans le cas où l'une de ces propriétés est vérifiée, $\mu = \mathbb{E}X_n$.

Remarque 1.2. *Des généralisations de la LGN en dimension infinie existent (voir e.g. Kuelbs-Zinn, Ann. Probab. '79). Cependant, ces généralisations ne couvrent pas les cas qui nous intéresseront, soit parce que les espaces fonctionnels sont trop restrictifs, soit parce que nos variables ne seront plus indépendantes.*

Rappel sur le CLT : on a de la convergence pour des v.a. réelles, iid, et de variance finie. On va étudier des situations plus complexes, où les variables aléatoires sont dans des espaces fonctionnels et/ou interagissent entre elles (i.e. elles sont corrélées).

Le TCL fournit une vitesse de convergence dans la loi des grands nombres.

Théorème 1.3 (Théorème Central Limit). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles i.i.d. ayant une variance σ^2 finie. Alors*

$$\sqrt{n} \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}X_i)}{\sigma} \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, 1).$$

1. Mathématicien russe, 1903-1987. Célèbre pour sa formalisation des fondements de la théorie des probabilités, les *Grundbegriffe der Wahrscheinlichkeitsrechnung* parus en 1933. D'importantes contributions également en théorie de la turbulence.

On observe ici un premier phénomène dit “d’universalité” : quelque soit la loi des X_n (de second moment fini), la loi limite dans le TCL est une gaussienne. Nous en rencontrons d’autres.

1.1.1 Théorème de Donsker

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires *i.i.d.* telles que $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$. La suite

$$S_n = \sum_{i=0}^n X_i, \quad n \in \mathbb{N},$$

est une marche aléatoire simple sur $\mathbb{Z}$.

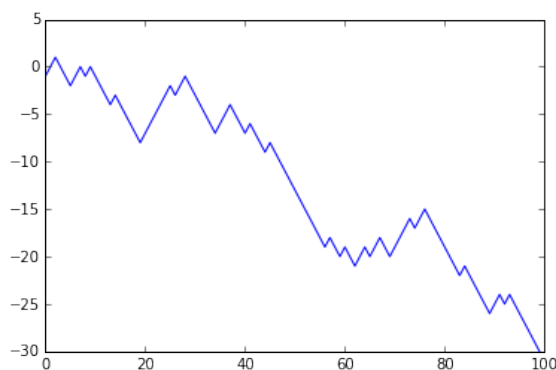


FIGURE 1 – Marche aléatoire sur $\mathbb{Z}$.

Considérons une extension en temps continu de $\{S_n\}_{n \in \mathbb{N}}$. Ainsi, on notera

$$\begin{aligned} \widetilde{W}_t^{(1)} &= S_{[t]}, \quad t \in \mathbb{R}_+, \\ W_t^{(1)} &= S_{[t]} + (t - [t]) X_{[t]+1}, \quad t \in \mathbb{R}_+. \end{aligned}$$

On a $W^{(1)} \in \mathcal{C}(\mathbb{R}_+)$ et $\widetilde{W}^{(1)} \in D(\mathbb{R}_+)$, l’espace des fonctions càdlàg². On voit que c’est $W^{(1)}$ qui est représentée dans la Figure 1.

Réalisons désormais l’expérience suivante : Accélérons le temps d’un facteur n dans le processus $\widetilde{W}^{(1)}$. A $t > 0$ fixé, si on laisse tendre $n \rightarrow +\infty$, on sait par la LGN que $\frac{1}{n} \widetilde{W}_{nt}^{(1)} \rightarrow 0$ p.s. et que $\frac{1}{\sqrt{n}} \widetilde{W}_{nt}^{(1)} \xrightarrow{(d)} \mathcal{N}(0, \sqrt{t})$ par le TCL.

Ainsi, si on souhaite observer de loin une marche aléatoire, le bon changement d’échelle doit se faire à la vitesse $\sqrt{n}$ sur l’axe des ordonnées lorsque l’accélération est en n en temps. C’est le “bon” changement, au sens où tout autre vitesse induirait une limite dégénérée ($+\infty$ si on va trop vite, 0 sinon).

On notera désormais

$$\begin{aligned} W_t^{(n)} &= \frac{1}{\sqrt{n}} W_{nt}^{(1)}, \quad t \in \mathbb{R}_+, \\ \widetilde{W}_t^{(n)} &= \frac{1}{\sqrt{n}} \widetilde{W}_{nt}^{(1)}, \quad t \in \mathbb{R}_+. \end{aligned}$$

2. continues à droite avec une limite à gauche en chaque point.

Pour des temps $0 \leq t_1 < t_2 < \dots < t_k$, une généralisation multidimensionnelle du TCL entraîne que

$$\left(\widetilde{W}_{t_1}^{(n)}, \dots, \widetilde{W}_{t_k}^{(n)} \right) \xrightarrow{(d)} (N_1, \dots, N_1 + \dots + N_k),$$

où $N_1, \dots, N_k$ sont des gaussiennes centrées indépendantes de variance respectives $t_1, t_2 - t_1, \dots, t_k - t_{k-1}$.

Bien qu'intéressante, cette limite ne permet néanmoins pas d'approcher des fonctionnelles telles que $\sup_{t \in [0,1]} W_t^{(\infty)}$, où $W^{(\infty)}$ représente une éventuelle limite de $W^{(n)}$.

Le théorème de Donsker va un cran plus loin, puisqu'il ne considère pas des limites de variables aléatoires comme évoquées précédemment avec le TCL, mais donne la convergence des processus $\{W_t^{(n)}\}_{t \in \mathbb{R}_+}$ et $\{\widetilde{W}_t^{(n)}\}_{t \in \mathbb{R}_+}$.

Théorème 1.4 (Donsker). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles i.i.d. centrées ayant une variance σ^2 finie. Alors*

- ▷ *La suite de processus $(\frac{1}{\sigma} \widetilde{W}^{(n)})_{n \in \mathbb{N}}$ converge en loi dans l'espace $D(\mathbb{R}_+)$ vers le mouvement brownien standard ;*
- ▷ *La suite de processus $(\frac{1}{\sigma} W^{(n)})_{n \in \mathbb{N}}$ converge en loi dans l'espace $\mathcal{C}(\mathbb{R}_+)$ vers le mouvement brownien standard.*

Ce théorème se généralise assez naturellement dans $\mathbb{R}^d$. La limite est alors un mouvement brownien de $\mathbb{R}^d$, c'est-à-dire un processus à valeurs dans $\mathbb{R}^d$ dont les composantes sont des browniens indépendants.

On verra que la preuve repose sur un argument de compacité dans $\mathcal{P}(\mathcal{C}(\mathbb{R}_+))$ (resp. $\mathcal{P}(D(\mathbb{R}_+))$), l'espace des mesures de probabilité sur $\mathcal{C}(\mathbb{R}_+)$ (resp. $D(\mathbb{R}_+)$).

1.1.2 Théorie de Stroock-Varadhan sur l'approximation des diffusions

Au-delà de la convergence de la marche aléatoire simple vers le MB, on peut s'intéresser à la convergence de chaînes de Markov renormalisées vers des processus de diffusions.

Nous prenons l'exemple ici de l'urne d'Ehrenfest³. L'expérience est la suivante : on considère deux urnes A et B reliées entre elles par un étroit tunnel et remplies de N molécules de gaz. A chaque instant (discret), on tire aléatoirement une molécule et on la change d'urne.

Notons Y_n^N le nombre de molécules dans l'urne A au temps $n \in \mathbb{N}$. Y^N est une chaîne de Markov à valeurs dans $\{0, \dots, N\}$ dont les probabilités de transition sont données par :

$$\begin{cases} \mathbb{P}(Y_{n+1}^N = m+1 \mid Y_n^N = m) = \frac{N-m}{N} \\ \mathbb{P}(Y_{n+1}^N = m-1 \mid Y_n^N = m) = \frac{m}{N} \end{cases}, \quad n \in \mathbb{N}, \quad m \in \{0, \dots, N\}.$$

On devine assez facilement que lorsque N et n sont grands, $\frac{1}{N} Y^N$ devrait tendre vers $\frac{1}{2}$ par une sorte de loi des grands nombres. On peut être plus précis que cela en considérant,

3. Physicien autrichien, 1880-1933. Etudiant de Ludwig Boltzmann. A propos de ces deux-là, et en guise d'avertissement, voici une citation de D.L. Goodstein au début de son livre "States of Matter" :

Boltzmann, who spent much of his life studying statistical mechanics, died in 1906, by his own hand. Paul Ehrenfest, carrying on the work, died similarly in 1933. Now it is our turn to study statistical mechanics. Perhaps it will be wise to approach the subject cautiously.

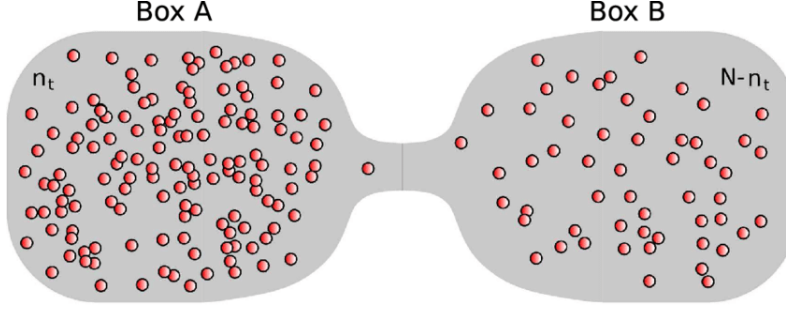


FIGURE 2 – Modèle d’urne d’Ehrenfest.

comme pour la marche aléatoire dans la section précédente, un processus changé de temps et renormalisé. Soit donc $\{X_t^N\}_{t \in [0,1]}$ défini pour tout $N \in \mathbb{N}^*$ par

$$X_t^N = \frac{Y_{[Nt]}^N - N/2}{\sqrt{N}}, \quad t \in [0, 1].$$

Théorème 1.5. *La suite de processus $(X^N)_{N \in \mathbb{N}^*}$ converge en loi dans l’espace $D([0, 1])$ vers le processus d’Ornstein-Uhlenbeck, solution de l’EDS :*

$$dX_t = -X_t dt + dW_t.$$

Ainsi le processus d’Ornstein-Uhlenbeck décrit les fluctuations autour de la moyenne ($\sim \frac{N}{2}$) du nombre de molécules dans l’urne A lorsque N est grand et le temps est accéléré d’un facteur N .

Le théorème précédent sera énoncé et prouvé de façon plus générale pour diverses chaînes de Markov et nous identifierons à chaque fois la limite comme solution d’une équation différentielle stochastique (EDS). Nous verrons que la preuve repose à nouveau sur un argument de compacité. Il faudra ensuite montrer que toute-suite qui converge, converge vers le même objet. Pour identifier cette objet, nous aurons recours à ce qu’on appelle un problème de martingales.

Comme son nom l’indique, cette théorie est issue des travaux de Stroock et Varadhan (voir notamment le chapitre 11 de [18]).

1.2 Limites champ moyen

1.2.1 Particules déterministes et équation de Vlasov

Considérons un ensemble de particules ponctuelles toutes chargées négativement, par exemple des électrons. Les lois fondamentales de la dynamique donnent pour le mouvement de ces électrons, le système d’équations suivant, où X_t^i représente la position de la particule i :

$$\frac{d^2}{dt^2} X_t^i = \sum_{j=1}^N F^0(X_t^i - X_t^j), \quad i = 1 \dots N, \quad t > 0, \quad (1.1)$$

avec $F^0(x) = \nabla V(x)$ pour tout $x \in \mathbb{R}^3$ et V le potentiel coulombien $V(x) = \alpha \frac{1}{|x|}$, $\alpha \in \mathbb{R}$. En introduisant $V_t^i = \frac{d}{dt} X_t^i$, on obtient le système d'EDO d'ordre 1 suivant :

$$\begin{cases} \frac{d}{dt} X_t^i = V_t^i \\ \frac{d}{dt} V_t^i = \sum_{j=1}^N F^0(X_t^i - X_t^j) \end{cases}, \quad i = 1 \dots N, \quad t > 0. \quad (1.2)$$

Quitte à changer légèrement le formalisme du problème, posons $z = (x, v)$, $z' = (x', v') \in \mathbb{R}^6$,

$$\bar{F}^0(z, z') = \begin{pmatrix} v \\ F^0(x - x') \end{pmatrix} \in \mathbb{R}^6,$$

avec x représentant la coordonnée de position, v la coordonnée de vitesse. En posant $Z_t^i = \begin{pmatrix} X_t^i \\ V_t^i \end{pmatrix}$, le problème s'écrit alors

$$\frac{d}{dt} Z_t^i = \sum_{j=1}^N \bar{F}^0(Z_t^i, Z_t^j), \quad i = 1 \dots N, \quad t > 0. \quad (1.3)$$

On considérera donc des équations sous cette forme plutôt que sous la forme cinétique (1.1)-(1.2).

Suivant la nature du noyau d'interaction F^0 , les théorèmes de Cauchy-Lipschitz donneront l'existence et l'unicité globale ou locale d'une solution de (1.3). Ainsi, la solution n'est que locale dans le cas du potentiel de Coulomb. Par ailleurs, le temps T^* d'existence de la solution dépend *a priori* du nombre N de particules...

Remise à l'échelle du problème.

Lorsque le système est composé d'un grand nombre de particules, il est naturel de considérer la limite $N \rightarrow +\infty$. Mais au préalable, il est nécessaire de s'interroger sur l'ordre de grandeur de la somme dans l'équation (1.3). *A priori*, on aimerait que cette somme reste un $O(1)$ lorsque $N \rightarrow +\infty$.

Première solution, accélérer le temps d'un facteur $\sqrt{N}$ dans l'équation (1.1) pour obtenir une équation satisfaite par $\tilde{X}_t^i := X_{\sqrt{N}t}^i$:

$$\frac{d^2}{dt^2} \tilde{X}_t^i = \frac{1}{N} \sum_{j=1}^N F^0(\tilde{X}_t^i - \tilde{X}_t^j), \quad i = 1 \dots N, \quad t > 0.$$

La justification physique d'un tel changement d'échelle n'apparaît pas forcément évidente.

Deuxième solution, on considère le système de N particules à "masse" totale constante égale à 1. Alors il faut tenir compte de la masse $1/N$ de chaque particule quand on écrit le principe fondamental de la dynamique (1.1), de même qu'il faut tenir compte des caractéristiques "physiques" de la force F^0 . Ainsi, pour F^0 une force de type gravitationnelle, le produit des masses des deux particules intervient dans chaque terme $F^0(X_t^i - X_t^j)$. Par conséquent, $F^0(X_t^i - X_t^j) =: \frac{1}{N^2} \tilde{F}^0(X_t^i - X_t^j)$, où on considère que $\tilde{F}^0$ ne dépend plus de N , contrairement à F^0 . Alors

$$\frac{d^2}{dt^2} X_t^i = \frac{1}{N} \sum_{j=1}^N \tilde{F}^0(\tilde{X}_t^i - \tilde{X}_t^j), \quad i = 1 \dots N, \quad t > 0,$$

ce qu'on résume de la façon suivante :

$$\frac{d}{dt} Z_t^i = \frac{1}{N} \sum_{j=1}^N F(Z_t^i, Z_t^j), \quad i = 1 \dots N, \quad t > 0. \quad (1.4)$$

Limite $N \rightarrow +\infty$. Ce type de limite est très étudiée en théorie cinétique des gaz, et constitue un bon modèle pour les plasmas.

On considère désormais uniquement l'équation (1.4) avec des particules à valeurs dans $\mathbb{R}^d$. Le bon objet à considérer pour étudier la limite $N \rightarrow +\infty$ est la mesure empirique du système, donnée par :

$$\mu_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{Z_t^i}.$$

Si F est de classe $\mathcal{C}^1$, cette mesure est solution au sens des distributions de l'équation de Vlasov ⁴ ([20, Prop. 4]) :

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (\mu F[\mu]) = 0, \quad (1.5)$$

où

$$F[\mu](x) = \int_{\mathbb{R}^d} F(x, y) \mu(dy).$$

On observe ici que la condition initiale est donnée par la position des particules. La question qui se pose est donc la suivante :

En supposant que pour une mesure μ_0 donnée on a $\mu_0^N \xrightarrow{} \mu_0$, peut-on en déduire que $\mu_t^N \xrightarrow{*} \mu_t$ pour tout $t > 0$, où $\{\mu_t\}_{t \geq 0}$ est solution de l'équation de Vlasov (1.5) partant de μ_0 ?*

Stratégie #1. Une stratégie très générale consiste à montrer que la famille (μ^N) est compacte dans $\mathcal{P}(\mathcal{C}(\mathbb{R}_+; \mathbb{R}^d))$. Pour cela on utilisera le théorème de Prokhorov (voir Chapitre 2) qui donne un critère de compacité dans les espaces de mesures.

Puisque toute sous-suite de μ^N qui converge, converge vers une solution de (1.5), il reste à démontrer l'unicité dans cette équation (chose qu'on ne fera pas dans ce cours). Ainsi, de toute sous-suite on peut extraire une sous-suite qui converge vers l'unique solution μ de l'équation de Vlasov, donc $\mu^N \xrightarrow{*} \mu$.

Stratégie #2. On introduira dans les prochains chapitres certaines distances, dites de Wasserstein, sur les mesures. L'inégalité suivante, obtenue indépendamment par Braun and Hepp [4] et Dobrušin [5] à la fin des années 70, permet aisément de répondre à la question précédente.

Théorème 1.6. *Soit $\mu_0, \nu_0 \in \mathcal{P}(\mathbb{R}^d)$. Soit $\{\mu_t\}_{t \geq 0}$ une solution de (1.5) associée à la CI μ_0 , et soit $\{\nu_t\}_{t \geq 0}$ une solution de (1.5) associée à la CI ν_0 . Alors*

$$\mathcal{W}_1(\mu_t, \nu_t) \leq e^{2(1 \vee \|F\|_{Lip})t} \mathcal{W}_1(\mu_0, \nu_0), \quad \forall t > 0.$$

On note qu'un tel théorème donne directement l'unicité des solutions (au sens des distributions) dans l'équation de Vlasov.

4. Physicien russe, 1908-1975. Connue pour ses travaux en mécanique statistique et en physique des plasmas.

Remarque 1.7. *Tout ce qui précède ne fonctionne que pour des noyaux d'interaction réguliers. Divers progrès ont eu lieu depuis les années 70 pour traiter de noyaux moins réguliers. Une avancée majeure a été réalisée récemment par Serfaty [17] qui a introduit une nouvelle méthode permettant de traiter des noyaux d'interaction coulombiens répulsifs.*

1.2.2 Particules stochastiques et équations de McKean-Vlasov

On considère désormais un système de particules bruitées. Soient $W^1, \dots, W^N$ des $\mathbb{R}^d$ -mouvements browniens indépendants. L'analogie stochastique de l'équation (1.4) est

$$dX_t^i = \frac{1}{N} \sum_{j=1}^N F(X_t^i - X_t^j) dt + dW_t^i, \quad i = 1, \dots, N, \quad t > 0. \quad (1.6)$$

Cette fois-ci, la mesure empirique ne peut être solution d'une EDP pour au moins deux raisons : d'une part, cette mesure est *a priori* aléatoire du fait de la présence des browniens, il y a donc peu de chances qu'elle puisse résoudre une EDP. D'autre part, on peut s'attendre à ce que l'équivalent de l'équation de Vlasov dans ce cadre fasse intervenir un laplacien, et donc s'écrive :

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (\mu F[\mu]) = \frac{1}{2} \Delta \mu. \quad (1.7)$$

Or par régularisation parabolique, la solution d'une telle équation ne peut être une somme de Dirac, μ^N n'est donc certainement pas solution de (1.7). On reparlera de cette équation dans la sous-section suivante.

Remarquons pour l'instant qu'en notant μ_t^N la mesure empirique de (1.6), on peut réécrire ce système comme suit :

$$dX_t^i = \mu_t^N * F(X_t^i) dt + dW_t^i, \quad i = 1, \dots, N, \quad t > 0.$$

Un candidat naturel au passage à la limite est donc

$$\begin{cases} dX_t = \mu_t * F(X_t) dt + dW_t \\ \mathcal{L}(X_t) = \mu_t \end{cases}, \quad t \geq 0. \quad (1.8)$$

C'est l'EDS dite de McKean-Vlasov, et on nomme souvent la solution “processus non-linéaire”. On verra comment résoudre cette équation par un argument de point fixe.

La question qui se pose est de savoir si lorsque N est grand, les trajectoires des particules (1.6) ressemblent à la solution de l'équation de McKean-Vlasov.

Pour répondre, plaçons-nous sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel est définie une famille de browniens $(W^i)_{i \in \mathbb{N}}$ ainsi que des variables aléatoires intégrables $(X_0^i)_{i \in \mathbb{N}}$. On considère le système de particules (1.6) dirigées par cette famille de browniens et dont la condition initiale est donnée par les X_0^i . Définissons également $\tilde{X}^i$ la solution de l'équation (1.8) issue de X_0^i et dirigée par le brownien W^i .

Par un argument de couplage que nous verrons par la suite, l'inégalité suivante est vérifiée (toujours sous l'hypothèse que F est Lipschitz) :

$$\mathbb{E} \left[\sup_{t \in [0, T]} |X_t^i - \tilde{X}_t^i| \right] \leq \frac{C}{\sqrt{N}}, \quad \forall N \in \mathbb{N}^*.$$

Une référence classique pour ces questions ainsi que les notions de propagation du chaos de la sous-section suivante est le cours de Sznitman [19].

1.2.3 Propagation du chaos

Lorsque nous avons introduit le processus nonlinéaire au paragraphe précédent, nous sommes finalement restés au niveau microscopique lors du passage à la limite $N \rightarrow +\infty$.

Pour passer au niveau macroscopique, et donc au niveau des EDP, regardons la loi des particules. Elles satisfont un système d'EDP de Fokker-Planck (linéaire, obtenues classiquement par application de la formule d'Itô). Nous allons désormais faire une première hypothèse très importante sur la loi des particules : supposons qu'initialement, les particules sont échangeables, c'est-à-dire que pour tout N , la loi de $(X_0^1, \dots, X_0^N)$ est invariante par permutation de l'ordre de ces variables aléatoires. Alors en tout temps $t > 0$, les particules sont encore échangeables.

On pourrait alors essayer de chercher une équation satisfaite par μ_t^N en l'évaluant contre des fonctions test, mais on aboutirait à une EDS de grande dimension. A la place, considérons une fonction test φ et observons que sous l'hypothèse d'échangeabilité, on a

$$\mathbb{E}\langle \mu_t^N, \varphi \rangle = \mathbb{E}\varphi(X_t^1).$$

La loi de μ_t^N se ramène donc à celle de X_t^1 , or pour la loi de X_t^1 on trouve par le calcul (par la formule d'Itô) que $\mu_t = \mathcal{L}(X_t^1)$ vérifie

$$\frac{\partial}{\partial t} \mu_t(x) = \frac{1}{2} \Delta \mu_t(x) - \nabla \cdot \left(\int_{\mathbb{R}^d} F(x-y) \bar{\mu}_t(x, dy) \right),$$

où $\bar{\mu}_t = \mathcal{L}(X_t^1, X_t^2)$. Cette équation n'est donc pas fermée.

Observons néanmoins que si les particules étaient indépendantes, on aurait $\bar{\mu}_t = \mathcal{L}(X_t^1) \otimes \mathcal{L}(X_t^2)$ et l'équation précédente donnerait précisément l'EDP de McKean-Vlasov (1.7).

En supposant que les particules sont initialement indépendantes, i.e. pour tout $N \in \mathbb{N}$,

$$\mathcal{L}(X_0^1, \dots, X_0^N) = \mathcal{L}(X_0^1) \otimes \dots \otimes \mathcal{L}(X_0^N),$$

on dira qu'un système *propage le chaos* si pour tout $k \in \mathbb{N}^*$, et pour tout $t > 0$,

$$\begin{aligned} \mathcal{L}(X_t^{1,N}, \dots, X_t^{k,N}) &\xrightarrow{N \rightarrow \infty} \mathcal{L}(X_t^{1,\infty}) \otimes \dots \otimes \mathcal{L}(X_t^{k,\infty}) \\ &= \mathcal{L}(X_t)^{\otimes k}. \end{aligned}$$

On pourra alors démontrer la convergence (en un sens à préciser) de la mesure empirique (aléatoire) vers la solution de l'EDP (1.7).

L'idée de propagation du chaos était déjà présente dans les travaux de Boltzmann⁵. Il utilisait la notion de chaos moléculaire pour décrire le fait qu'un très grand nombre de particules se comportent statistiquement comme des particules indépendantes.

La notion de propagation du chaos a été formalisée par Kac [8]⁶.

5. Physicien autrichien, 1844-1906. Père fondateur de la physique statistique et de la théorie cinétique des gaz.

6. Mathématicien américain, 1914-1984.

1.3 Limites hydrodynamiques

Mentionnons une dernière classe de modèles très intéressants, que nous n'étudierons pas davantage dans ce cours. Il s'agit de modèles de systèmes de particules sur un espace discret (souvent fini, typiquement $\mathbb{T}_N^d$), dont on observe la dynamique par changement d'échelle.

Nous allons présenter le processus d'exclusion sur $\mathbb{T}_N = \{1, \dots, N\}$. Dans ce modèle, chaque site dispose d'une horloge poissonnienne et ne peut être occupé au plus que par une particule. Lorsqu'une horloge sonne et que le site est occupé, la particule se déplace d'un site aléatoirement vers sa droite ou sa gauche, si le site est libre.

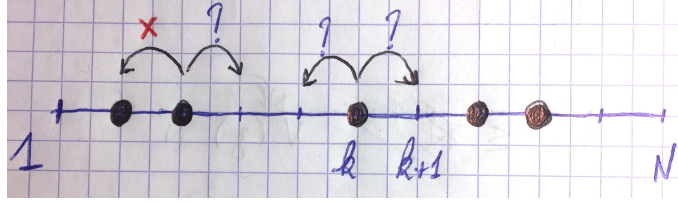


FIGURE 3 – Processus d'exclusion.

Ce système de particules est un processus de Markov $\{\eta(t)\}_{t \geq 0}$ à valeurs dans $\{0, 1\}^{\mathbb{T}_N}$. Supposons que la loi de Poisson des horloges est de paramètre 1 et que la probabilité pour une particule de se déplacer vers la gauche ou la droite est $\frac{1}{2}$. Alors le générateur infinitésimal de ce processus s'écrit :

$$L_N f(\eta) = \frac{1}{2} \sum_{|x-y|=1} (f(\eta^{x,y}) - f(\eta)), \quad \forall \eta \in \{0, 1\}^{\mathbb{T}_N},$$

pour une fonction $f : \{0, 1\}^{\mathbb{T}_N} \rightarrow \mathbb{R}$, où

$$(\eta^{x,y})_z = \begin{cases} \eta_y & \text{si } z = x \\ \eta_x & \text{si } z = y \\ \eta_z & \text{si } z \notin \{x, y\} \end{cases}.$$

Correctement renormalisée, la mesure empirique de ce système converge vers la solution de l'équation de la chaleur. C'est ce qu'on appelle une limite hydrodynamique. On peut aussi s'intéresser aux fluctuations autour de la limite, et montrer qu'elles convergent vers une EDP stochastique.

Comparé aux systèmes de particules de la section précédente, on observe que toutes les particules n'interagissent pas les unes avec les autres, mais uniquement avec les particules voisines.

Processus de contact (long-range ou pas), processus de vote, processus d'exclusion faiblement asymétrique sont autant de variantes de cette dynamique. Le dernier mentionné fait l'objet de nombreuses attentions ces derniers temps, car ses fluctuations convergent par changement d'échelle vers l'EDPS de Kardar-Parisi-Zhang :

$$\partial_t h = \partial_{xx}^2 h + \lambda (\partial_x h)^2 - \infty + \xi,$$

où ξ est un bruit blanc espace-temps.

Une référence classique pour ce type de problèmes est le livre de Kipnis and Landim [10].

Chapitre 2

Convergence des mesures de probabilité

Ce chapitre est essentiellement inspiré du premier chapitre du livre de Billingsley [2]. Certains éléments proviennent également du très riche monographie de Bogachev [3]. On pourra aussi consulter les notes de cours de Miermont [14] et Méliot [13].

2.1 Rappels sur la convergence des variables aléatoires

On rappelle brièvement les différentes notions de convergences pour les suites de variables aléatoires. Sauf mention contraire, on munira tout espace topologique de sa tribu de Borel. Sur un tel espace (E, τ) , on notera $\mathcal{C}_b(E)$ l'espace des fonctions continues et bornées de E dans $\mathbb{R}$.

- ▷ *Convergence presque sûre* : Soit (E, τ) un espace topologique et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités. Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ à valeurs dans (E, τ) (toutes définies sur Ω) converge p.s. vers une variable aléatoire $X : (\Omega, \mathcal{F}) \rightarrow (E, \tau)$ si

$$\mathbb{P}(X_n \longrightarrow X) = 1.$$

- ▷ *Convergence en norme L^p* : Soit $(E, \|\cdot\|)$ un espace de Banach et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités. Pour $p \geq 1$, on construit $L^p(\Omega; E)$ en complétant l'espace des fonctions étagées à valeurs dans E . Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ à valeurs dans $(E, \|\cdot\|)$ (toutes définies sur Ω) converge en norme $L^p(\Omega; E)$, $p \geq 1$, vers une variable aléatoire $X : (\Omega, \mathcal{F}) \rightarrow (E, \|\cdot\|)$ si

$$\lim_{n \rightarrow +\infty} \mathbb{E} [\|X_n - X\|^p] = 0.$$

- ▷ *Convergence en probabilité* : Soit (S, d) un espace métrique et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités. Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ à valeurs dans (S, d) (toutes définies sur Ω) converge en probabilité vers une variable aléatoire $X : (\Omega, \mathcal{F}) \rightarrow (S, d)$ si

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

ou de manière équivalente si

$$\lim_{n \rightarrow +\infty} \mathbb{E}[1 \wedge d(X_n, X)] = 0.$$

- ▷ *Convergence en loi* : Soit (E, τ) un espace topologique, qu'on munit de sa tribu de Borel. Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ à valeurs dans (E, τ) converge en loi vers une variable aléatoire X si pour tout $\varphi \in \mathcal{C}_b(E)$,

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)].$$

Quelques remarques :

Remarque 2.1. *Les implications et contre-exemples entre les différentes notions de convergence ont été vue en première année, nous ne reviendrons pas dessus.*

Remarque 2.2. *Observez que ces notions ne nécessitent pas toutes le même niveau d'information sur l'espace d'arrivée des variables aléatoires.*

En outre, les variables aléatoires n'ont même pas besoin d'être définies sur le même espace de probabilités pour la convergence en loi.

Remarque 2.3. *En termes de loi des variables aléatoires (i.e. de mesures), on peut réexprimer la convergence en loi ainsi : si pour tout $\varphi \in \mathcal{C}_b(E)$,*

$$\lim_{n \rightarrow +\infty} \int_E \varphi(x) \mathbb{P}_{X_n}(dx) = \int_E \varphi(x) \mathbb{P}_X(dx),$$

on dira dans la prochaine section que $(\mathbb{P}_{X_n})_{n \in \mathbb{N}}$ convergence étroitement vers $\mathbb{P}_X$. On aura fréquemment recours à plusieurs notations équivalentes pour les intégrales ci-dessus : pour μ une mesure et $\varphi \in L^1(\mu)$, on notera $\mu(\varphi) \equiv \langle \mu, \varphi \rangle \equiv \int_E \varphi d\mu$.

Enfin, rappelons un résultat classique qui permet de passer d'estimées sur la loi d'une suite d'événements $(A_n)_{n \in \mathbb{N}}$ à une information presque sûre concernant la réalisation de $\limsup A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k = \{A_n \text{ infiniment souvent}\}$.

Lemme 2.4 (Borel-Cantelli). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $(A_n)_{n \in \mathbb{N}}$ une suite d'événements. On a*

$$(i) \text{ si } \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty, \text{ alors } \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k\right) = 0;$$

(ii) (loi de Borel du 0-1) *si on suppose en plus que les $(A_n)_{n \in \mathbb{N}}$ sont indépendants, alors*

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k\right) \text{ vaut } 0 \text{ ou } 1 \text{ suivant que } \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty \text{ ou } \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \infty.$$

Démonstration. Exercice. □

Exercice 2.5. *On considère la suite de v.a. à valeur dans $\mathbb{R}$ définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ par*

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2n \log(n+1)} \quad \text{et} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n \log(n+1)}.$$

Soit $Y_n = \frac{1}{n} \sum_{k=0}^n X_k$. Montrer que Y_n converge en probabilité mais ne converge pas p.s.

Exercice 2.6. *Montrer que si une suite $(X_n)_{n \in \mathbb{N}}$ à valeurs dans un espace de Banach converge en probabilité vers X , on peut en extraire une sous-suite qui converge p.s.*

Remarque 2.7. La convergence p.s. et en probabilité sont des notions topologiques. C'est évident pour la convergence p.s., ça l'est un peu moins pour la convergence en probabilités : cela signifie que si d et d' sont deux distances définissant la même topologie, alors la suite converge pour la distance d si et seulement si elle converge pour d' (utiliser l'exercice 2.6). En conséquence, on peut définir la convergence en probabilité sur des espaces métrisables.

2.2 Convergence en loi : cas des variables aléatoires dans un espace euclidien

Vous avez déjà rencontré plusieurs caractérisations de la convergence en loi pour les v.a. à valeurs dans $\mathbb{R}$ ou dans $\mathbb{R}^d$. Ainsi, le théorème de convergence de Lévy donne une caractérisation par les fonctions caractéristiques (voir Exercice 2.9).

Intéressons-nous plus particulièrement à une caractérisation déjà rencontrée en première année pour les variables aléatoires à valeurs dans $\mathbb{R}$. On a l'équivalence suivante :

- ▷ $X^n \xrightarrow{(d)} X$;
- ▷ Pour tout x tel que $\mathbb{P}_X(\{x\}) = 0$, $\mathbb{P}_{X^n}((-\infty, x]) \rightarrow \mathbb{P}_X((-\infty, x])$.

Nous n'allons pas démontrer ce résultat immédiatement (admis en première année), mais il fera partie d'un théorème plus général sur les espaces métriques que nous traiterons dans la prochaine section.

Exemple 2.8. Soit $X_k \sim \mathcal{B}(p)$ des variables aléatoires de Bernoulli iid. Par le TCL, on a que

$$S_n := \sqrt{n} \frac{\frac{1}{n} \sum_{k=1}^n X_k - p}{\sqrt{p(1-p)}} \xrightarrow{(d)} \mathcal{N}(0, 1).$$

Considérons l'ensemble

$$A = \left\{ \sqrt{n} \frac{\frac{k}{n} - p}{\sqrt{p(1-p)}}, n \in \mathbb{N}^*, k \in \{0, \dots, n\} \right\},$$

alors $\mathbb{P}(S_n \in A) = 1$ pour tout n , et pourtant $\mathbb{P}(\mathcal{N}(0, 1) \in A) = 0$. Cependant, cela ne contredit pas la convergence de S_n vers $\mathcal{N}(0, 1)$. Nous verrons en effet que $\mathbb{P}(S_n \in A) \rightarrow \mathbb{P}(\mathcal{N}(0, 1) \in A)$ lorsque $\mathbb{P}(\mathcal{N}(0, 1) \in \partial A) = 0$, ce qui n'est pas le cas ici puisque $\mathbb{P}(\mathcal{N}(0, 1) \in \partial A) = \mathbb{P}(\mathcal{N}(0, 1) \in \mathbb{R}) = 1$.

Exercice 2.9. Le but de cet exercice est de démontrer le Théorème de Lévy, dont on rappelle l'énoncé :

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ telles que : (i) pour tout $t \in \mathbb{R}^d$, $\mathbb{E}[e^{i\langle t, X_n \rangle}]$ converge vers une limite notée $L(t)$; (ii) la fonction $t \mapsto L(t)$ est continue en 0.

Alors il existe une variable aléatoire X telle que $L(t) = \mathbb{E}[e^{i\langle t, X \rangle}]$ et $X_n \xrightarrow{(d)} X$.

Montrons le résultat dans le cas $d = 1$.

1) Montrer qu'il existe $C > 0$ telle que pour toute mesure de probabilité μ sur $\mathbb{R}$ et pour tout $K > 0$,

$$\mu(\mathbb{R} \setminus [-K, K]) \leq C K \int_{[-\frac{1}{K}, \frac{1}{K}]} (1 - \operatorname{Re}(\mathbb{E}_\mu[e^{itX}])) dt.$$

2) En utilisant la question précédente, montrer que pour tout $\varepsilon > 0$, il existe $K > 0$ tel que

$$\sup_{n \in \mathbb{N}} \mathbb{P}_{X_n}(\mathbb{R} \setminus [-K, K]) \leq \varepsilon.$$

On verra plus tard que cette propriété donne de la compacité et implique qu'il existe une sous-suite $(X_{\varphi(n)})_{n \in \mathbb{N}}$ qui converge en loi vers une certaine variable aléatoire X . On admet ce résultat pour l'instant.

3) Conclure.

On voit que la preuve repose sur un schéma classique : existence de sous-suites convergentes par un argument de compacité, puis identification de la limite grâce à une forme d'unicité.

Dans la suite, on va s'intéresser aux propriétés des mesures sur un espace métrique, et à leur convergence. Un des objectifs sera d'obtenir la métrisabilité de la convergence étroite. Une telle propriété nous permettra par la suite d'avoir des caractérisations séquentielles sur l'espace des mesures de probabilité, ce qui sera bien utile quand on mettra en place (à plusieurs reprises) la stratégie suivante pour prouver la convergence : existence de sous-suites convergentes par compacité, puis identification de la limite.

2.3 Convergence étroite des mesures de probabilité sur un espace métrique

2.3.1 Propriétés des mesures

Définition 2.10. Soit (E, τ) un espace topologique. On note $\mathcal{B}(E)$ la tribu de Borel, i.e. la plus petite tribu contenant les ouverts ($\mathcal{B}(E) = \sigma(\tau)$). On appelle

- (i) Mesure de Borel, toute mesure définie sur $(E, \mathcal{B}(E))$;
- (ii) Mesure régulière, toute mesure de Borel μ telle que pour tout $A \in \mathcal{B}(E)$,

$$\begin{aligned} \mu(A) &= \sup\{\mu(F) : F \text{ fermé inclus dans } A\} \\ &= \inf\{\mu(U) : U \text{ ouvert contenant } A\} ; \end{aligned}$$

- (iii) Mesure de Radon finie, toute mesure de Borel μ finie telle que pour tout $A \in \mathcal{B}(E)$ et tout $\varepsilon > 0$, il existe un compact $K_\varepsilon \subset A$ tel que $\mu(A \setminus K_\varepsilon) < \varepsilon$.

Sauf mention contraire, on ne considérera par la suite que des mesures de probabilité. On notera $\mathcal{P}(E)$ l'ensemble des mesures de probabilité de Borel, donc sur $(E, \mathcal{B}(E))$.

Proposition 2.11. Soit (E, τ) un espace métrisable.

- 1) Toute mesure $\mu \in \mathcal{P}(E)$ est régulière.
- 2) Soit d une distance qui métrise la topologie de E . Alors, toute mesure $\mu \in \mathcal{P}(E)$ est caractérisée par les valeurs de $\mu(f) \equiv \int_E f d\mu$, pour $f \in \operatorname{Lip}(E, d)$, l'ensemble des fonctions lipschitziennes pour d .

Démonstration. Commençons par la première assertion. Soit d une distance qui métrise la topologie de E . Si A est fermé, on peut prendre $F = A$ et $U = A^\delta := \{x : d(x, A) < \delta\}$ pour un $\delta > 0$ suffisamment petit. En effet, on rappelle que par définition des mesures, $\mu(A^\delta)$ décroît vers $\mu(A)$.

Considérons l'ensemble $\mathcal{E}$ des éléments de $\mathcal{B}(E)$ qui vérifie la propriété désirée. Puisque $\mathcal{E}$ contient les fermés, il nous suffit de montrer que $\mathcal{E}$ est une tribu.

On commence par remarquer que $\emptyset \in \mathcal{E}$. De plus, $\mathcal{E}$ est stable par passage au complémentaire. En effet, si $A \in \mathcal{E}$ alors pour tout $\varepsilon > 0$, il existe $F \subset A \subset U$ avec F fermé et U ouvert tels que $\mu(U \setminus F) < \varepsilon$. On a également par passage au complémentaire que $E \setminus U \subset E \setminus A \subset E \setminus F$ et $\mu((E \setminus F) \setminus (E \setminus U)) = \mu(U \setminus F) < \varepsilon$. Donc $E \setminus A \in \mathcal{E}$.

Soit $(A_n)_{n \in \mathbb{N}}$ une suite de $\mathcal{E}$. Soit $\varepsilon > 0$ et soient $(F_n)_{n \in \mathbb{N}}$ une suite de fermés et $(U_n)_{n \in \mathbb{N}}$ une suite d'ouverts tels que pour tout n , $F_n \subset A_n \subset U_n$ et $\mu(U_n \setminus F_n) < \frac{\varepsilon}{2^{n+1}}$. Notons $U = \bigcup_{n \in \mathbb{N}} U_n$ et $F = \bigcup_{n \leq N} F_n$, où N est tel que $\mu(\bigcup_{n \in \mathbb{N}} F_n \setminus F) < \frac{\varepsilon}{2}$. On a alors $F \subset \bigcup_{n \in \mathbb{N}} A_n \subset U$ et

$$\begin{aligned} \mu(U \setminus F) &\leq \mu\left(U \setminus \bigcup_{n \in \mathbb{N}} F_n\right) + \mu\left(\bigcup_{n \in \mathbb{N}} F_n \setminus F\right) \\ &\leq \sum_{n \in \mathbb{N}} \mu(U_n \setminus \bigcup_{n \in \mathbb{N}} F_n) + \frac{\varepsilon}{2} \\ &\leq \sum_{n \in \mathbb{N}} \mu(U_n \setminus F_n) + \frac{\varepsilon}{2} \\ &< \varepsilon. \end{aligned}$$

Donc $\mathcal{E}$ est stable par union dénombrable, donc c'est une tribu. Puisqu'elle contient tous les fermés, elle contient $\mathcal{B}(E)$.

Montrons maintenant la deuxième partie de cette proposition. Nous venons de voir que μ est déterminée par ses valeurs sur les fermés (ou sur les ouverts). Il suffit donc de montrer que pour tout fermé F , l'indicatrice $\mathbb{1}_F$ peut être approchée par des fonctions lipschitziennes. Pour $\varepsilon > 0$, posons

$$f_{\varepsilon, F}(x) = \max(0, 1 - \varepsilon^{-1} d(x, F)). \quad (2.1)$$

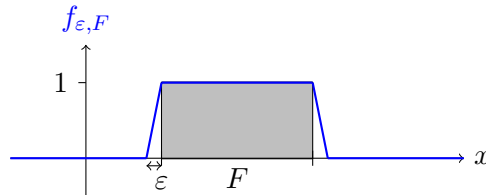


FIGURE 1 – “J’ai montré mon chef-d’oeuvre aux grandes personnes et je leur ai demandé si mon dessin leur faisait peur.”

Alors cette fonction est ε^{-1} -lipschitzienne, bornée par 1, et converge (ponctuellement) en décroissant vers $\mathbb{1}_F$ quand $\varepsilon \rightarrow 0$. Par le théorème de convergence dominée, en en déduit que

$$\mu(\mathbb{1}_F) = \mu(F) = \lim_{\varepsilon \rightarrow 0} \mu(f_{\varepsilon, F}).$$

Ainsi, les intégrales des fonctions lipschitziennes contre μ caractérisent bien les valeurs de μ sur les fermés, et donc caractérisent μ . $\square$

Proposition 2.12. *Soit (E, d) un espace métrique complet et séparable. Toute mesure de Borel finie est de Radon.*

Démonstration. Soit μ une mesure de Borel finie. Soit $\varepsilon > 0$. Par séparabilité de E , on peut, pour tout $n \in \mathbb{N}^*$, recouvrir E par une famille de boules ouvertes $(U_{j,n})_{j \in \mathbb{N}}$ de rayon $\frac{1}{n}$. Par σ -additivité de μ , on peut trouver une suite $m_n \in \mathbb{N}$ telle que

$$\mu \left(E \setminus \bigcup_{j=0}^{m_n} U_{j,n} \right) < \frac{\varepsilon}{2^n}.$$

Ainsi, l'ensemble $K := \bigcap_{n \in \mathbb{N}^*} \bigcup_{j \leq m_n} U_{j,n}$ vérifie

$$\mu(E \setminus K) \leq \sum_{n \in \mathbb{N}^*} \mu \left(E \setminus \bigcup_{j=0}^{m_n} U_{j,n} \right) < \varepsilon.$$

Enfin, on remarque que $\overline{K}$ est précompact¹, et est donc compact puisque E est complet (voir Lemme A.1). Donc $\mu(E \setminus \overline{K}) < \varepsilon$ et μ est de Radon. $\square$

2.3.2 Convergence étroite des mesures

Comme mentionné précédemment, la convergence étroite est l'analogie pour les mesures de la convergence en loi des variables aléatoires. Cette notion n'est pas restreinte aux seules mesures de probabilité. Toutefois, dans le cadre de ce cours, nous ne travaillerons désormais qu'avec des mesures de probabilité.

Définition 2.13. *Soit $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$ et $\mu \in \mathcal{P}(E)$. On dit que $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ , noté $\mu_n \xrightarrow{(e)} \mu$ ou $\mu_n \Rightarrow \mu$, si pour tout $\varphi \in \mathcal{C}_b(E)$, on a*

$$\mu_n(\varphi) \xrightarrow{n \rightarrow \infty} \mu(\varphi).$$

Remarque 2.14 (Petits rappels d'analyse fonctionnelle). *On munit $\mathcal{C}_b(E)$ de la topologie de la convergence uniforme sur les compacts, ce qui en fait un espace de Fréchet. On note $\mathcal{C}_b(E)'$ le dual topologique de $\mathcal{C}_b(E)$, i.e. l'ensemble des applications linéaires continues de $\mathcal{C}_b(E) \rightarrow \mathbb{R}$.*

La topologie faible- sur $\mathcal{C}_b(E)'$ est la topologie la moins fine rendant continues les formes linéaires $L \in \mathcal{C}_b(E)' \mapsto L(\varphi)$, pour tout $\varphi \in \mathcal{C}_b(E)$. Ainsi, une suite $(L_n)_{n \in \mathbb{N}} \subset \mathcal{C}_b(E)'$ converge faiblement-* vers $L \in \mathcal{C}_b(E)'$ si $L_n(\varphi) \rightarrow L(\varphi)$, pour tout $\varphi \in \mathcal{C}_b(E)$.*

Remarquons qu'à toute mesure $\mu \in \mathcal{P}(E)$, on peut associer une forme linéaire de la façon suivante :

$$L_\mu : \varphi \in \mathcal{C}_b(E) \mapsto \mu(\varphi) = \int_E \varphi d\mu,$$

cette application étant continue puisque $|L_\mu(\varphi)| \leq \|\varphi\|_\infty \mu(E) = \|\varphi\|_\infty$.

Ceci justifie l'abus de langage qui consiste parfois à considérer une mesure finie comme élément de $\mathcal{C}_b(E)'$, et à noter $\mu_n \xrightarrow{} \mu$ pour la convergence étroite.*

1. i.e. pour tout $\varepsilon > 0$, on peut recouvrir $\overline{K}$ par un nombre fini de boules de rayon ε .

Remarque 2.15. *Un des intérêts majeurs des topologies faibles est d'avoir plus de compacts, comme par exemple les boules fermées en dimension infinie.*

Cependant, la topologie faible(-) n'est pas métrisable pour les Banach de dimension infinie (cf. cours de Distributions et Opérateurs). On verra pourtant que si E est séparable et complet, la topologie de la convergence étroite est métrisable (pourquoi n'est-ce pas en contradiction avec l'assertion précédente?).*

Exemple 2.16. • Soit $(x_n)_{n \in \mathbb{N}}$ une suite de réels et $x \in \mathbb{R}$. Alors $x_n \rightarrow x$ si et seulement si $\delta_{x_n} \xrightarrow{(e)} \delta_x$.

• La suite de mesures sur $[0, 1]$ définie pour tout $n \in \mathbb{N}^*$ par

$$\mu_n = \frac{1}{n+1} \sum_{k=0}^n \delta_{\frac{k}{n}}$$

converge étroitement vers la mesure de Lebesgue sur $[0, 1]$.

En fait, il n'est pas toujours aisé de tester une mesure contre toutes les fonctions continues et bornées pour obtenir la convergence des mesures.

Le théorème suivant, dont vous avez déjà rencontré des applications, donne plusieurs caractérisations de la convergence étroite.

Théorème 2.17 (Portmanteau²). Soit (E, d) un espace métrique. Soit $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$ et $\mu \in \mathcal{P}(E)$. Les assertions suivantes sont équivalentes :

- (i) $\mu_n \xrightarrow{(e)} \mu$;
- (ii) pour toute fonction φ bornée et uniformément continue, $\mu_n(\varphi) \rightarrow \mu(\varphi)$;
- (iii) pour toute fonction φ bornée et Lipschitz continue, $\mu_n(\varphi) \rightarrow \mu(\varphi)$;
- (iv) pour tout fermé F de E , $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$;
- (v) pour tout ouvert U de E , $\liminf_{n \rightarrow \infty} \mu_n(U) \geq \mu(U)$;
- (vi) pour tout borélien $A \in \mathcal{B}(E)$ tel que $\mu(\partial A) = 0$, $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$.

Démonstration. L'implication (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) et l'équivalence (iv) $\Leftrightarrow$ (v) sont évidentes.

Montrons que (iii) $\Rightarrow$ (iv). Soit F un fermé de E et soit $\varepsilon > 0$. La fonction $f_{\varepsilon, F}$ définie en (2.1) est lipschitzienne et vérifie $\mathbb{1}_F \leq f_{\varepsilon, F} \leq 1$. Ainsi pour tout n , on a $\mu_n(F) \leq \mu_n(f_{\varepsilon, F})$ et par passage à la limite, $\limsup_n \mu_n(F) \leq \mu(f_{\varepsilon, F})$ puisque nous avons supposé (iii). D'autre part, on obtient par convergence dominée que $\mu(f_{\varepsilon, F}) \rightarrow \mu(F)$ lorsque $\varepsilon \rightarrow 0$. Ainsi (iv) est vérifié.

Montrons que (iv)-(v) $\Rightarrow$ (vi). Soit $A \in \mathcal{B}(E)$ tel que $\mu(\partial A) = 0$, et donc tel que $\mu(A) = \mu(\overline{A}) = \mu(A^\circ)$. En appliquant (iv), on obtient

$$\mu(A) = \mu(\overline{A}) \geq \limsup_n \mu_n(\overline{A}) \geq \limsup_n \mu_n(A),$$

et de même en appliquant (v), on obtient

$$\mu(A) = \mu(A^\circ) \leq \liminf_n \mu_n(A^\circ) \leq \liminf_n \mu_n(A).$$

2. Vraisemblablement dû au mathématicien russe Alexandrov, bien que Billingsley l'attribue à Jean-Pierre Portmanteau, de l'Université de Felletin dans la Creuse (~ 1500 habitants).

Ainsi,

$$\limsup_n \mu_n(A) \leq \mu(A) \leq \liminf_n \mu_n(A),$$

et les inégalités précédentes sont donc des égalités.

Montrons finalement que $(vi) \Rightarrow (i)$. Soit $\varphi \in \mathcal{C}_b(E)$. Sans perte de généralité, on peut supposer que $0 \leq \varphi \leq 1$. En effet, si φ n'est pas positive, on peut la décomposer en $\varphi = \varphi_+ - \varphi_-$ et raisonner sur chacun des termes séparément. De plus, on peut toujours se ramener par linéarité à une fonction inférieure à 1. On a alors

$$\mu_n(\varphi) = \int_E \int_{[0,1]} \mathbb{1}_{[0,\varphi(x)]}(y) dy \mu_n(dx).$$

Puisque la mesure $dy \otimes \mu_n(dx)$ est σ -finie (puisque finie) et l'application $(x, y) \mapsto \mathbb{1}_{[0,\varphi(x)]}(y)$ est mesurable et positive, le théorème de Fubini-Tonelli donne :

$$\mu_n(\varphi) = \int_{[0,1]} \mu_n(\{x \in E : \varphi(x) \geq y\}) dy. \quad (2.2)$$

La fonction $\psi : y \mapsto \mu(\{x \in E : \varphi(x) \geq y\})$ est décroissante, donc admet au plus un nombre dénombrable de discontinuité (si vous ne connaissez pas cette propriété, remarquez qu'il s'agit de l'ensemble $\{y : \mu(\{x \in E : \varphi(x) = y\}) > 0\}$ et cet ensemble est l'union des $\{y : \mu(\{x \in E : \varphi(x) = y\}) \geq \frac{1}{n}\}$, $n \in \mathbb{N}^*$ et que $\{y : \mu(\{x \in E : \varphi(x) = y\}) \geq \frac{1}{n}\}$ contient au plus n points).

Soit y un point de continuité de ψ et notons $N_y := \{\varphi \geq y\} = \varphi^{-1}([y, 1])$. Etant l'image réciproque d'un fermé par une fonction continue, N_y est fermé. Si $x \in \partial N_y$, alors il existe une suite (x_n) de $E \setminus N_y$ qui converge vers x . Or $x_n \in E \setminus N_y$ signifie que $\varphi(x_n) < y$, et en passant à la limite, on obtient $\varphi(x) \leq y$, donc $\varphi(x) = y$. Ainsi, $\partial N_y \subset \varphi^{-1}(\{y\})$. Or nous avons supposé que y est un point de continuité de ψ , donc $\mu(\varphi^{-1}(\{y\})) = 0$, et donc $\mu(\partial N_y) = 0$.

La propriété (vi) nous donne alors que $\mu_n(\varphi \geq y) \rightarrow \mu(\varphi \geq y)$. Puisqu'il y a au plus un nombre dénombrable de discontinuité de ψ , on en déduit que

$$\mu_n(\varphi \geq \cdot) \rightarrow \mu(\varphi \geq \cdot) \quad \lambda - p.p.$$

Donc par convergence dominée dans l'intégrale (2.2), on obtient le résultat. $\square$

Exercice 2.18. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles dont les fonctions de répartition sont notées F_{X_n} , et X une v.a. réelle de fonction de répartition F_X . Montrer que $X_n \xrightarrow{(d)} X$ si et seulement si $F_{X_n}(x) \rightarrow F_X(x)$, en tout point de continuité x de F_X .

Exercice 2.19. Soit (E, τ) un espace métrisable.

(i) Soit $(x_n)_{n \in \mathbb{N}} \subset E$ et $x \in E$. On suppose que de toute sous-suite de $(x_n)_{n \in \mathbb{N}}$, on peut extraire une sous-suite qui converge vers x . Montrer que $(x_n)_{n \in \mathbb{N}}$ converge vers x .

(ii) On considère maintenant une suite $(\mu_n)_{n \in \mathbb{N}}$ de mesures de probabilités de Borel sur E . On suppose que de toute sous-suite de $(\mu_n)_{n \in \mathbb{N}}$, on peut extraire une sous-suite qui converge étroitement vers $\mu \in \mathcal{P}(E)$. Montrer que $(\mu_n)_{n \in \mathbb{N}}$ converge vers μ . Commenter le résultat obtenu.

Dans la suite de ce chapitre, l'objectif sera de mettre en évidence des conditions assurant qu'une famille de mesures est relativement compacte, et d'assurer également que la propriété de Bolzano-Weierstrass (qui donne la caractérisation séquentielle de la compacité) est vérifiée sur $\mathcal{P}(E)$. Pour garantir cette dernière propriété, on cherchera à obtenir la métrisabilité de la convergence étroite sur $\mathcal{P}(E)$.

Exercice 2.20. Soit (E, τ) un espace topologique métrisable. Montrer que $\mathcal{C}_b(E)$ est séparable si et seulement si E est compact.

On rappelle que $\mathcal{C}_b(E)$ est séparable si et seulement si la topologie faible-* sur la boule unité de $\mathcal{C}_b(E)'$ est métrisable (cf. [16, Thm 3.16]). Notons τ^* cette topologie. L'exercice précédent indique donc que

$$\tau^* \text{ est métrisable} \iff E \text{ est compact.}$$

On verra pourtant dans la sous-section suivante des conditions beaucoup moins restrictives assurant la métrisabilité de la convergence étroite des mesures de probabilité sur E .

2.3.3 Espaces polonais et propriétés topologiques de $\mathcal{P}(E)$

Espaces polonais

Définition 2.21. On dit qu'un espace topologique (E, τ) est un espace polonais s'il est séparable et s'il existe une distance d sur E qui métrise la topologie τ et pour laquelle (E, d) est complet.

Remarque 2.22. On a souvent accès directement à un espace métrique (i.e. la distance est fournie avec l'espace), mais ce n'est pas toujours le cas.

De plus, il n'est pas toujours intéressant de fixer une distance plus qu'une autre, et on peut même ne pas avoir besoin d'une distance en particulier. Par exemple, il peut nous suffire de tester contre des fonctions pour caractériser telle ou telle convergence, plutôt que de tenter de contrôler une distance qui n'est pas toujours simple à exprimer. En revanche, il est très utile de savoir qu'un espace est métrisable, car cela nous donne accès à plusieurs caractérisations séquentielles : de la continuité des fonctions, de la compacité par la propriété de Bolzano-Weierstrass, etc.

Enfin, une remarque importante qui justifie à nouveau de parler d'espace métrisable plutôt que d'espace métrique : deux distances équivalentes (i.e. qui engendrent la même topologie) peuvent donner ou non la complétude de l'espace. Par exemple $\mathbb{R}$ muni de sa topologie usuel est évidemment polonais. Or on peut munir $\mathbb{R}$ de la distance $d(x, y) = |\arctan(x) - \arctan(y)|$ qui est équivalente à la distance usuelle. Pourtant $(\mathbb{R}, d)$ n'est pas complet. Néanmoins l'espace métrique $(\mathbb{R}, d)$ est polonais !

Distance sur l'espace des mesures

Un certain nombre de distances sont connues pour métriser la convergence étroite. La distance de Lévy-Prokhorov en fait partie et on verra qu'elle a en plus l'avantage de rendre $\mathcal{P}(E)$ complet.

Lemme 2.23. Soit (E, d) un espace métrique. On rappelle la notation $A^\varepsilon = \{x \in E : d(x, A) < \varepsilon\}$ pour toute partie A de E et tout $\varepsilon > 0$ (notez que A^ε est un ouvert, donc mesurable). L'application $d_{LP} : \mathcal{P}(E) \times \mathcal{P}(E) \rightarrow \mathbb{R}_+$ définie par :

$$d_{LP}(\mu, \nu) := \inf \{ \varepsilon > 0 : \forall A \in \mathcal{B}(E), \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ et } \nu(A) \leq \mu(A^\varepsilon) + \varepsilon \},$$

est une distance sur $\mathcal{P}(E)$, appelée distance de Lévy-Prokhorov.

Démonstration. L'application $d_{LP} : \mathcal{P}(E) \times \mathcal{P}(E) \rightarrow \mathbb{R}_+$ est symétrique et l'inégalité triangulaire s'obtient grâce à la relation $(A^\varepsilon)^\eta \subset A^{\varepsilon+\eta}$. Il reste à montrer que $d_{LP}(\mu, \nu) = 0$ si et seulement si $\mu = \nu$. Grâce à la Proposition 2.11, il suffit de vérifier l'égalité sur les fermés de E . Or si A est un fermé de E , on a $A = \bigcap_{\varepsilon>0} A^\varepsilon$. Comme $\mu(A) \leq \nu(A^\varepsilon) + \varepsilon$ pour tout $\varepsilon > 0$ (puisque $d_{LP}(\mu, \nu) = 0$), on obtient par passage à la limite que $\mu(A) \leq \nu(A)$. De manière symétrique, on obtient que $\mu(A) = \nu(A)$. On a donc bien une distance.

Au passage, remarquons que d_{LP} est définie plus simplement par :

$$d_{LP}(\mu, \nu) := \inf \{ \varepsilon > 0 : \forall A \in \mathcal{B}(E), \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \}. \quad (2.3)$$

En effet, soit $\varepsilon > 0$ tel que $\mu(A) \leq \nu(A^\varepsilon) + \varepsilon$. Commençons par remarquer que pour toute partie A de E et tout $\varepsilon > 0$, on a

$$((A^\varepsilon)^c)^\varepsilon \subset A^c.$$

Par conséquent,

$$\begin{aligned} \nu(A) &= 1 - \nu(A^c) \leq 1 - \nu(((A^\varepsilon)^c)^\varepsilon) \\ &\leq 1 - \mu((A^\varepsilon)^c) + \varepsilon = \mu(A^\varepsilon) + \varepsilon, \end{aligned}$$

ce qui montre bien l'égalité entre les deux définitions. $\square$

Premières propriétés topologiques

Proposition 2.24. *Soit (E, τ) un espace métrisable et séparable. Alors $\mathcal{P}(E)$ muni de la topologie de la convergence étroite est métrisable.*

Démonstration. Soit d une distance sur E qui métrise τ . Montrons que d_{LP} (construite à partir de d) métrise la convergence étroite sur $\mathcal{P}(E)$. On prend pour cela une suite $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$ telle que $d_{LP}(\mu_n, \mu) \rightarrow 0$. Soit A un fermé et $\varepsilon > 0$. Ainsi il existe $N \in \mathbb{N}$ tel que

$$\forall n \geq N, \quad \mu_n(A) \leq \mu(A^\varepsilon) + \varepsilon.$$

On a donc $\limsup_n \mu_n(A) \leq \mu(A^\varepsilon) + \varepsilon$. Puisque cette inégalité est valable $\forall \varepsilon > 0$, et que A^ε décroît vers $\bigcap_{\varepsilon>0} A^\varepsilon = A$ (car A fermé), la σ -additivité de μ implique que $\mu(A^\varepsilon) \rightarrow \mu(A)$, et donc

$$\limsup_n \mu_n(A) \leq \mu(A).$$

D'après le théorème Portmanteau, on a donc que $\mu_n \xrightarrow{(e)} \mu$.

Réciproquement, supposons qu'on dispose d'une suite $(\mu_n)_{n \in \mathbb{N}}$ telle que $\mu_n \xrightarrow{(e)} \mu$, et montrons que $d_{LP}(\mu_n, \mu) \rightarrow 0$. Soit $\varepsilon > 0$. Puisque E est séparable, il existe une suite de boules ouvertes de rayon $\varepsilon > 0$ recouvrant E . Par σ -additivité de μ , on peut trouver un nombre fini de telles boules $A_1, \dots, A_N$ telles que

$$\mu \left(E \setminus \bigcup_{k=1}^N A_k \right) \leq \varepsilon. \quad (2.4)$$

Soit $A \in \mathcal{B}(E)$. On note

$$U_A := \bigcup_{k \leq N : A \cap A_k \neq \emptyset} A_k$$

et on recouvre alors A comme suit :

$$A \subset U_A \bigcup (A \setminus U_A).$$

Or U_A est ouvert donc par le théorème Portmanteau, $\liminf_n \mu_n(U_A) \geq \mu(U_A)$. De plus on remarque que le cardinal de l'ensemble $\{U_A : A \in \mathcal{B}(E)\}$ est fini, ainsi il existe N_0 tel que pour tout $n \geq N_0$ et pour tout $A \in \mathcal{B}(E)$, $\mu_n(U_A) \geq \mu(U_A) - \varepsilon$, de sorte que

$$\forall A \in \mathcal{B}(E), \quad \mu(A) \leq \mu_n(U_A) + \varepsilon + \mu(A \setminus U_A).$$

De plus, $U_A \subset A^{2\varepsilon}$ et $A \setminus U_A \subset E \setminus \bigcup_{k=1}^N A_k$, donc

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

Par la caractérisation non symétrique (2.3) de la distance de Lévy-Prokhorov, on en déduit que pour tout $n \geq N_0$, $d_{LP}(\mu_n, \mu) \leq 2\varepsilon$, ce qui donne la convergence au sens de la distance de Lévy-Prokhorov. $\square$

Toujours à l'aide de la distance de Lévy-Prokhorov, nous allons montrer qu'il existe une famille dénombrable de mesures discrètes (i.e. des sommes finies de Dirac) telle que toute mesure de $\mathcal{P}$ peut être approchée par une mesure de cette famille.

Proposition 2.25. *Soit (E, τ) un espace métrisable et séparable. Alors $\mathcal{P}(E)$ muni de la topologie de la convergence étroite est séparable.*

Démonstration. Soit $(x_n)_{n \in \mathbb{N}}$ une suite dense de E . Considérons le sous-ensemble de $\mathcal{P}(E)$ donné par

$$\left\{ \sum_{k=1}^N r_k \delta_{x_k}; N \in \mathbb{N}, r_1, \dots, r_N \in \mathbb{Q}_+ \text{ tels que } \sum_{k=1}^N r_k = 1 \right\},$$

qui est clairement dénombrable. Montrons qu'il est dense dans $\mathcal{P}(E)$.

Soit $\varepsilon > 0$ et $\mu \in \mathcal{P}(E)$. Comme dans la preuve précédente, on considère N boules ouvertes $A_1, \dots, A_N$ de rayon $\varepsilon > 0$ et centrées en $x_1, \dots, x_N$ telles que (2.4) soit vérifié. A partir de ces N boules, on construit N boréliens disjoints $A'_1, \dots, A'_N$ (en prenant des intersections des boules) tels que $x_k \in A'_k$ pour tout $k \in \{1, \dots, N\}$ et $\bigcup_{k=1}^N A'_k = \bigcup_{k=1}^N A_k$. On choisit alors des rationnels $r_1, \dots, r_N$ tels que

$$\sum_{k=1}^N |r_k - \mu(A'_k)| < \varepsilon,$$

et on utilisera la mesure suivante pour approcher μ :

$$\nu := \sum_{k=1}^N r_k \delta_{x_k}.$$

Alors, pour $A \in \mathcal{B}(E)$, on définit $U_A = \bigcup_{k \leq N: A \cap A'_k \neq \emptyset} A'_k$ (notez que l'union est disjointe cette fois-ci, et que $U_A \in \mathcal{B}(E)$) et on obtient

$$\mu(A) \leq \mu(U_A) + \mu(A \setminus U_A) \leq \mu(U_A) + \varepsilon.$$

Or

$$|\mu(U_A) - \nu(U_A)| \leq \sum_{k \leq N: A \cap A'_k \neq \emptyset} |\mu(A'_k) - r_k| < \varepsilon.$$

Donc

$$\mu(A) \leq \nu(U_A) + 2\varepsilon.$$

Enfin, il reste à remarquer que $U_A \subset A^{2\varepsilon}$ pour conclure que $d_{LP}(\mu, \nu) \leq 2\varepsilon$. $\square$

On verra plus tard, une fois que nous aurons établi le théorème de Prokhorov, que d_{LP} rend $\mathcal{P}(E)$ complet, en faisant donc un espace polonais.

2.4 Compacité dans $\mathcal{P}(E)$

D'abord, observons une première situation où $\mathcal{P}(E)$ est compact :

Proposition 2.26. *Soit (E, τ) un espace topologique métrisable. L'ensemble $\mathcal{P}(E)$ muni de la topologie de la convergence étroite est compact si et seulement si E est compact. Dans ce cas, $\mathcal{P}(E)$ est également métrisable.*

D'un point de vue pratique, ce résultat nous dit que si E est compact, on peut extraire de toute suite de mesures de probabilité sur E une sous-suite qui converge.

Nous ne ferons pas la preuve dans ce cours (voir les références citées en début de chapitre). On notera néanmoins que le sens réciproque repose sur un intéressant théorème de Riesz, qui stipule que lorsque (E, d) est compact, les formes linéaires positives sur $\mathcal{C}(E)$ qui valent 1 pour la fonction constante égale à 1, sont précisément de la forme $f \mapsto \mu(f)$, pour $\mu \in \mathcal{P}(E)$.

Intéressons-nous désormais à un critère plus général, qui ne nécessite pas de supposer que E est compact. Nous allons voir que les parties relativement compactes de $\mathcal{P}(E)$ seront données par les familles tendues de mesures :

Définition 2.27. *Une famille $\mathcal{M} \subset \mathcal{P}(E)$ de mesures est tendue si pour tout $\varepsilon > 0$, il existe un compact K_ε de E tel que*

$$\sup_{\mu \in \mathcal{M}} \mu(E \setminus K_\varepsilon) \leq \varepsilon.$$

Exemple 2.28. $\triangleright$ *Si E est un espace métrique complet et séparable, et $\mathcal{M}$ est réduit à une seule mesure $\mu \in \mathcal{P}(E)$, alors puisque μ est de Radon (Proposition 2.12), $\mathcal{M}$ est tendue.*

$\triangleright$ *Si $E = \mathbb{R}$, alors la famille $\mathcal{M} = \{\delta_n, n \in \mathbb{Z}\}$ n'est pas tendue.*

Venons-en désormais à un résultat important de ce cours.

Théorème 2.29 (Prokhorov). *Soit (E, τ) un espace polonais et $\mathcal{P}(E)$ muni de la topologie de la convergence étroite. Soit $\mathcal{M} \subset \mathcal{P}(E)$ une famille de mesures. Alors $\mathcal{M}$ est relativement compacte si et seulement si $\mathcal{M}$ est tendue.*

Remarque 2.30. $\triangleright$ *Une démonstration de ce théorème (une version légèrement plus générale) est disponible sous forme de problème sur l'espace en ligne du cours. La preuve repose sur une approche analyse fonctionnelle, et est accessible à partir des connaissances du cours de 2A "Distributions et Opérateurs".*

- ▷ On comprend désormais mieux la fin de la démonstration du théorème de Lévy (cf Exercice 2.9).
- ▷ En pratique, on utilisera plus souvent le sens réciproque.

Démonstration. Commençons par le sens direct (on remarquera que la deuxième partie de l'argument est très similaire à la preuve de la Proposition 2.12). Soit $\mathcal{M}$ une famille relativement compacte. Quitte à considérer $\overline{\mathcal{M}}$ à la place de $\mathcal{M}$, on peut supposer $\mathcal{M}$ compacte. En effet, si $\overline{\mathcal{M}}$ est tendue, alors $\mathcal{M}$ l'est également.

Soit $(U_n)_{n \in \mathbb{N}}$ une suite croissante d'ouverts tels que $\bigcup_{n \in \mathbb{N}} U_n = E$. Dans un premier temps, montrons que pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que $\inf_{\mu \in \mathcal{M}} \mu(U_N) \geq 1 - \varepsilon$. Par l'absurde, si ce n'était pas le cas, il existerait un $\varepsilon > 0$ et une suite $(\mu_n)_{n \in \mathbb{N}}$ de $\mathcal{M}$ tels que pour tout $n \in \mathbb{N}$, $\mu_n(U_n) < 1 - \varepsilon$. Or par hypothèse, E est polonais, donc $\mathcal{P}(E)$ est métrisable par la Proposition 2.24. De plus $\mathcal{M}$ est compact, donc il existe $\mu \in \mathcal{P}(E)$ et une sous-suite de (μ_n) telle que $\mu_{\varphi(n)} \xrightarrow{(e)} \mu$. D'après le théorème Portmanteau, on a donc pour tout $n \in \mathbb{N}$,

$$\liminf_{m \rightarrow \infty} \mu_{\varphi(m)}(U_{\varphi(m)}) \geq \liminf_{m \rightarrow \infty} \mu_{\varphi(m)}(U_{\varphi(n)}) \geq \mu(U_{\varphi(n)}),$$

où la première inégalité résulte de la croissance des U_n . Donc $\mu(U_{\varphi(n)}) \leq 1 - \varepsilon$ pour tout $n \in \mathbb{N}$, ce qui est absurde puisque $\mu(E) = 1$.

Soit d une distance rendant E complet. Soit $\varepsilon > 0$ et $(x_n)_{n \in \mathbb{N}}$ une suite dense de E . D'après le paragraphe précédent, on a que pour tout $n \geq 1$, il existe un entier N_n tel que

$$\inf_{\mu \in \mathcal{M}} \mu \left(\bigcup_{k=1}^{N_n} B_d(x_k, \frac{1}{n}) \right) \geq 1 - \frac{\varepsilon}{2^n}.$$

Alors en posant

$$K = \bigcap_{n \geq 1} \bigcup_{k=1}^{N_n} \overline{B}_d(x_k, \frac{1}{n}),$$

on a que K est fermé et précompact, donc compact (puisque (E, d) est complet). De plus

$$\sup_{\mu \in \mathcal{M}} \mu(E \setminus K) \leq \sum_{n=1}^{\infty} \mu \left(E \setminus \bigcup_{k=1}^{N_n} \overline{B}_d(x_k, \frac{1}{n}) \right) \leq \varepsilon.$$

Donc $\mathcal{M}$ est tendu.

Réciproquement, supposons que $\mathcal{M}$ est tendue. Puisque $\mathcal{P}(E)$ est métrisable (Proposition 2.24), la précompacité est donnée par la caractérisation séquentielle suivante : il suffit de montrer que de toute suite de $\mathcal{M}$, on peut extraire une sous-suite qui converge (avec une limite pas nécessairement dans $\mathcal{M}$).

Le but ici va être de se ramener au cadre compact de la Proposition 2.26. On utilise un lemme de topologie (voir Lemme A.2 en Annexe) : il existe une distance $\tilde{d}$ telle que $(E, \tilde{d})$ est précompact. Soit $\tilde{E}$ le complété de E pour la distance $\tilde{d}$ (voir Annexe). Alors $(\tilde{E}, \tilde{d})$ est compact, et nous allons désormais travailler sur cet espace. Ainsi, toute mesure $\mu \in \mathcal{P}(E)$ s'étend en une mesure $\tilde{\mu}$ sur $(\tilde{E}, \mathcal{B}(\tilde{E}))$ de la façon suivante :

$$\forall A \in \mathcal{B}(\tilde{E}), \quad \tilde{\mu}(A) = \mu(A \cap E).$$

Vous pourrez vérifier que $\mathcal{B}(E) = \{A \cap E : A \in \mathcal{B}(\tilde{E})\}$ et que $\tilde{\mu}$ ainsi définie est bien une mesure de probabilité.

Soit $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{M}$. La suite associée $(\tilde{\mu}_n)_{n \in \mathbb{N}}$ est un sous-ensemble de $\mathcal{P}(\tilde{E})$ qui est compact d'après la Proposition 2.26. Donc il existe $\tilde{\mu}_\infty \in \mathcal{P}(\tilde{E})$ et une extraction φ telle que $\tilde{\mu}_{\varphi(n)} \xrightarrow{(e)} \tilde{\mu}_\infty$.

Notez qu'on n'a pas utilisé la tension jusqu'ici, cette propriété va nous servir à montrer que le support de $\tilde{\mu}_\infty$ est dans E , et on pourra alors construire la restriction de $\tilde{\mu}_\infty$ à une mesure $\mu_\infty \in \mathcal{P}(E)$.

Grâce à la tension, on a que pour tout $\varepsilon > 0$, il existe $K_\varepsilon \subset E$ compact tel que $\inf_n \mu_n(K_\varepsilon) \geq 1 - \varepsilon$. Comme K_ε est fermé, la propriété (iv) du Théorème Portmanteau nous donne que

$$\tilde{\mu}_\infty(K_\varepsilon) \geq \limsup_n \tilde{\mu}_n(K_\varepsilon) \geq 1 - \varepsilon.$$

Ainsi l'ensemble $K := \bigcup_{n \in \mathbb{N}^*} K_{1/n}$ est un borélien de E (pas nécessairement compact) et vérifie

$$\tilde{\mu}_\infty(K) = \lim_{N \rightarrow \infty} \tilde{\mu}_\infty \left(\bigcup_{n=1}^N K_{1/n} \right) = 1.$$

Donc $\tilde{\mu}_\infty$ peut être considérée comme une mesure sur E puisque pour tout $A \in \mathcal{B}(\tilde{E})$, on a

$$\tilde{\mu}_\infty(A) = \tilde{\mu}_\infty(A \cap K) = \tilde{\mu}_\infty(A \cap E)$$

et $A \cap E \in \mathcal{B}(E)$. On notera $\mu_\infty \in \mathcal{P}(E)$. Enfin, on a $\mu_n \xrightarrow{(e)} \mu_\infty$. En effet, soit f une fonction uniformément continue et bornée sur $(E, \tilde{d})$. Cette fonction se prolonge de manière unique (voir Lemme A.3) en une fonction uniformément continue sur $(\tilde{E}, \tilde{d})$. On note $\tilde{f}$ ce prolongement. Alors, puisque le support de $\tilde{\mu}_n$ et $\tilde{\mu}_\infty$ est dans E ,

$$\mu_{\varphi(n)}(f) = \tilde{\mu}_{\varphi(n)}(\tilde{f}) \rightarrow \tilde{\mu}_\infty(\tilde{f}) = \mu(f),$$

ce qui fournit la convergence recherchée, et la relative compacité de $\mathcal{M}$. $\square$

Remarque 2.31. *Nous n'avons pas utilisé la complétude dans le sens réciproque. Ce sens reste d'ailleurs vrai dans un cadre beaucoup plus général (voir [3]), contrairement au sens direct.*

En conséquence du théorème de Prokhorov, on obtient la complétude de $(\mathcal{P}(E), d_{LP})$, ce qui est toujours une information particulièrement utile.

Théorème 2.32. *Soit (E, τ) un espace polonais. Alors $\mathcal{P}(E)$ muni de la topologie de la convergence étroite est polonais.*

Démonstration. Grâce aux résultats des Propositions 2.24 et 2.25, il ne reste plus qu'à démontrer que la distance d_{LP} rend $\mathcal{P}(E)$ complet.

Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de Cauchy pour la distance d_{LP} . La preuve repose sur le principe suivant : il suffit d'avoir la convergence d'une sous-suite pour que $(\mu_n)_{n \in \mathbb{N}}$ converge, et par le théorème de Prokhorov, il suffit de montrer que $(\mu_n)_{n \in \mathbb{N}}$ est tendue pour avoir la convergence d'une sous-suite. Montrons donc la tension de $(\mu_n)_{n \in \mathbb{N}}$.

Soit $\varepsilon > 0$. Par la propriété de Cauchy, on a que pour tout $n \in \mathbb{N}$, il existe $\varphi(n) \in \mathbb{N}$ tel que pour tout $p \geq \varphi(n)$,

$$d_{LP}(\mu_p, \mu_{\varphi(n)}) \leq \varepsilon_{n+1}, \quad (2.5)$$

où on note, pour tout $n \in \mathbb{N}$, $\varepsilon_n := \frac{\varepsilon}{2^n}$. Remarquons que quelque soit n , l'ensemble $\{\mu_1, \dots, \mu_{\varphi(n)}\}$ est (relativement) compact, donc tendu d'après Prokhorov. Ainsi, il existe $K_n \subset E$ un compact tel que $\max_{1 \leq k \leq \varphi(n)} \mu_k(E \setminus K_n) \leq \varepsilon_{n+1}$. Par (2.5) et ce qui précède, il s'ensuit que pour tout $p \geq \varphi(n)$,

$$\mu_p(K_n^{\varepsilon_n}) \geq \mu_{\varphi(n)}(K_n) - \varepsilon_{n+1} \geq 1 - \varepsilon_n$$

et pour tout $p \in \{1, \dots, \varphi(n)\}$, la même propriété est vérifiée par construction.

On pose alors $K = \bigcap_{n \in \mathbb{N}} K_n^{\varepsilon_{n+1}}$. On remarque que K_n est précompact, donc il existe N et des points $x_1, \dots, x_N$ de E tels que $K_n \subset \bigcup_{k=1}^N B(x_k, \varepsilon_{n+1})$. Par conséquent, $K_n^{\varepsilon_{n+1}} \subset \bigcup_{k=1}^N B(x_k, \varepsilon_n)$ et donc K est précompact. Ainsi K est fermé et précompact dans un complet, donc compact. On a alors que

$$\mu_p(E \setminus K) \leq \sum_{n \in \mathbb{N}} \mu_p(E \setminus K_n^{\varepsilon_{n+1}}) \leq \varepsilon.$$

Donc $(\mu_n)_{n \in \mathbb{N}}$ est bien tendue. □

Théorème 2.33 (Représentation de Skorokhod). *Soit $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires à valeurs dans un espace polonais (E, τ) , telles que $X_n \xrightarrow{(e)} X$.*

Alors il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et des variables aléatoires $(\tilde{X}_n)_{n \in \mathbb{N}}$ et $\tilde{X}$ définies de cet espace vers (E, τ) telles que

- ▷ $\tilde{X}_n \stackrel{(d)}{=} X_n, \forall n \in \mathbb{N};$
- ▷ $\tilde{X}_n \rightarrow \tilde{X} \text{ p.s.}$

Nous ne démontrerons pas ce théorème, la preuve étant assez longue et technique (mais néanmoins faisable). Cependant, ce résultat est souvent très utile, ce qui justifiait de l'énoncer. En voici un exemple d'application.

Exercice 2.34. *Montrer qu'une suite de variables aléatoires à valeurs réelles $(X_n)_{n \in \mathbb{N}}$ qui converge en loi et qui est uniformément intégrable vérifie $\mathbb{E}X_n \rightarrow \mathbb{E}X$.*

[On rappelle la définition de l'uniforme intégrabilité : $\lim_{A \rightarrow +\infty} \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geq A}] = 0.$]

Chapitre 3

Convergence des processus continus (et un peu càdlàg)

Dans ce chapitre, nous allons nous intéresser à des processus stochastiques, c'est-à-dire des variables aléatoires à valeurs dans un espace de fonction. On traitera essentiellement de processus continus, avec dans un premier temps des processus d'un espace métrique compact (K, d_K) vers un espace polonais E . Cet ensemble $\mathcal{C}(K, E)$ sera noté de façon plus concise $\mathcal{C}$, et sera muni de la distance uniforme. Dans un second temps, on traitera plus généralement de fonctions $\mathcal{C}(\mathbb{R}_+, E)$ définies sur tout $\mathbb{R}_+$. Enfin, on évoquera le cas de processus discontinus, mais ayant en tout point une limite à gauche et étant continus à droite (Section 3.4).

3.1 Convergence étroite dans l'espace des fonctions continues sur un compact

3.1.1 L'espace $\mathcal{C}$

On munit $\mathcal{C}$ de la topologie de la convergence uniforme : pour une distance d métrisant la topologie de E et le rendant complet,

$$d_\infty(f, g) = \sup_{t \in K} d(f(t), g(t)), \quad \forall f, g \in \mathcal{C}.$$

Il est clair que la définition de d_∞ dépend du choix de d , mais on peut vérifier que la topologie sur $\mathcal{C}$ n'en dépend pas.

Introduisons les notions de *module de continuité* d'une fonction et d'*équicontinuité* d'une famille de fonctions, qui nous seront utiles dans la preuve du théorème suivant et dans la suite de ce chapitre.

Définition 3.1. Soit (X, d_X) et (Y, d_Y) des espaces métriques. On définit :

▷ le module de continuité de $f \in \mathcal{C}(X, Y)$ par

$$\omega(f, \eta) = \sup \{d_Y(f(x), f(y)) : x, y \in X, d_X(x, y) \leq \eta\}, \quad \forall \eta > 0;$$

▷ l'équicontinuité d'une famille $\mathcal{A} \subset \mathcal{C}(X, Y)$ par la propriété

$$\forall \varepsilon > 0, \exists \eta > 0 \text{ tel que } \sup_{f \in \mathcal{A}} \omega(f, \eta) \leq \varepsilon.$$

Théorème 3.2. *Soit E un espace polonais. Alors l'espace $\mathcal{C} = \mathcal{C}(K, E)$ est polonais.*

Démonstration. La complétude est un résultat classique qui découle de la complétude de (E, d) . Montrons la séparabilité. Pour tous $m, n \in \mathbb{N}^*$, posons

$$C_{m,n} = \left\{ f \in \mathcal{C} : \omega(f, \frac{1}{m}) < \frac{1}{n} \right\}.$$

Par compacité de K , il existe pour tout $m \in \mathbb{N}^*$ un ensemble fini de points $K_m = \{x_1, \dots, x_{N_m}\}$ tels que

$$K \subset \bigcup_{j=1}^{N_m} B(x_j, \frac{1}{m}).$$

De plus, il existe un ensemble dénombrable $D_{n,m} \subset C_{n,m}$ tel que pour tout $f \in C_{n,m}$, pour tout $\varepsilon > 0$, il existe $g \in D_{n,m}$ tel que pour tout $x \in K_m$, $d(f(x), g(x)) < \varepsilon$. En effet, la séparabilité de E assure que pour tout $p \in \mathbb{N}^*$, on peut recouvrir E par un nombre dénombrable de boules de rayon $\frac{1}{p}$, notées $(B_l^p)_{l \in \mathbb{N}}$. Pour tout N_m -uplet $(l_1, \dots, l_{N_m})$ d'entiers strictement positifs, on considère le sous-ensemble de $C_{n,m}$ suivant :

$$C_{n,m,p,l_1,\dots,l_{N_m}} := \left\{ f \in C_{n,m} : f(x_j) \in B_{l_j}^p, \forall j = 1, \dots, N_m \right\}.$$

Si $C_{n,m,p,l_1,\dots,l_{N_m}} \neq \emptyset$, on choisit un unique élément de cet ensemble qu'on note $c_{n,m,p,l_1,\dots,l_{N_m}}$ et on définit

$$D_{n,m} = \bigcup_{p \in \mathbb{N}^*} \bigcup_{(l_1,\dots,l_{N_m}) \in (\mathbb{N}^*)^{N_m}} \{c_{n,m,p,l_1,\dots,l_{N_m}}\}.$$

Alors nous allons montrer que l'ensemble dénombrable suivant

$$D := \bigcup_{n,m \in \mathbb{N}^*} D_{n,m}$$

est dense dans $\mathcal{C}$. Pour cela, soit $\varepsilon > 0$ et $f \in \mathcal{C}$. Soit un entier $n > \frac{3}{\varepsilon}$. Puisque f est uniformément continue (par le théorème de Heine), il existe $m \in \mathbb{N}^*$ tel que $f \in C_{n,m}$. Soit $g \in D_{n,m}$ tel que $d(f(x), g(x)) < \frac{1}{n}$ pour tout $x \in K_m$. Pour x quelconque dans K , soit $y \in K_m$ tel que $d_K(x, y) < \frac{1}{m}$. Alors

$$\begin{aligned} d(f(x), g(x)) &\leq d(f(x), f(y)) + d(f(y), g(y)) + d(g(y), g(x)) \\ &< \frac{3}{n} < \varepsilon. \end{aligned}$$

□

Rappelons, sans le démontrer¹, un résultat classique d'analyse fonctionnelle caractérisant les sous-ensembles relativement compacts de $\mathcal{C}$.

Théorème 3.3 (Arzelà-Ascoli). *Soit K compact et E métrique. On se donne une famille de fonctions $\mathcal{A} \subset \mathcal{C}(K, E)$. Pour K_0 un sous-ensemble dénombrable dense de K , on a l'équivalence suivante :*

▷ $\mathcal{A}$ est relativement compacte ;

1. cf. cours de *Distributions et Opérateurs* en 2A, ou [14, Théorème 2.1].

- ▷ $\mathcal{A}$ est équicontinue et pour tout $x \in K_0$, $\{f(x), f \in \mathcal{A}\}$ est relativement compact dans E .

Lorsque les boules fermées de E sont compactes, on peut simplifier le deuxième point de l'assertion précédente.

Corollaire 3.4. *Soit K compact connexe et E métrique dont les boules fermées sont compactes. On considère une famille $\mathcal{A} \subset \mathcal{C}(K, E)$. On a l'équivalence suivante :*

- ▷ $\mathcal{A}$ est relativement compacte ;
 ▷ $\mathcal{A}$ est équicontinue et il existe un $x_0 \in K$ tel que $\{f(x_0), f \in \mathcal{A}\}$ est relativement compact dans E .

Démonstration. Supposons que $\mathcal{A}$ est équicontinue et qu'il existe un $x_0 \in K$ tel que $\{f(x_0), f \in \mathcal{A}\}$ est relativement compact dans E . Montrons que pour tout $x \in K$, $\{f(x), f \in \mathcal{A}\}$ est relativement compact dans E .

Par équicontinuité, il existe $\eta > 0$ tel que $\omega(f, \eta) \leq 1$ pour tout $f \in \mathcal{A}$. Par compacité de K , il existe un nombre fini N_η de points $x_1, \dots, x_{N_\eta}$ tels que $K \subset \bigcup_{j=1}^{N_\eta} B(x_j, \eta/2)$. Ainsi, pour tout $x \in K$, il existe $N \leq N_\eta$ points $\{y_1, \dots, y_N\} \subset \{x_1, \dots, x_{N_\eta}\}$ tels que $d(x_0, y_1) < \eta$, $d(y_j, y_{j+1}) < \eta$ pour $1 \leq j \leq N-1$ et $d(y_N, x) < \eta$. On a donc, pour tout $f \in \mathcal{A}$,

$$\begin{aligned} d(f(x_0), f(x)) &\leq d(f(x_0), f(y_1)) + \sum_{j=1}^{N-1} d(f(y_j), f(y_{j+1})) + d(f(y_N), f(x)) \\ &\leq N + 1 \leq N_\eta + 1. \end{aligned}$$

Puisque $\{f(x_0), f \in \mathcal{A}\}$ est borné, on en déduit que $\{f(x), f \in \mathcal{A}\}$ est borné aussi, et donc relativement compact. $\square$

3.1.2 Cylindres et lois finies-dimensionnelles

Etant donné un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathcal{T}$ un ensemble quelconque, on appelle *processus stochastique* une famille de variables aléatoires $(X_t)_{t \in \mathcal{T}}$ définies sur cet espace et à valeurs dans E . On va voir que de manière équivalente, un processus stochastique est une application mesurable de $(\Omega, \mathcal{F})$ vers l'ensemble de fonctions $E^\mathcal{T}$, muni d'une tribu adéquate.

Définition 3.5. *Soit $\mathcal{T}$ un ensemble quelconque et E un espace polonais. On appelle tribu produit et on note $E^{\otimes \mathcal{T}}$ la plus petite tribu sur $E^\mathcal{T}$ qui rend mesurables les applications coordonnées :*

$$\pi_t^{(E)} : \begin{array}{ccc} E^\mathcal{T} & \rightarrow & E \\ f & \mapsto & f(t) \end{array}, \quad t \in \mathcal{T}.$$

En conséquence, les *cylindres*, c'est-à-dire les ensembles de la forme

$$\{f \in E^\mathcal{T} : f(t_1) \in A_1, \dots, f(t_n) \in A_n\} = (\pi_{t_1}^{(E)})^{-1}(A_1) \cap \dots \cap (\pi_{t_n}^{(E)})^{-1}(A_n),$$

sont mesurables pour tous $t_1, \dots, t_n \in \mathcal{T}$, $A_1, \dots, A_n \in \mathcal{B}(E)$, et $n \in \mathbb{N}^*$.

On peut aussi exprimer simplement la mesurabilité d'une application de $(\Omega, \mathcal{F}) \rightarrow (E^\mathcal{T}, E^{\otimes \mathcal{T}})$: ainsi, on dira que X est une v.a. à valeurs dans $E^\mathcal{T}$ si c'est une application mesurable, ce qui est équivalent vue la définition précédente à supposer que pour tout $t \in \mathcal{T}$, $X(t)$ est une variable aléatoire de $(\Omega, \mathcal{F})$ vers $(E, \mathcal{B}(E))$.

Définition 3.6. Soit $\mu \in \mathcal{P}(E^{\mathcal{T}})$. Pour tout $n \in \mathbb{N}^*$ et tout $(t_1, \dots, t_n) \in \mathcal{T}^n$, on note $\mu_{t_1, \dots, t_n}$ la mesure image de μ par la projection sur les coordonnées

$$\pi_{t_1, \dots, t_n}^{(E)} : \begin{array}{ccc} E^{\mathcal{T}} & \rightarrow & E^n \\ f & \mapsto & (f(t_1), \dots, f(t_n)) \end{array} .$$

Il s'agit donc d'une mesure sur E^n , appelée marginale de μ . L'ensemble des marginales

$$\{\mu_{t_1, \dots, t_n} : (t_1, \dots, t_n) \in \mathcal{T}^n, n \in \mathbb{N}^*\}$$

est appelée ensemble des marginales finies-dimensionnelles de μ .

Dans le vocabulaire des probabilités, si X est une v.a. à valeurs dans $E^{\mathcal{T}}$ de loi μ , $\mu_{t_1, \dots, t_n}$ est la loi de $(X(t_1), \dots, X(t_n))$ et on appelle loi finie-dimensionnelle une telle marginale.

Proposition 3.7. Toute mesure $\mu \in \mathcal{P}(E^{\mathcal{T}})$ est caractérisée par ses marginales finies-dimensionnelles.

Démonstration. Supposons que μ et ν ont les mêmes marginales finies-dimensionnelles. Alors elles sont égales sur tous les cylindres. Or ces ensembles forment un π -système (ensemble stable par intersections finies) et engendrent la tribu produit. Comme μ et ν sont des mesures finies, un corollaire du lemme de classe monotone [11, Corollaire 1.4.2] indique que $\mu = \nu$. $\square$

On pourrait continuer ainsi à travailler sur $(E^{\mathcal{T}}, E^{\otimes \mathcal{T}})$, construire des processus grâce au Théorème d'extension de Kolmogorov² (qui nous donnerait l'existence de processus de Markov ou du mouvement brownien dans $\mathbb{R}^{\mathbb{R}^+}$), mais remarquons que $E^{\mathcal{T}}$ n'est pas polonais en général. Ce cadre n'est donc pas adapté à l'étude de la convergence étroite, ce qui justifie de considérer un espace plus petit que $E^{\mathcal{T}}$.

On pourrait aussi légitimement se demander si les trajectoires $t \mapsto X_t(\omega)$ sont continues pour presque tout $\omega \in \Omega$.

Remarque 3.8. L'ensemble $\mathcal{C}(K, E)$ n'appartient pas en général à la tribu produit $E^{\otimes K}$. Ce résultat est laissé en **exercice** (on pourra considérer le processus défini sur $[0, 1]$ par $X(t) = \mathbb{1}_{U=t}$, $U \sim \text{Unif}([0, 1])$, calculer ses lois finies-dimensionnelles et chercher une contradiction).

Rappelons que $\mathcal{C}$ est muni de la topologie de la convergence uniforme (possible ici car K est supposé compact), et notons $\mathcal{C}$ la tribu borélienne associée. Un *processus stochastique continu* est une variable aléatoire d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ vers $(\mathcal{C}, \mathcal{C})$.

Définition 3.9. On note $\Pi_{\mathcal{C}}$ la tribu des cylindres sur $\mathcal{C}$, c'est-à-dire la trace de la tribu $E^{\otimes K}$ sur $\mathcal{C}$: $\Pi_{\mathcal{C}} = \sigma(\{A \cap \mathcal{C} : A \in E^{\otimes K}\})$. On a encore que $\Pi_{\mathcal{C}}$ est la plus petite tribu qui rend mesurables les applications coordonnées :

$$\pi_t : \begin{array}{ccc} \mathcal{C} & \rightarrow & E \\ f & \mapsto & f(t) \end{array}, t \in K.$$

Proposition 3.10. Les deux tribus que nous avons définies sur $\mathcal{C}$ coïncident : $\Pi_{\mathcal{C}} = \mathcal{C}$.

2. cf. cours de Probabilités Avancées en 2A ou [9, Theorem 5.16].

Démonstration. Commençons par montrer que $\Pi_{\mathcal{C}} \subset \mathcal{C}$. L'application π_t définie précédemment de $(\mathcal{C}, d_\infty)$ vers (E, d) est continue, donc mesurable pour la tribu borélienne. Donc $\Pi_{\mathcal{C}} \subset \mathcal{C}$.

Montrons maintenant l'inclusion réciproque. Il suffit de prouver que toute boule fermée de $\mathcal{C}$ est dans $\Pi_{\mathcal{C}}$. En effet, la tribu engendrée par les boules fermées est égale à la tribu borélienne.

Soit donc $f \in \mathcal{C}$ et $r > 0$. L'ensemble K est compact, il est donc en particulier séparable. Soit $(x_n)_{n \in \mathbb{N}}$ une suite dense de K , alors

$$\begin{aligned} \overline{B}_{d_\infty}(f, r) &= \{g \in \mathcal{C} : d(f(x), g(x)) \leq r, \forall x \in (x_n)_{n \in \mathbb{N}}\} \\ &= \bigcap_{x \in (x_n)_{n \in \mathbb{N}}} \pi_x^{-1}(\overline{B}_d(f(x), r)). \end{aligned}$$

Il s'agit d'une intersection dénombrable, donc $\overline{B}_{d_\infty}(f, r) \in \Pi_{\mathcal{C}}$. □

3.1.3 Convergence étroite dans $\mathcal{P}(\mathcal{C})$

On cherche dans cette partie et la suivante à mettre en évidence des critères de convergence étroite dans $\mathcal{P}(\mathcal{C})$, plus simples que la définition de base qui consiste à tester la convergence de $\mu_n(f)$ pour tout $f \in \mathcal{C}_b(\mathcal{C})$.

Définition 3.11. On dit qu'une suite $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(\mathcal{C})$ converge au sens des lois finies-dimensionnelles vers $\mu \in \mathcal{P}(\mathcal{C})$ si pour tout $j \in \mathbb{N}^*$ et tous $(t_1, \dots, t_j) \in K^j$, on a

$$(\mu_n)_{t_1, \dots, t_j} \xrightarrow{(e)} \mu_{t_1, \dots, t_j}.$$

On note $\mu_n \xrightarrow{(fd)} \mu$ cette convergence. De manière équivalente pour une suite de processus continus $(X_n)_{n \in \mathbb{N}}$, on a la convergence des lois finies-dimensionnelles si pour tout $j \in \mathbb{N}^*$ et tous $(t_1, \dots, t_j) \in K^j$,

$$(X_n(t_1), \dots, X_n(t_j)) \xrightarrow{(d)} (X(t_1), \dots, X(t_j)).$$

Remarque 3.12. La convergence étroite implique la convergence *fd*, puisque les projections $\pi_{t_1, \dots, t_j}$ sont continues.

En revanche, la réciproque est fausse, comme l'illustre l'exemple classique suivant : Considérons $K = [0, 1]$, U une variable aléatoire uniforme et pour tout $n \in \mathbb{N}$, $X_n(t) = \max(0, 1 - n|U - t|)$, $t \in [0, 1]$, une suite de processus continus.

Exercice : Vérifiez que $X_n \xrightarrow{(fd)} 0$ (le processus constant égal à 0), mais que X_n ne converge pas en loi vers le processus nul.

La convergence au sens des finies-dimensionnelles est donc plus faible, mais le résultat suivant permet de mieux comprendre en quoi cette convergence va nous être utile.

Proposition 3.13. Soit $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(\mathcal{C})$ et $\mu \in \mathcal{P}(\mathcal{C})$. Alors les deux propriétés suivantes sont équivalentes :

- ▷ $\mu_n \xrightarrow{(e)} \mu$;
- ▷ $\mu_n \xrightarrow{(fd)} \mu$ et $(\mu_n)_{n \in \mathbb{N}}$ est tendue.

Démonstration. Si $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement, alors $\{\mu_n\}_{n \in \mathbb{N}} \cup \{\mu\}$ est compact, donc tendu d'après le Théorème de Prokhorov. Par ailleurs, nous avons vu dans la remarque précédente que la convergence étroite implique la convergence des finies-dimensionnelles.

Etudions maintenant le sens réciproque. Puisque $(\mu_n)_{n \in \mathbb{N}}$ est tendue, elle est relativement compacte d'après le Théorème de Prokhorov (on rappelle que $\mathcal{C}$ est bien un espace polonais d'après le Théorème 3.2). Considérons donc $(\mu_{\varphi(n)})_{n \in \mathbb{N}}$ une sous-suite qui converge étroitement et $\nu \in \mathcal{P}(\mathcal{C})$ sa limite. Par hypothèse, $\mu_{\varphi(n)} \xrightarrow{(fd)} \mu$, donc μ et ν ont les mêmes marginales. Donc la Proposition 3.7 implique que $\mu = \nu$. Ainsi nous avons montré que toute sous-suite de $(\mu_n)_{n \in \mathbb{N}}$ qui converge, converge vers μ . Donc $(\mu_n)_{n \in \mathbb{N}}$ converge vers μ (cf. Exercice 2.19). $\square$

Ce résultat simple nous sera extrêmement utile lorsqu'on voudra par exemple démontrer le théorème de Donsker. Nous allons désormais chercher des critères rendant $(\mu_n)_{n \in \mathbb{N}}$ tendue. Pour cela, utilisons les compacts de $\mathcal{C}$ identifiés dans la Section 3.1.1.

3.1.4 Tension dans $\mathcal{P}(\mathcal{C})$

Proposition 3.14. *Un sous-ensemble $\mathcal{M} \subset \mathcal{P}(\mathcal{C})$ est tendu si et seulement si*

- (i) *pour tout $t \in K$, l'ensemble $\{\mu_t : \mu \in \mathcal{M}\}$ est tendu ;*
- (ii) *pour tous $\varepsilon > 0$, $\eta > 0$, il existe $\delta > 0$ tel que*

$$\sup_{\mu \in \mathcal{M}} \mu(\{f \in \mathcal{C} : \omega(f, \delta) \geq \eta\}) \leq \varepsilon.$$

Lorsque les boules fermées de E sont compactes, il suffit de trouver un $t \in K$ tel que $\{\mu_t : \mu \in \mathcal{M}\}$ soit tendu.

Démonstration. Supposons que (i) et (ii) sont vérifiés et montrons la tension (la réciproque est laissée en exercice). Soit $\varepsilon > 0$ et $(t_n)_{n \in \mathbb{N}}$ une suite dénombrable dense de K . Pour tout $n \in \mathbb{N}$, on peut grâce à (i) trouver un compact $K_n^E \subset E$ tel que

$$\sup_{\mu \in \mathcal{M}} \mu_{t_n}(E \setminus K_n^E) \leq \frac{\varepsilon}{2^{n+2}}.$$

Et grâce à (ii), pour tout $m \in \mathbb{N}$, il existe $\delta_m > 0$ tel que

$$\sup_{\mu \in \mathcal{M}} \mu\left(\left\{f : \omega(f, \delta_m) > \frac{1}{2^m}\right\}\right) \leq \frac{\varepsilon}{2^{m+2}}.$$

On définit alors $\mathcal{K}_\varepsilon = \{f \in \mathcal{C} : \forall n \in \mathbb{N}, f(t_n) \in K_n^E \text{ et } \forall m \in \mathbb{N}, \omega(f, \delta_m) \leq \frac{1}{2^m}\}$. D'après le Théorème 3.3 (Arzelà-Ascoli), $\mathcal{K}_\varepsilon$ est relativement compact. On observe aussi que $\mathcal{K}_\varepsilon$ est une intersection de fermés (comme images réciproques de fermés par des fonctions continues), ainsi $\mathcal{K}_\varepsilon$ est fermé, et donc compact. Enfin, on a

$$\sup_{\mu \in \mathcal{M}} \mu(\mathcal{C} \setminus \mathcal{K}_\varepsilon) \leq \sum_{n \in \mathbb{N}} \frac{\varepsilon}{2^{n+2}} + \sum_{m \in \mathbb{N}} \frac{\varepsilon}{2^{m+2}} = \varepsilon.$$

Donc $\mathcal{M}$ est tendu. $\square$

Remarque 3.15. *La convergence des marginales finies-dimensionnelles implique le (i) de la proposition précédente.*

On peut transposer la proposition précédente pour des suites de processus stochastiques, en remplaçant notamment le $\sup_{\mathcal{M}}$ par une $\limsup_{n \rightarrow +\infty}$.

On dira qu'une suite de processus $(X_n)_{n \in \mathbb{N}}$ à valeurs dans $\mathcal{C}$ est tendue si ses lois le sont, i.e. si la suite des mesures de probabilité $(\mathbb{P}_{X_n})_{n \in \mathbb{N}}$ (parfois aussi notée $(\mathcal{L}(X_n))_{n \in \mathbb{N}}$) est tendue.

Corollaire 3.16. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. à valeurs dans $\mathcal{C}$. Alors cette suite est tendue si et seulement si*

- (i) *pour tout $t \in K$, la suite $(X_n(t))_{n \in \mathbb{N}}$ est tendue ;*
- (ii) *pour tous $\varepsilon > 0$, $\eta > 0$, il existe $\delta > 0$ tel que*

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

Lorsque les boules fermées de E sont compactes, il suffit de trouver un $t \in K$ tel que $(X_n(t))_{n \in \mathbb{N}}$ soit tendue.

Démonstration. Il suffit de montrer que si pour tous $\varepsilon > 0$, $\eta > 0$, il existe $\delta > 0$ tel que $\limsup_{n \rightarrow +\infty} \mathbb{P}(\omega(X_n, \delta) \geq \eta) \leq \varepsilon$, alors il existe $\delta' > 0$ tel que $\sup_{n \in \mathbb{N}} \mathbb{P}(\omega(X_n, \delta') \geq \eta) \leq \varepsilon$.

Soit $\delta > 0$ tel que $\limsup_{n \rightarrow +\infty} \mathbb{P}(\omega(X_n, \delta) \geq \eta) \leq \varepsilon/2$. Ainsi il existe un rang N tel que $\sup_{n \geq N} \mathbb{P}(\omega(X_n, \delta) \geq \eta) \leq \varepsilon$. Quitte à choisir δ plus petit, on a également

$$\sup_{n < N} \mathbb{P}(\omega(X_n, \delta) \geq \eta) \leq \varepsilon.$$

D'où le résultat. □

Exemple 3.17. *Reprenons l'exemple de la remarque 3.12. Le critère (ii) n'est pas vérifié pour la suite de processus (X_n) , donc elle n'est pas tendue.*

Bien que le corollaire précédent simplifie considérablement l'étude de la convergence en loi des processus continus, le module de continuité n'est pas toujours facile à contrôler, notamment car il faut contrôler la loi de toute la trajectoire d'un processus. Nous allons voir qu'on peut se ramener à l'étude de deux marginales grâce au critère de Kolmogorov. On étudie ici des processus indexés par un compact de $\mathbb{R}$, disons $[0, 1]$ sans perte de généralité. On note $\mathcal{C} = \mathcal{C}([0, 1], E)$.

Nous admettrons le lemme technique suivant, dont la preuve n'est pas d'une grande difficulté mais n'est pas non plus d'un intérêt essentiel dans ce cours (vous trouverez cette preuve dans [14, Lemme 2.2]). En revanche il est intéressant de noter que ce lemme mène également au théorème de continuité de Kolmogorov (voir Remarque 3.20).

Lemme 3.18. *Soit une fonction $f : [0, 1] \rightarrow E$. On note $D_n = \{\frac{k}{2^n}, 0 \leq k \leq 2^n\}$ l'ensemble des dyadiques d'ordre n , et $D = \bigcup_{n \in \mathbb{N}} D_n$ l'ensemble des dyadiques. On suppose qu'il existe $C, \alpha > 0$ tels que pour tout $n \in \mathbb{N}$, pour tous dyadiques $s, t \in D_n$ consécutifs, on ait $d(f(s), f(t)) \leq C|s - t|^\alpha$.*

Alors pour tous $s, t \in D$,

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

Théorème 3.19 (Critère de tension de Kolmogorov). Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. à valeurs dans $\mathcal{C}$. Supposons que

- ▷ pour tout $t \in [0, 1]$, $(X_n(t))_{n \in \mathbb{N}}$ est tendue ;
- ▷ il existe $p, \beta > 0$ et $C > 0$ tels que

$$\sup_{n \in \mathbb{N}} \mathbb{E} [d(X_n(s), X_n(t))^p] \leq C |s - t|^{1+\beta}, \quad \forall s, t \in [0, 1].$$

Alors $(X_n)_{n \in \mathbb{N}}$ est tendue. On a même plus précisément que pour tout $\alpha \in (0, \frac{\beta}{p})$ et tout $\varepsilon > 0$, il existe un $M > 0$ tel que

$$\sup_{n \in \mathbb{N}} \mathbb{P} \left(\sup_{s, t \in [0, 1], s \neq t} \frac{d(X_n(s), X_n(t))}{|s - t|^\alpha} \geq M \right) \leq \varepsilon.$$

Lorsque les boules fermées de E sont compactes, il suffit de trouver un $t \in [0, 1]$ tel que $(X_n(t))_{n \in \mathbb{N}}$ soit tendue.

Démonstration. Notons

$$[f]_\alpha = \sup_{s, t \in [0, 1], s \neq t} \frac{d(f(s), f(t))}{|s - t|^\alpha}$$

la seminorme Hölder de paramètre α , et remarquons que si $f \in \mathcal{C}$,

$$\sup_{s, t \in D, s \neq t} \frac{d(f(s), f(t))}{|s - t|^\alpha} = [f]_\alpha.$$

Ainsi, par le Lemme 3.18, on a pour tout $n \in \mathbb{N}$ et pour tout $M > 0$,

$$\mathbb{P} \left([X_n]_\alpha > \frac{2M}{1 - 2^{-\alpha}} \right) \leq \mathbb{P} (\exists m \in \mathbb{N}, \exists k \in \{0, \dots, 2^m - 1\}, d(X_n(k2^{-m}), X_n((k+1)2^{-m})) > M2^{-m\alpha}).$$

On obtient donc

$$\begin{aligned} \mathbb{P} \left([X_n]_\alpha > \frac{2M}{1 - 2^{-\alpha}} \right) &\leq \sum_{m \in \mathbb{N}} 2^m \max_{k \in \{0, \dots, 2^m - 1\}} \mathbb{P} (d(X_n(k2^{-m}), X_n((k+1)2^{-m})) > M2^{-m\alpha}) \\ &\leq \sum_{m \in \mathbb{N}} 2^m \max_{k \in \{0, \dots, 2^m - 1\}} \frac{\mathbb{E} [d(X_n(k2^{-m}), X_n((k+1)2^{-m}))^p]}{M^p 2^{-pm\alpha}}, \end{aligned}$$

en appliquant l'inégalité de Markov à la deuxième ligne. Par hypothèse, il vient donc

$$\mathbb{P} \left([X_n]_\alpha > \frac{2M}{1 - 2^{-\alpha}} \right) \leq \frac{C}{M^p} \sum_{m \in \mathbb{N}} 2^{m-m(1+\beta)+pm\alpha},$$

et on s'assure que cette somme est finie en choisissant $\alpha \in (0, \frac{\beta}{p})$. De la sorte, il existe $C' > 0$ (qui dépend uniquement de α, β, p et C) telle que

$$\mathbb{P} \left([X_n]_\alpha > \frac{2M}{1 - 2^{-\alpha}} \right) \leq \frac{C'}{M^p}. \quad (3.1)$$

Enfin, il suffit de remarquer que pour tout $\delta > 0$, $\omega(X_n, \delta) \leq [X_n]_\alpha \delta^\alpha$. On déduit de cette inégalité et de (3.1) que pour $\varepsilon > 0$ et $\eta > 0$ fixés, on a

$$\begin{aligned} \mathbb{P} (\omega(X_n, \delta) \geq \eta) &\leq \mathbb{P} ([X_n]_\alpha \geq \eta \delta^{-\alpha}) \leq \frac{2^p C' \delta^{\alpha p}}{(1 - 2^{-\alpha})^p \eta^p} \\ &\leq \varepsilon \end{aligned}$$

en choisissant δ suffisamment petit. Donc par la Proposition 3.14, on conclut que (X_n) est tendue. $\square$

Remarque 3.20. On peut déduire du Lemme 3.18 et d'un calcul similaire à la preuve précédente le Théorème de continuité de Kolmogorov. Notons

$$[f]_\alpha^\sim = \sup_{s,t \in D, s \neq t} \frac{d(f(s), f(t))}{|s - t|^\alpha}.$$

On va travailler avec un processus X à valeurs dans $E^{[0,1]}$ (on ne le suppose donc pas continu), qui vérifie

$$\mathbb{E}[d(X(s), X(t))^p] \leq C|s - t|^{1+\beta}, \quad \forall s, t \in [0, 1]. \quad (3.2)$$

En reproduisant la preuve précédente jusqu'à (3.1), on obtient que pour $\alpha \in (0, \frac{\beta}{p})$,

$$\mathbb{P}\left([X]_\alpha^\sim > \frac{2M}{1 - 2^{-\alpha}}\right) \leq \frac{C'}{M^p}.$$

Ceci implique (vérifiez-le !) que pour tout $q \in (0, p)$, $\mathbb{E}[(X]_\alpha^\sim)^q] < \infty$. Ainsi sur un ensemble $\Omega_0 \in \mathcal{F}$ de probabilité 1, $[X]_\alpha^\sim < \infty$, ce qui revient à dire que $\forall \omega \in \Omega_0$, $X(\omega)$ est α -Hölder continu sur D . On appelle $\tilde{X}(\omega)$ le prolongement par continuité de $X(\omega)$ à tout l'intervalle $[0, 1]$. Nous allons vérifier que pour tout $t \in [0, 1]$, $\mathbb{P}(\tilde{X}_t = X_t) = 1$. Le théorème de Kolmogorov stipule donc que si X vérifie (3.2), alors X admet une modification Hölder continue.

Montrons donc que pour tout $t \in [0, 1]$, $\mathbb{P}(\tilde{X}_t = X_t) = 1$. Soit $t \in [0, 1]$ et $(t_n)_{n \in \mathbb{N}} \subset D$ telle que $t_n \rightarrow t$. Alors (3.2) et l'inégalité de Markov impliquent que $X_{t_n} \rightarrow X_t$ en probabilité. D'autre part, la construction même de $\tilde{X}$ indique que $X_{t_n} \rightarrow \tilde{X}_t$ presque sûrement. D'où le résultat.

3.2 Convergence étroite dans l'espace des fonctions continues sur $\mathbb{R}_+$

Les processus que nous étudierons par la suite seront souvent indicés par $\mathbb{R}_+$ plutôt que par un compact. Ceci pose quelques problèmes techniques : en premier lieu, la "distance" uniforme n'est pas nécessairement à valeurs dans $\mathbb{R}_+$ (elle peut être infinie, par exemple si $E = \mathbb{R}$, ce n'est donc plus une distance) ; en outre, l'espace $\mathcal{C}(\mathbb{R}_+; E)$ n'est pas séparable, donc *a fortiori* pas polonais.

On munit en fait l'espace $\mathcal{C}(\mathbb{R}_+; E)$ d'une distance de Fréchet, induite par les distances sur $\mathcal{C}([0, N]; E)$, $N > 0$:

$$d_\infty(f, g) = \sum_{N \in \mathbb{N}^*} 2^{-N} \left(1 \wedge \sup_{t \in [0, N]} d(f(t), g(t)) \right), \quad \forall f, g \in \mathcal{C}(\mathbb{R}_+; E).$$

Cette distance engendre la topologie dite de la convergence uniforme sur les compacts, ou topologie *compacte-ouverte*. On peut vérifier que si E est polonais, $\mathcal{C}(\mathbb{R}_+; E)$ muni de cette distance l'est également (*exercice*).

En pratique, on peut montrer que vérifier la tension dans $\mathcal{C}(\mathbb{R}_+; E)$ revient à vérifier la tension dans chacun des espaces $\mathcal{C}([0, N]; E)$, $N \in \mathbb{N}^*$:

Proposition 3.21. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathcal{C}(\mathbb{R}_+; E)$. Alors on a l'équivalence suivante :

- ▷ $(X_n)_{n \in \mathbb{N}}$ est tendue dans $\mathcal{C}(\mathbb{R}_+; E)$;
- ▷ pour tout $N \in \mathbb{N}^*$, la suite des restrictions $(\{X_n(t), t \in [0, N]\})_{n \in \mathbb{N}}$ est tendue dans $\mathcal{C}([0, N]; E)$.

3.3 Théorème de Donsker

On dispose désormais de tous les outils et notions nécessaires pour démontrer le théorème de Donsker, que nous avons présenté dans la partie 1.1.1.

Sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, on se donne une suite de variables aléatoires $(X_k)_{k \in \mathbb{N}}$ i.i.d. et à valeurs dans $\mathbb{R}^d$, de moyenne nulle et de matrice de covariance l'identité.

On définit la marche aléatoire $(S_n)_{n \in \mathbb{N}}$ par $S_0 = 0$ et pour tout $n \geq 1$, $S_n = \sum_{k=1}^n X_k$. On définit alors le processus continu $(W_t^{(1)})_{t \in \mathbb{R}_+}$, qui étend la marche aléatoire, par

$$W_t^{(1)} = S_{[t]} + (t - [t])X_{[t]+1}.$$

Le processus changé d'échelle est alors, pour tout $n \geq 1$,

$$W_t^{(n)} = \frac{1}{\sqrt{n}} W_{nt}^{(1)}, \quad t \geq 0,$$

et on a clairement $W^{(n)} \in \mathcal{C}(\mathbb{R}_+; \mathbb{R}^d)$.

Théorème 3.22. *Soit $(X_k)_{k \in \mathbb{N}}$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{R}^d$, admettant un moment d'ordre 2. On suppose que les X_k sont de moyenne nulle et de matrice de covariance l'identité.*

Alors la suite de processus $(W^{(n)})_{n \in \mathbb{N}^}$ définie ci-dessus converge en loi dans $\mathcal{C}(\mathbb{R}_+; \mathbb{R}^d)$ vers un processus $(W_t)_{t \in \mathbb{R}_+}$, appelé mouvement brownien standard (de dimension d). Ce processus est caractérisée par la propriété que pour tous $0 \leq t_1 \leq \dots \leq t_l$, les v.a. $W_{t_{k+1}} - W_{t_k}$, $1 \leq k \leq l-1$ sont des gaussiennes indépendantes, centrées et de variances respectives $(t_{k+1} - t_k) I_d$.*

Démonstration. Nous allons démontrer ce résultat sous des hypothèses plus restrictives que celles annoncées : nous supposons que les X_k ont des moments d'ordre 4 finis. La tension dans le cadre générale demande un peu plus de travail. Vous pouvez consulter cette preuve dans [2, 13, 14].

Le schéma de preuve repose sur la caractérisation de la convergence en loi des processus à valeurs dans $\mathcal{C}$ donnée dans la Proposition 3.13 : nous allons donc montrer la convergence des lois finies-dimensionnelles (ce qui nous donnera au passage les lois finies-dimensionnelles de la loi limite), puis la tension.

Première étape : convergence des lois finies-dimensionnelles. Soit $l \in \mathbb{N}^*$ et $0 \leq t_1 \leq \dots \leq t_l$. Il est clair que les variables aléatoires $S_{[nt_{k+1}]} - S_{[nt_k]}$, $1 \leq k \leq l$, sont indépendantes, centrées et de variances respectives $([nt_{k+1}] - [nt_k]) I_d$. Ainsi, d'après le Théorème central limit multidimensionnel³, on a

$$\left(\frac{1}{\sqrt{n}}(S_{[nt_2]} - S_{[nt_1]}), \dots, \frac{1}{\sqrt{n}}(S_{[nt_l]} - S_{[nt_{l-1}]}) \right) \xrightarrow[n \rightarrow \infty]{(d)} (N_1, \dots, N_l),$$

3. on peut aussi démontrer cette convergence "à la main", en utilisant l'indépendance et la caractérisation de la convergence en loi par le Théorème de Lévy.

où $N_1, \dots, N_l$ sont des gaussiennes indépendantes, centrées et de variances respectives $(t_{k+1} - t_k) I_d$. On en déduit par transformation continue que

$$\left(\frac{1}{\sqrt{n}} S_{\lfloor nt_1 \rfloor}, \frac{1}{\sqrt{n}} S_{\lfloor nt_2 \rfloor}, \dots, \frac{1}{\sqrt{n}} S_{\lfloor nt_l \rfloor} \right) \xrightarrow[n \rightarrow \infty]{(d)} (N_1, N_2, \dots, N_1 + \dots + N_l).$$

Pour passer à la convergence des finies-dimensionnelles de $W^{(n)}$, observons que

$$\left| W_t^{(n)} - \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} \right| \leq \frac{|X_{\lfloor nt \rfloor + 1}|}{\sqrt{n}}$$

converge p.s. vers 0 (et donc aussi en probabilité). Donc par le Lemme de Slutsky (voir DM), il vient que

$$(W_{t_1}^{(n)}, \dots, W_{t_l}^{(n)}) \xrightarrow[n \rightarrow \infty]{(d)} (N_1, N_2, \dots, N_1 + \dots + N_l).$$

Deuxième étape : tension. On a vu dans la Proposition 3.21 qu'il suffit de montrer la tension de la restriction $(\{W_t^{(n)}, t \in [0, N]\})_{n \in \mathbb{N}}$ pour tout $N \in \mathbb{N}^*$. Nous allons considérer uniquement le cas $N = 1$ afin d'alléger les notations, mais vous pouvez vérifier que cela n'engendre aucune perte de généralité.

Nous allons chercher à appliquer le critère de tension de Kolmogorov (Théorème 3.19). Soient $s < t \in [0, 1]$. On commence par supposer que $ns \in \mathbb{N}$ et $nt \in \mathbb{N}$. On a alors deux cas. *Premier cas* : si $n(t-s) > 1$, alors en notant $X_k^{(1)}, \dots, X_k^{(d)}$ les composantes du vecteur aléatoire X_k , on obtient

$$\begin{aligned} \mathbb{E} [|W_s^{(n)} - W_t^{(n)}|^4] &= \frac{1}{n^2} \mathbb{E} [|S_{ns} - S_{nt}|^4] = \frac{1}{n^2} \mathbb{E} [|S_{n(t-s)}|^4] \\ &= \frac{1}{n^2} \sum_{1 \leq j_1, j_2 \leq d} \mathbb{E} \left[\sum_{1 \leq k_1, k_2 \leq n(t-s)} X_{k_1}^{(j_1)} X_{k_2}^{(j_1)} \sum_{1 \leq l_1, l_2 \leq n(t-s)} X_{l_1}^{(j_2)} X_{l_2}^{(j_2)} \right]. \end{aligned}$$

Simplifions cette somme en tenant compte des covariances nulles entre coordonnées d'un même vecteur et de l'indépendance entre vecteurs différents (on rappelle que les vecteurs sont d'espérance nulle). On obtient alors

$$\begin{aligned} \mathbb{E} [|W_s^{(n)} - W_t^{(n)}|^4] &= \frac{1}{n^2} \sum_{j_1 \neq j_2} \sum_{1 \leq k_1, k_2 \leq n(t-s)} \mathbb{E} [X_{k_1}^{(j_1)} X_{k_2}^{(j_1)}] \sum_{1 \leq l_1, l_2 \leq n(t-s)} \mathbb{E} [X_{l_1}^{(j_2)} X_{l_2}^{(j_2)}] \\ &\quad + \frac{1}{n^2} \sum_{1 \leq j \leq d} \sum_{1 \leq k_1, k_2, k_3, k_4 \leq n(t-s)} \mathbb{E} [X_{k_1}^{(j)} X_{k_2}^{(j)} X_{k_3}^{(j)} X_{k_4}^{(j)}] \\ &= \frac{1}{n^2} \sum_{j_1 \neq j_2} \sum_{1 \leq k \leq n(t-s)} \mathbb{E} [(X_k^{(j_1)})^2] \sum_{1 \leq l \leq n(t-s)} \mathbb{E} [(X_l^{(j_2)})^2] \\ &\quad + \frac{1}{n^2} \sum_{1 \leq j \leq d} \left\{ \sum_{1 \leq k \leq n(t-s)} \mathbb{E} [(X_k^{(j)})^4] + \sum_{k_1 \neq k_2} \mathbb{E} [(X_{k_1}^{(j)})^2] \mathbb{E} [(X_{k_2}^{(j)})^2] \right\} \\ &= \frac{1}{n^2} \frac{d(d-1)}{2} (n(t-s))^2 + \frac{1}{n^2} \left(M_4 + d \frac{(n(t-s))(n(t-s)-1)}{2} \right), \end{aligned}$$

où on a noté M_4 la somme (en j et en k) des moments d'ordre 4. Or cette quantité est clairement majorée par $n(t-s) d \max_{1 \leq j \leq d} \mathbb{E} [(X_k^{(j)})^4]$, où le max ne dépend pas de k

(puisqu'on a supposé que les X_k sont i.i.d.). Par conséquent, il existe une constante $C > 0$ qui ne dépend que de d et des moments d'ordre 4 de X_k (pas de n), telle que

$$\begin{aligned}\mathbb{E} \left[|W_s^{(n)} - W_t^{(n)}|^4 \right] &\leq \frac{d(d-1) + d}{2} (t-s)^2 + \frac{d\mathbb{E}[(X_k^{(j)})^4]}{n} (t-s) \\ &\leq C (t-s)^2,\end{aligned}\tag{3.3}$$

en utilisant dans la dernière inégalité l'hypothèse de ce premier cas, à savoir $n(t-s) > 1$. On a maintenant un *deuxième cas*, avec $n(t-s) \leq 1$. Comme on suppose toujours que $n(t-s)$ est entier, la seule possibilité est $n(t-s) = 1$. On a alors

$$\mathbb{E} \left[|W_s^{(n)} - W_t^{(n)}|^4 \right] = \mathbb{E} \left[\frac{1}{n^2} |X_{nt}|^4 \right] = \frac{1}{n} \mathbb{E}[|X_i|^4] = C(t-s)^2,$$

avec $C = \mathbb{E}[|X_i|^4]$. Ainsi dans les deux cas considérés, $\mathbb{E} \left[|W_s^{(n)} - W_t^{(n)}|^4 \right] \leq C(t-s)^2$, quelque soit $s < t$ avec ns et nt entiers.

Considérons maintenant le cas où ns et nt ne sont plus nécessairement entiers. Si $\lfloor ns \rfloor = \lfloor nt \rfloor$, alors $W_t^{(n)} - W_s^{(n)} = \sqrt{n}(t-s)X_{\lfloor nt \rfloor + 1}$, et donc

$$\mathbb{E} \left[|W_s^{(n)} - W_t^{(n)}|^4 \right] = n^2(t-s)^4 \mathbb{E}[|X_1|^4] \leq C(t-s)^2,$$

en utilisant le fait que $t-s \leq 1/n$. Si $\lfloor ns \rfloor < \lfloor nt \rfloor$, alors on écrit

$$\sqrt{n}|W_s^{(n)} - W_t^{(n)}| \leq |W_{ns}^{(1)} - W_{\lfloor ns \rfloor + 1}^{(1)}| + |W_{\lfloor ns \rfloor + 1}^{(1)} - W_{\lfloor nt \rfloor}^{(1)}| + |W_{\lfloor nt \rfloor}^{(1)} - W_{nt}^{(1)}|.$$

Par l'inégalité de Minkowski et les deux résultats précédents, on a alors

$$\begin{aligned}\mathbb{E} \left[|W_s^{(n)} - W_t^{(n)}|^4 \right]^{\frac{1}{4}} &\leq C^{\frac{1}{4}} \left(|\lfloor ns \rfloor + 1 - ns|^{\frac{1}{2}} + |\lfloor nt \rfloor - \lfloor ns \rfloor - 1|^{\frac{1}{2}} + |nt - \lfloor nt \rfloor|^{\frac{1}{2}} \right) \\ &\leq 3C^{\frac{1}{4}} |t-s|^{\frac{1}{2}}.\end{aligned}$$

Nous pouvons maintenant conclure grâce aux deux étapes précédentes et la Proposition 3.13. La loi limite est caractérisée ses lois finies-dimensionnelles (Proposition 3.7), décrites ici dans l'énoncé du théorème. On appelle une telle loi la *mesure de Wiener*. $\square$

Remarque 3.23. *On voit que c'est l'application du critère de Kolmogorov qui nécessite des moments d'ordre strictement supérieur à 2 pour les X_k afin d'obtenir une borne en $(t-s)^{1+\beta}$, $\beta > 0$, dans l'inégalité (3.3).*

Grâce au théorème de Donsker, on peut passer d'une information sur la loi du mouvement brownien à une information sur la loi de la marche aléatoire simple, et vice versa. C'est ce qu'illustrent les deux exercices suivants.

Exercice 3.24. *Soit $(X_k)_{k \in \mathbb{N}}$ des v.a. réelles i.i.d. centrées et de variance $\sigma^2 > 0$. Montrer que la suite de v.a. $(\frac{\max_{1 \leq k \leq n} S_k}{\sigma\sqrt{n}})_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire de fonction de répartition donnée par $F(x) = 2 \int_0^x \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$ [on pourra utiliser le principe de réflexion pour le brownien, i.e. $\sup_{t \in [0,1]} |W_t| \stackrel{(d)}{=} |W_1|$]. Notez qu'un tel résultat n'aurait pu être obtenu par simple application du TCL.*

Exercice 3.25. Loi de l'arcsinus : le but de l'exercice est de démontrer que si $(W_t)_{t \in [0,1]}$ est un mouvement brownien standard et que $R = \sup\{t \in [0,1] : W_t = 0\}$ est le dernier temps de passage en 0 du brownien, alors la loi de R est donnée par $\mathbb{P}(R \leq t) = \frac{2}{\pi} \arcsin(\sqrt{t})$, $t \in [0,1]$.

Soit $(X_k)_{k \in \mathbb{N}}$ une suite de v.a. i.i.d. de loi donnée par $\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = \frac{1}{2}$. On note $S_n = \sum_{k=1}^n X_k$ la marche aléatoire associée ($S_0 = 0$). Soit $R_n = \max\{k \in \{1, \dots, n\} : S_k = 0\}$.

1) Le but de cette question est de montrer que pour tout $n \geq 1$, $\mathbb{P}(S_1 \neq 0, \dots, S_{2n} \neq 0) = \mathbb{P}(S_{2n} = 0)$.

a) Exprimer $\mathbb{P}(S_1 \neq 0, \dots, S_{2n} \neq 0)$ en fonction de $\mathbb{P}(S_1 > 0, \dots, S_{2n} > 0)$.

b) Montrer que $\mathbb{P}(S_1 > 0, \dots, S_{2n} > 0) = \frac{1}{2} \mathbb{P}(S_1 \geq 0, \dots, S_{2n-1} \geq 0)$.

c) En déduire que $\mathbb{P}(S_1 \neq 0, \dots, S_{2n} \neq 0) = \mathbb{P}(S_1 \geq 0, \dots, S_{2n} \geq 0)$.

d) Déduire de la question précédente que $\mathbb{P}(S_1 \neq 0, \dots, S_{2n} \neq 0) = \mathbb{P}(S_{2n} = 0)$.

On cherchera pour cela à établir une transformation bijective entre les chemins $\{S_k\}_{1 \leq k \leq 2n}$ tels que $S_{2n} = 0$, et les chemins qui restent positifs (i.e. $S_k \geq 0$ pour tout $k \in \{1, \dots, 2n\}$). On pourra utiliser le premier temps où le minimum de la marche (celle qui vérifie $S_{2n} = 0$) atteint son minimum.

2) a) Déduire de la question 1) un lien entre $\mathbb{P}(R_{2n} = 2k)$ d'une part, et $\mathbb{P}(S_{2j} = 0)$ et $\mathbb{P}(S_{2(n-k)} = 0)$ d'autre part.

b) Montrer que $\lim_{k \rightarrow +\infty} \sqrt{\pi k} \mathbb{P}(S_{2k} = 0) = 1$.

c) En déduire que pour tout $x \in (0,1)$, $\lim_{n \rightarrow +\infty} n\pi \sqrt{x(1-x)} \mathbb{P}(R_{2n} = 2\lfloor xn \rfloor) = 1$.

3) On veut maintenant utiliser le théorème de Donsker pour démontrer la loi de l'arcsinus.

a) On note $\Phi : f \mapsto \sup\{t \in [0,1] : f(t) = 0\}$. L'application Φ est-elle continue de $\mathcal{C}([0,1]; \mathbb{R})$ vers $\mathbb{R}$?

b) On admettra la propriété suivante du brownien : presque sûrement, le brownien prend des valeurs positives et négatives dans tout voisinage de ses zéros, et $W_1 \neq 0$. En déduire que W est p.s. un point de continuité de Φ .

c) Conclure.

3.4 Bref aperçu de la tension dans l'espace des fonctions càdlàg

Reprenons l'exemple de la convergence de la marche aléatoire par changement d'échelle dans le théorème de Donsker. On peut être tenté de ne pas interpoler linéairement ce processus⁴ et de considérer plus simplement

$$\widetilde{W}_t^{(1)} = S_{\lfloor t \rfloor} \quad \text{et} \quad \widetilde{W}_t^{(n)} = \frac{1}{\sqrt{n}} \widetilde{W}_{nt}^{(1)}, \quad n \in \mathbb{N}^*, \quad t \geq 0.$$

Cependant la suite de processus $(\widetilde{W}^{(n)})_{n \in \mathbb{N}}$ n'est pas dans $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ car ses éléments sont discontinus. Les critères de tension établis dans les parties 3.1.4 et 3.2 ne sont donc pas

4. ce qui sera le cas dans le prochain chapitre.

applicables. Le bon espace pour travailler est plutôt $\mathcal{D}(\mathbb{R}_+; \mathbb{R}^d)$, l'espace des fonctions dites càdlàg, c'est-à-dire *continue à droite avec une limite à gauche* en tout point (composante par composante). Grâce au théorème de Donsker, on peut montrer la convergence suivante :

Exercice 3.26. *Montrer que sous les hypothèses du Théorème 3.22, on a pour tout $T > 0$ et tout $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{t \in [0, T]} |\widetilde{W}_t^{(n)} - W_t^{(n)}| \geq \varepsilon \right) = 0.$$

En déduire que la suite de processus $(\widetilde{W}^{(n)})_{n \in \mathbb{N}}$ converge en loi dans l'espace $\mathcal{D}(\mathbb{R}_+; \mathbb{R}^d)$, muni de la topologie de la convergence uniforme sur les compacts, vers le mouvement brownien.

Cependant l'espace $\mathcal{D}(\mathbb{R}_+; \mathbb{R}^d)$ n'est pas séparable si on le munit de la topologie de la convergence uniforme sur les compacts ; penser par exemple aux fonctions $f_y = \mathbb{1}_{[y, 1]}$ définies sur $[0, 1]$ pour tout $y \in (0, 1)$, qui vérifient $\|f_y - f_z\|_\infty = 1$ dès que $y \neq z$.

On peut munir cet espace d'une topologie qui le rend polonais, et nous présentons ici succinctement la topologie J1 qui est une des quatre topologies de Skorokhod. Comme pour l'espace des fonctions continues, on se restreint dans un premier temps à un intervalle de temps compact, disons $[0, T]$, $T > 0$.

Définition 3.27 (Topologie de Skorokhod). *On dit qu'une suite de fonctions $(f_n)_{n \in \mathbb{N}} \subset \mathcal{D}([0, T]; \mathbb{R}^d)$ converge vers $f \in \mathcal{D}([0, T]; \mathbb{R}^d)$ pour la topologie de Skorokhod s'il existe une suite de fonctions $\varphi_n : [0, T] \rightarrow [0, T]$ strictement croissantes telle que*

$$\|f_n \circ \varphi_n - f\|_\infty \xrightarrow{n \rightarrow \infty} 0 \quad \text{et} \quad \|\varphi_n - \text{Id}\|_\infty \xrightarrow{n \rightarrow \infty} 0.$$

La distance associée à cette notion de convergence est donnée ainsi : pour $f, g \in \mathcal{D}([0, 1]; \mathbb{R}^d)$, on définit

$$d_S(f, g) = \inf_{\varphi \in \text{Homeo}_+} \max(\|f - g \circ \varphi\|_\infty, \|\text{Id} - \varphi\|_\infty),$$

où Homeo_+ est l'ensemble des homéomorphismes croissant de $[0, 1]$ vers $[0, 1]$. On peut vérifier que la restriction de cette distance à l'espace des fonctions continues engendre la topologie uniforme.

Exercice 3.28. *Pour $n \geq 3$, soient $f_n = \mathbb{1}_{[\frac{1}{2} - \frac{1}{n}, 1]}$ et $g_n = \mathbb{1}_{[\frac{1}{2} + \frac{1}{n}, 1]}$. Montrer que ces deux suites convergent vers la fonction $h = \mathbb{1}_{[\frac{1}{2}, 1]}$ pour la topologie de Skorokhod.*

A l'aide de cette exemple, déduire que $\mathcal{D}([0, 1]; \mathbb{R})$, muni de la topologie de Skorokhod, n'est pas un espace vectoriel topologique (!).

La distance d_S définie ci-dessus rend $\mathcal{D}([0, 1]; \mathbb{R}^d)$ séparable mais pas complet. Néanmoins on peut trouver une distance (qu'on n'écrira pas, voir par exemple [2]) qui engendre la même topologie et qui rend $\mathcal{D}([0, 1]; \mathbb{R}^d)$ complet. Ainsi $\mathcal{D}([0, 1]; \mathbb{R}^d)$ est polonais.

Il est ensuite possible d'étudier des critères de tension sur $\mathcal{D}([0, 1]; \mathbb{R}^d)$ comme nous l'avons fait pour les fonctions continues, en remplaçant le module de continuité par un module adapté aux discontinuités des fonctions càdlàg. Nous n'entrerons pas dans les détails et renvoyons le lecteur à [2], nous contentant ici du critère énoncé dans le prochain paragraphe.

Dans le chapitre suivant, nous approcherons des diffusions par des chaînes de Markov. Comme pour le Théorème de Donsker, la question se posera de savoir comment on étend les chaînes de Markov (temps discret) à des processus en temps continu. On penchera pour l'extension en des trajectoires constantes par morceaux, comme pour $\widetilde{W}$ ci-dessus, afin de préserver la propriété de Markov, qui sera essentielle pour identifier la limite. Cependant, les objets limites auront des trajectoires continues, et dans ce contexte, on pourra utiliser le critère de tension suivant.

Théorème 3.29. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathcal{D}(\mathbb{R}_+; E)$. On a l'équivalence suivante :*

- ▷ $(X_n)_{n \in \mathbb{N}}$ est tendue pour la topologie de Skorokhod et toutes ses valeurs d'adhérence sont des v.a. à valeurs dans $\mathcal{C}(\mathbb{R}_+; E)$ (ou de manière équivalente, les valeurs d'adhérence des lois de (X_n) sont des mesures de probabilités chargeant l'espace $\mathcal{C}(\mathbb{R}_+; E)$) ;
- ▷ pour tout $N \in \mathbb{N}^*$, tout $\varepsilon > 0$ et tout $\eta > 0$, il existe $n_0 \in \mathbb{N}^*$ et $A > 0$ tels que

$$\text{si } n \geq n_0, \text{ alors } \mathbb{P} \left(\sup_{t \in [0, N]} |X_n(t)| > A \right) \leq \varepsilon;$$

et il existe en plus $\delta > 0$ tel que

$$\text{si } n \geq n_0, \text{ alors } \mathbb{P} \left(\omega(X_n|_{[0, N]}, \delta) > \eta \right) \leq \varepsilon.$$

Lorsque les boules fermées de E sont compactes, il suffit d'avoir que $\mathbb{P}(|X_n(0)| > A) \leq \varepsilon$ pour que la première condition soit vérifiée.

Démonstration. Voir Proposition 3.26, chapitre VI de [7]. □

Chapitre 4

Théorie de Stroock-Varadhan de l'approximation des diffusions

Dans cette partie des notes, nous présentons une introduction à la théorie de Stroock-Varadhan de l'approximation des diffusions, complétant ainsi l'exemple de l'urne d'Ehrenfest exposé dans l'introduction du cours. Ces notes s'inspirent du cours de Berestycki [1] et du livre de Durrett [6]. Pour un traitement plus exhaustif, on pourra se référer au chapitre 11 du livre de Stroock et Varadhan [18].

Partant d'une suite de chaînes de Markov $\{(Y_k^{(n)})_{k \in \mathbb{N}}, n \in \mathbb{N}\}$, on considère pour tout $n \in \mathbb{N}$ l'extension en temps continu dilatée d'un facteur n en temps :

$$X_t^{(n)} = Y_{[nt]}^{(n)}, \quad t \geq 0. \quad (4.1)$$

On s'intéresse à la convergence de la suite de processus càdlàg $(X^{(n)})_{n \in \mathbb{N}}$ vers une diffusion. On préférera travailler avec les extensions càdlàg plutôt que continues car les premières préservent la propriété de Markov, qui sera essentielle pour identifier la limite. A titre d'exemple d'une telle suite de chaînes de Markov, on a avec les notations de la section précédente que la marche aléatoire simple renormalisée $\frac{1}{\sqrt{n}}S_k$ correspond à $Y_k^{(n)}$ et le processus $\widetilde{W}^{(n)}$ correspond ici à $X^{(n)}$.

Rappelons l'approche générale présentée dans ce cours pour obtenir des passages à la limite d'une suite :

- (i) Montrer une forme de compacité de la suite, en l'occurrence la tension de la suite de chaînes de Markov renormalisées.
- (ii) Identifier la limite : montrer que toute sous-suite admet une sous-suite qui converge, et que la limite est toujours la même.

4.1 Rappels sur les processus de Markov et les martingales

Dans cette partie et la suivante, on travaillera sur des espaces de probabilité filtrés, c'est-à-dire $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ satisfaisant les conditions usuelles : la tribu $\mathcal{F}$ est $\mathbb{P}$ -complète (si A est un sous-ensemble de $B \in \mathcal{F}$ et $\mathbb{P}(B) = 0$, alors $A \in \mathcal{F}$) et la filtration est continue à droite :

$$\mathcal{F}_t = \bigcap_{s > t} \mathcal{F}_s, \quad \forall t \in \mathbb{R}_+.$$

Définition 4.1. Soit $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ un espace de probabilité filtré. On appelle martingale locale tout processus stochastique $M : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}^d$ qui est $(\mathcal{F}_t)$ -adapté et pour lequel il existe une suite de temps d'arrêt $(T_n)_{n \in \mathbb{N}}$ tels que

- ▷ pour tout $n \in \mathbb{N}$, $T_n \leq T_{n+1}$ p.s. et $\lim_{n \rightarrow +\infty} T_n = +\infty$ p.s. ;
- ▷ pour tout $n \in \mathbb{N}$, le processus arrêté $M^{T_n} := (M_{t \wedge T_n})_{t \in \mathbb{R}_+}$ est une vraie martingale.

On notera $\mathcal{M}_{c,loc}$ l'espace des martingales locales continues.

On rappelle la définition et l'existence du crochet de martingales, donné par la proposition suivante.

Proposition 4.2. Soient $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ un espace de probabilité filtré et $M, N \in \mathcal{M}_{c,loc}$ des $(\mathcal{F}_t)$ -martingales locales. Alors il existe un unique¹ processus noté $[M, N]$ qui est $(\mathcal{F}_t)$ -adapté, continu, et tel que $MN - [M, N]$ soit une martingale locale partant de 0.

En outre, si on définit pour tout $n \in \mathbb{N}$ et pour tout $t \geq 0$ le processus

$$[M, N]_t^{(n)} = \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}}) (N_{(k+1)2^{-n}} - N_{k2^{-n}}),$$

alors $[M, N]^{(n)}$ converge uniformément sur les compacts en probabilité vers $[M, N]$.

Enfin, la notion de chaîne de Markov sera centrale dans ce chapitre.

Définition 4.3. Soient $(E, \mathcal{E})$ un espace mesurable, $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une filtration discrète. On appelle chaîne de Markov tout processus $X : \Omega \times \mathbb{N} \rightarrow (E, \mathcal{E})$ qui est $(\mathcal{F}_n)$ -adapté et vérifie

$$\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A \mid X_n), \quad \forall n \in \mathbb{N}, \forall A \in \mathcal{E}.$$

On rappelle qu'une chaîne de Markov est *homogène*² si pour tout $n \in \mathbb{N}$, la loi de X_{n+1} conditionnellement à X_n ne dépend pas de n .

La loi d'une chaîne de Markov peut être décrite par la donnée d'une loi initial $\mu \in \mathcal{P}(E)$ et d'un *noyau de transition*. On rappelle qu'un noyau de transition est une application $Q : E \times \mathcal{E} \rightarrow [0, 1]$ telle que

- ▷ pour tout $A \in \mathcal{E}$, $x \mapsto Q(x, A)$ est mesurable ;
- ▷ pour tout $x \in E$, $A \mapsto Q(x, A)$ est une mesure de probabilité.

Ainsi, étant donnés μ et Q , on peut construire une chaîne de Markov telle que $\mathbb{P}(X_1 \in A) = \mu Q(A) = \int_E Q(x, A) \mu(dx)$.

A l'inverse, pour une chaîne de Markov donnée, $\mathbb{P}(X_1 \in A \mid X_0) =: Q(X_0, A)$ fournit un noyau de transition.

Définition 4.4. On appelle mouvement brownien standard un processus stochastique continu à valeurs dans $\mathbb{R}^d$ qui vérifie : (i) $W_0 = 0$ p.s. ; (ii) pour tous $0 \leq t_1 \leq \dots \leq t_l$, les v.a. $W_{t_{k+1}} - W_{t_k}$, $1 \leq k \leq l-1$ sont des gaussiennes indépendantes, centrées et de variances respectives $(t_{k+1} - t_k) I_d$.

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ un espace de probabilité filtré. On appelle $\mathcal{F}_t$ -mouvement brownien tout mouvement brownien adapté à $(\mathcal{F}_t)$ tel que pour tout $t \in \mathbb{R}_+$, $(W_{t+s} - W_t)_{s \in \mathbb{R}_+}$ est indépendant de $\mathcal{F}_t$.

Le théorème de Donsker assure l'existence du mouvement brownien.

1. à indistinguabilité près.

2. sauf mention contraire, toutes les chaînes de Markov seront supposées homogènes, on ne précisera donc plus qu'elles le sont.

4.2 Compléments sur les équations différentielles stochastiques et les générateurs

Dans tout ce chapitre, nous noterons $\mathbb{R}^{d \times d}$ l'espace des matrices carrées de taille d . On considérera $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ et $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ des applications mesurables. Par la suite, nous allons travailler avec des équations différentielles stochastiques (EDS) de la forme suivante : pour W un mouvement brownien standard de $\mathbb{R}^d$,

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dW_s, \quad t \geq 0, \quad (4.2)$$

ce qui peut se lire composante par composante de la façon suivante :

$$X_t^i = X_0^i + \int_0^t b_i(X_s) ds + \sum_{j=1}^d \int_0^t \sigma_{ij}(X_s) dW_s^j, \quad t \geq 0, \quad i \in \{1, \dots, d\}.$$

Définition 4.5 (Solution d'une EDS). *Une solution de l'EDS précédente consiste en la donnée des éléments suivants :*

- (i) *un espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ satisfaisant les conditions usuelles ;*
- (ii) *un $(\mathcal{F}_t)_{t \geq 0}$ -mouvement brownien $(W_t)_{t \geq 0}$ à valeurs dans $\mathbb{R}^d$;*
- (iii) *un processus continu $(X_t)_{t \geq 0}$, $(\mathcal{F}_t)_{t \geq 0}$ -adapté, vérifiant l'équation (4.2).*

Pour $X_0 = x \in \mathbb{R}^d$ fixé, on parle d'EDS partant de x et on note (E_x) le problème correspondant.

Pour (4.2), on a les notions suivantes d'existence et d'unicité.

Définition 4.6. *Pour l'équation (4.2), on dit qu'il y a :*

- ▷ *existence faible si pour tout $x \in \mathbb{R}^d$, il existe une solution au problème (E_x) ;*
- ▷ *unicité en loi si pour tout $x \in \mathbb{R}^d$, toutes les solutions de (E_x) ont la même loi ;*
- ▷ *unicité trajectorielle si sur un même espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ et pour un même $(\mathcal{F}_t)_{t \geq 0}$ -mouvement brownien W sur cet espace, deux solutions X et X' telles que $X_0 = X'_0$ p.s. sont indistinguables.*
- ▷ *Une solution X de (E_x) est dite solution forte si X est adapté à la filtration naturelle de W .*

Plusieurs résultats importants relient ces définitions. Sans entrer dans les détails, si toute solution est forte, alors il y a unicité trajectorielle pour l'EDS. Par ailleurs, s'il existe des solutions et qu'il y a unicité trajectorielle, alors il y a aussi unicité en loi. Dans ce cas, l'unique solution est une solution forte, c'est le *théorème de Yamada-Watanabe*. On pourra se référer au Chapitre IX de [15] pour plus de détails.

On rappelle que si les coefficients b et σ sont Lipschitz continus, alors il existe une solution forte à l'EDS qui est trajectoriellement unique.

Dans la preuve du théorème 4.13 (partie unicité), nous aurons besoin de la notion de *générateur infinitésimal* d'un processus de Markov. Nous nous restreignons ici à la définition pour une solution X de (4.2). Cette définition permettra également de motiver l'introduction des notions de générateur discret et de problèmes de martingales qui apparaîtront dans les deux prochaines sections.

Définition 4.7. Soit $\mathcal{C}_0(\mathbb{R}^d; \mathbb{R})$ l'espace des fonctions continues de $\mathbb{R}^d \rightarrow \mathbb{R}$ qui tendent vers 0 à l'infini.

Pour tout $x \in \mathbb{R}^d$, on note X^x une solution de (E_x) et on pose

$$\mathcal{D}(\mathcal{L}) = \left\{ f \in \mathcal{C}_0(\mathbb{R}^d; \mathbb{R}) : \frac{\mathbb{E}f(X_t^x) - f(x)}{t} \text{ converge uniformément lorsque } t \rightarrow 0 \right\}.$$

Pour tout $f \in \mathcal{D}(\mathcal{L})$, on définit alors le générateur infinitésimal $\mathcal{L}$ de X par

$$\mathcal{L}f(x) = \lim_{t \searrow 0} \frac{\mathbb{E}f(X_t^x) - f(x)}{t}.$$

Exercice 4.8.

- ▷ Déterminer le générateur $\mathcal{L}$ (et le domaine $\mathcal{D}(\mathcal{L})$) pour un mouvement brownien standard.
- ▷ On suppose que b et σ sont continus bornés et qu'il existe une solution faible à l'EDS (4.2), unique en loi. Déterminer le générateur $\mathcal{L}$ de X pour $f \in \mathcal{C}_c^2(\mathbb{R}^d; \mathbb{R})$ (les fonctions $\mathcal{C}^2$ à support compact).
- ▷ Sous les mêmes hypothèses, montrer que pour tout $f \in \mathcal{D}(\mathcal{L})$, le processus

$$\left\{ f(X_t) - \int_0^t \mathcal{L}f(X_s) ds \right\}_{t \geq 0}$$

est une $(\mathcal{F}_t)_{t \geq 0}$ -martingale locale.

Revenons désormais aux processus discrets, et considérons plus spécifiquement la marche aléatoire simple renormalisée, à savoir le processus $\widetilde{W}^{(n)}$ rencontré dans le chapitre précédent.

Pour ce processus, on peut définir le *générateur discret* $\mathcal{L}^{(n)}$ comme suit. Pour $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable bornée, $x \in \mathbb{R}^d$,

$$\mathcal{L}^{(n)}f(x) = \frac{\mathbb{E} \left[f(\widetilde{W}_{\frac{1}{n}}^{(n)}) - f(\widetilde{W}_0^{(n)}) \mid \widetilde{W}_0^{(n)} = x \right]}{1/n}.$$

En notant μ la loi des accroissements de S_k (l'accroissement de la marche aléatoire simple), un simple calcul permet d'écrire

$$\mathcal{L}^{(n)}f(x) = n \int_{\mathbb{R}^d} \left(f\left(x + \frac{y}{\sqrt{n}}\right) - f(x) \right) \mu(dy). \quad (4.3)$$

Exercice 4.9.

- ▷ Montrer que le processus $\{f(\widetilde{W}_{\frac{k}{n}}^{(n)}) - \frac{1}{n} \sum_{j=0}^{k-1} \mathcal{L}^{(n)}f(\widetilde{W}_{\frac{j}{n}}^{(n)})\}_{k \in \mathbb{N}}$ est une martingale pour la filtration $\mathcal{F}_k = \sigma(X_j, j \leq k)$.
- ▷ Calculer $\mathcal{L}^{(n)}f$ pour $f_1 : x \mapsto x$ et $f_2 : x \mapsto xx^T$.

En dimension $d = 1$ et pour une mesure μ d'espérance nulle et de variance 1, observons à partir de (4.3) que pour $f \in \mathcal{C}_0^2$,

$$\begin{aligned} \mathcal{L}^{(n)}f(x) &= nf'(x) \int_{\mathbb{R}} \frac{y}{\sqrt{n}} \mu(dy) + \frac{n}{2} f''(x) \int_{\mathbb{R}} \frac{y^2}{n} \mu(dy) + o_{n \rightarrow \infty}(1) \\ &= \frac{1}{2} f''(x) + o_{n \rightarrow \infty}(1). \end{aligned}$$

On observe que $\mathcal{L}^{(n)}f$ converge uniformément vers $\frac{1}{2}f'' = \mathcal{L}f$, avec $\mathcal{L}$ qui n'est autre que le générateur infinitésimal du mouvement brownien standard. On verra dans la preuve du théorème 4.13 que la convergence du générateur infinitésimal suffit à caractériser la loi limite de la diffusion, lorsque la chaîne de Markov renormalisée qu'on étudie est tendue. Cette approche fournit une preuve alternative à la preuve du théorème de Donsker vue précédemment, qui ne nécessite pas l'utilisation du TCL.

4.3 Problèmes de martingales

Les problèmes de martingales permettent entre autres de résoudre des équations différentielles stochastiques. Ils nous permettront également d'identifier les lois limites de chaînes de Markov.

Dans toute la suite du chapitre, on notera $a = \sigma \sigma^T$. Bien qu'on puisse formuler la définition de problème de martingale plus généralement à l'aide des générateurs, on se restreint ici à la formulation suivante, en lien avec les EDS dirigées par un mouvement brownien standard.

Définition 4.10 (Problème de martingale). *Un processus continu X à valeurs dans $\mathbb{R}^d$ et un espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ forment une solution du problème de martingale $\mathbf{M}(a, b)$ si les processus*

$$Y = \left(X_t - \int_0^t b(X_s) ds \right)_{t \in \mathbb{R}_+}$$

et

$$\left(Y_t Y_t^T - \int_0^t a(X_s) ds \right)_{t \in \mathbb{R}_+}$$

sont des $\mathcal{F}_t$ -martingales locales. Puisqu'on s'intéresse uniquement au processus X , on supposera que $\mathcal{F}_t = \sigma(X_s, s \leq t)$.

Exercice 4.11. *Montrer que si b et σ sont des coefficients Lipschitz, alors il existe une EDS dont la solution produit une solution au problème de martingale $\mathbf{M}(\sigma \sigma^T, b)$.*

La proposition suivante montre que la réciproque est également vraie. Toute solution d'un problème de martingale donne une solution (faible) d'EDS.

Proposition 4.12. *Soient b et σ des fonctions mesurables. On suppose qu'il existe $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ et X solution du problème $\mathbf{M}(a, b)$ avec $a = \sigma \sigma^T$. Alors il existe un $\mathcal{F}_t$ -mouvement brownien $(W_t)_{t \in \mathbb{R}_+}$, possiblement défini sur un espace de probabilité élargi, tel que X soit solution de l'EDS (4.2).*

Démonstration. On ne fait la preuve que dans le cas où pour tout $x \in \mathbb{R}^d$, la matrice $\sigma(x)$ est inversible, et $x \mapsto \sigma(x)^{-1}$ est localement bornée. Dans ce cas, il n'est pas nécessaire d'élargir l'espace de probabilité et la filtration. Pour le cas général, voir [15, Theorem 2.7, p.299]. En particulier, la nécessité d'élargir la filtration est expliquée après [15, Prop. 3.8, p.202-203].

Par hypothèse, $Y_t^i := X_t^i - \int_0^t b_i(X_s) ds$ et $Y_t^i Y_t^j - \int_0^t \sigma_{i,j}(X_s) ds$ sont des martingales locales, pour tous i, j . Donc par la proposition 4.2, $[Y^i, Y^j]_t = \int_0^t \sigma_{i,j}(X_s) ds$. Puisque le

processus X est continu et σ^{-1} est localement bornée, on a que le processus $\sigma(X)^{-1}$ est localement borné et prévisible pour la filtration $(\mathcal{F}_t)$. Ainsi on peut définir

$$W_t^i = \sum_{k=1}^d \int_0^t (\sigma^{-1})_{i,k}(X_s) dY_s^k, \quad t \geq 0,$$

qui est une martingale locale continue. Calculons la variation quadratique de W :

$$\begin{aligned} [W^i, W^j]_t &= \sum_{k,l=1}^d \int_0^t (\sigma^{-1})_{i,k}(X_s) (\sigma^{-1})_{j,l}(X_s) a_{k,l}(X_s) ds \\ &= \int_0^t \delta_{i,j} ds = \delta_{i,j} t. \end{aligned}$$

Par le théorème de caractérisation de Lévy³, W est un mouvement brownien standard à valeurs dans $\mathbb{R}^d$. On a alors, par la formule de composition des intégrales stochastiques $(H \cdot (K \cdot M) = (HK) \cdot M)$:

$$\int_0^t \sigma(X_s) dW_s = Y_t - Y_0 = X_t - \int_0^t b(X_s) ds.$$

Donc X est solution de l'EDS. □

On déduit notamment de la proposition précédente que s'il y a unicité faible de l'EDS (4.2) (i.e. toutes les solutions partant d'une même condition initiale ont même loi), alors il y a unicité dans le problème de martingale associé, au sens où pour tout $x \in \mathbb{R}^d$, deux solutions de $\mathbf{M}(a, b)$ partant de x ont la même loi (sur l'espace des fonctions continues). C'est par exemple le cas lorsque b et σ sont Lipschitz.

4.4 Résultat principal

Reprenons les notations du début de ce chapitre. On se donne une suite de chaînes de Markov $\{(Y_k^{(n)})_{k \in \mathbb{N}}, n \in \mathbb{N}\}$ et l'extension en temps continu $X^{(n)} \in \mathcal{D}(\mathbb{R}_+; \mathbb{R}^d)$, dilatée d'un facteur n (voir équation (4.1)). On appelle Q_n^0 la matrice de transition de $Y^{(n)}$ et on note

$$Q_n(x, A) = n Q_n^0(x, A), \quad \forall x \in \mathbb{R}^d, \forall A \in \mathcal{B}(\mathbb{R}^d).$$

Le générateur discret de X^n est ici donné par

$$\mathcal{L}^{(n)} f(x) = n \int_{\mathbb{R}^d} (f(y) - f(x)) Q_n^0(x, dy) = \int_{\mathbb{R}^d} (f(y) - f(x)) Q_n(x, dy). \quad (4.4)$$

On note $B(x, r)$ la boule euclidienne de centre x et de rayon $r > 0$. En évaluant l'expression précédente pour les fonctions $f_2(x) = x x^T$ et $f_1(x) = x$, on obtient les opérateurs (tronqués) de covariance et de moyenne associés à Q_n (et donc à $X^{(n)}$) :

$$\begin{aligned} a_{i,j}^{(n)}(x) &= \int_{B(x,1)} (y_i - x_i) (y_j - x_j) Q_n(x, dy), \quad \forall x \in \mathbb{R}^d, \\ b_i^{(n)}(x) &= \int_{B(x,1)} (y_i - x_i) Q_n(x, dy), \quad \forall x \in \mathbb{R}^d. \end{aligned}$$

3. on rappelle qu'une martingale locale $(M^1, \dots, M^d)$ qui est continue est un mouvement brownien standard à valeurs dans $\mathbb{R}^d$ si et seulement si $[M^i, M^j]_t = \delta_{i,j} t$.

On note enfin, pour tout $\varepsilon > 0$,

$$\Delta_\varepsilon^{(n)}(x) = Q_n(x, B(x, \varepsilon)^c), \quad x \in \mathbb{R}^d,$$

qui correspond à la probabilité de faire un saut de taille $\geq \varepsilon$.

Théorème 4.13. *Supposons qu'il existe des fonctions continues $a : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ et $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ et que les conditions suivantes sont vérifiées :*

- (i) *le problème de martingale $\mathbf{M}(a, b)$ admet une unique solution ;*
- (ii) *pour tous i, j , tout $R > 0$, $\lim_{n \rightarrow +\infty} \sup_{x \in B(0, R)} |a_{i,j}(x) - a_{i,j}^{(n)}(x)| = 0$;*
- (iii) *pour tout i , tout $R > 0$, $\lim_{n \rightarrow +\infty} \sup_{x \in B(0, R)} |b_i(x) - b_i^{(n)}(x)| = 0$;*
- (iv) *pour tous $R, \varepsilon > 0$, $\lim_{n \rightarrow +\infty} \sup_{x \in B(0, R)} \Delta_\varepsilon^{(n)}(x) = 0$;*
- (v) *la condition initiale $X_0^{(n)}$ est déterministe et tend vers x .*

Alors $X^{(n)} \xrightarrow{(d)} X$ dans l'espace $\mathcal{D}(\mathbb{R}_+; \mathbb{R}^d)$ ⁴, où X est la solution du problème $\mathbf{M}(a, b)$ partant de x .

Remarque 4.14. *Toutes les conditions sont assez naturelles et intuitives, et on insistera sur la condition (iv) qui assure que les sauts disparaissent à la limite. Cela nous permettra notamment d'utiliser le critère de tension du théorème 3.29.*

Le reste de cette section est consacré à la preuve de ce théorème.

Démonstration. Par un argument de localisation qu'on ne décrira pas (voir les détails dans [6, p.304]), on peut se ramener à des coefficients vérifiant :

- (ii') pour tous i, j , $\lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}^d} |a_{i,j}(x) - a_{i,j}^{(n)}(x)| = 0$;
- (iii') pour tout i , $\lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}^d} |b_i(x) - b_i^{(n)}(x)| = 0$;
- (iv') pour tout $\varepsilon > 0$, $\lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}^d} \Delta_\varepsilon^{(n)}(x) = 0$;
- (vi) les fonctions $a^{(n)}$, $b^{(n)}$ et $\Delta_\varepsilon^{(n)}$ sont uniformément bornées en n et en x .

Étape 1 : Tension.

L'objectif est de pouvoir appliquer le critère de tension du théorème 3.29. En particulier, il suffit de montrer la tension sur tout intervalle de temps de la forme $[0, T]$, $T > 0$. Sans perte de généralité, nous le ferons pour $T = 1$. Dans le cas présent, l'espace polonais E n'est autre que $\mathbb{R}^d$. En particulier ses boules fermées sont compactes, donc le premier point à vérifier est que pour tout $\varepsilon > 0$, il existe $A > 0$ et $n_0 \in \mathbb{N}$ tels que $\mathbb{P}(|X_0^{(n)}| > A) \leq \varepsilon$ pour tout $n \geq n_0$. C'est trivial ici puisqu'on a supposé que la suite des conditions initiales est déterministe et converge.

4. pour la topologie de Skorokhod. Cette convergence implique la convergence $\overline{X}^{(n)} \xrightarrow{(d)} X$ dans $\mathcal{C}$, où $\overline{X}^{(n)}$ sont les interpolées linéaires.

Nous allons donc démontrer le deuxième point, c'est-à-dire contrôler les oscillations de $X^{(n)}$. Pour cela, considérons les variables aléatoires suivantes : pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned}\tau_0 &= 0, \quad \tau_k^{(n)} = \inf\{t \geq \tau_{k-1}^{(n)} : |X_t^{(n)} - X_{\tau_{k-1}^{(n)}}^{(n)}| \geq \frac{\varepsilon}{4}\}, \\ N^{(n)} &= \min\{k : \tau_k^{(n)} > 1\}, \\ \Delta\tau^{(n)} &= \min\{\tau_k^{(n)} - \tau_{k-1}^{(n)} : 1 \leq k \leq N^{(n)}\}, \\ \Delta X^{(n)} &= \max\{|X_t^{(n)} - X_{t-}^{(n)}| : t \in (0, 1]\}.\end{aligned}$$

On va alors montrer la propriété suivante : pour tous $\varepsilon > 0$, $\delta > 0$,

$$\text{Si } \Delta\tau^{(n)} > \delta \text{ et } \Delta X^{(n)} \leq \frac{\varepsilon}{4}, \text{ alors } \omega(X^{(n)}, \delta) \leq \varepsilon. \quad (4.5)$$

Ainsi, il suffira de montrer qu'il existe $n_0 \in \mathbb{N}$ tel que

$$\mathbb{P}\left(\Delta X^{(n)} > \frac{\varepsilon}{4}\right) \leq \frac{\varepsilon}{2}, \quad \forall n \geq n_0, \quad (4.6)$$

et qu'il existe $\delta > 0$ tel que

$$\mathbb{P}\left(\Delta\tau^{(n)} \leq \delta\right) \leq \frac{\varepsilon}{2}, \quad \forall n \geq n_0, \quad (4.7)$$

pour obtenir grâce à (4.5), (4.6) et (4.7) que

$$\mathbb{P}\left(\omega(X^{(n)}, \delta) > \varepsilon\right) \leq \mathbb{P}\left(\Delta X^{(n)} > \frac{\varepsilon}{4}\right) + \mathbb{P}\left(\Delta\tau^{(n)} \leq \delta\right) \leq \varepsilon, \quad \forall n \geq n_0, \quad (4.8)$$

soit le critère nécessaire pour conclure que $(X^{(n)})$ est tendue grâce au théorème 3.29.

Montrons pour commencer la propriété (4.5). Soient $s, t \in \mathbb{R}_+$ tels que $|s - t| < \delta$, et donc $|s - t| < \Delta\tau^{(n)}$. Il y a donc deux possibilités (sans perte de généralité, on suppose que $s < t$) :

▷ il existe k tel que $\tau_{k-1} \leq s < t < \tau_k$. Alors

$$|X_s^{(n)} - X_t^{(n)}| \leq |X_s^{(n)} - X_{\tau_{k-1}}^{(n)}| + |X_t^{(n)} - X_{\tau_{k-1}}^{(n)}| < \frac{\varepsilon}{2};$$

▷ il existe k tel que $\tau_{k-1} \leq s < \tau_k \leq t < \tau_{k+1}$. Alors

$$|X_s^{(n)} - X_t^{(n)}| \leq |X_s^{(n)} - X_{\tau_{k-1}}^{(n)}| + |X_{\tau_{k-1}}^{(n)} - X_{\tau_k^-}^{(n)}| + |X_{\tau_k^-}^{(n)} - X_{\tau_k}^{(n)}| + |X_{\tau_k}^{(n)} - X_t^{(n)}| \leq \varepsilon.$$

Montrons maintenant (4.6), ce qui assez immédiat. En effet, le processus $X^{(n)}$ a au plus n sauts sur l'intervalle de temps $[0, 1]$, donc

$$\begin{aligned}\mathbb{P}\left(\Delta X^{(n)} > \frac{\varepsilon}{4}\right) &\leq \sum_{|X_t^{(n)} - X_{t-}^{(n)}| > \frac{\varepsilon}{4}} \mathbb{P}\left(|X_t^{(n)} - X_{t-}^{(n)}| > \frac{\varepsilon}{4}\right) \\ &\leq n \sup_{x \in \mathbb{R}^d} Q_n^0\left(x, B(x, \frac{\varepsilon}{4})^c\right) = \sup_{x \in \mathbb{R}^d} \Delta_{\frac{\varepsilon}{4}}^{(n)}(x) \xrightarrow{n \rightarrow +\infty} 0,\end{aligned}$$

grâce à l'hypothèse (iv') .

Enfin, montrons (4.7), ce qui va nous demander significativement plus d'efforts. Commençons par observer que pour tout $k \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}_x \left(\Delta \tau^{(n)} \leq \delta \right) &\leq \mathbb{P}_x \left(\text{il existe } 1 \leq j \leq k \text{ tel que } \tau_j^{(n)} - \tau_{j-1}^{(n)} \leq \delta \text{ ou } N^{(n)} > k \right) \\ &\leq \sum_{j=1}^k \mathbb{P}_x (\tau_j^{(n)} - \tau_{j-1}^{(n)} \leq \delta) + \mathbb{P}_x (N^{(n)} > k) \\ &\leq k \sup_{y \in \mathbb{R}^d} \mathbb{P}_y (\tau_1^{(n)} \leq \delta) + \mathbb{P}_x (N^{(n)} > k), \end{aligned} \quad (4.9)$$

en utilisant la propriété de Markov forte à la dernière ligne.

On va donc chercher dans un premier temps à estimer $\mathbb{P}(\tau_1^{(n)} \leq \eta)$, $\eta > 0$. Pour le générateur défini en (4.4), on montre, comme à l'exercice 4.9, on a que pour tout n , le processus

$$\left\{ f(X_{k/n}^{(n)}) - \frac{1}{n} \sum_{j=0}^{k-1} \mathcal{L}^{(n)} f(X_{j/n}^{(n)}) \right\}_{k \in \mathbb{N}} \quad (4.10)$$

est une martingale pour la filtration naturelle de $\{Y_k^{(n)}\}_{k \in \mathbb{N}}$. Par ailleurs, si $\phi : \mathbb{R}_+ \rightarrow [0, 1]$ est une fonction de classe $\mathcal{C}^2$ telle que $\phi(0) = 1$ et $\phi(x) = 0, \forall x \geq 1$, on définit pour tout $\varepsilon > 0$:

$$f_{\varepsilon, y}(x) = \phi \left(\frac{|x - y|^2}{\varepsilon^2} \right), \quad x, y \in \mathbb{R}^d.$$

Alors on a qu'il existe C_ε qui ne dépend que de ε tel que

$$\sup_{x \in \mathbb{R}^d} |\mathcal{L}^{(n)} f_{\varepsilon, y}(x)| \leq C_\varepsilon. \quad (4.11)$$

La preuve de (4.11) est purement analytique et repose sur l'hypothèse (vi). Elle est laissée en **exercice**.

En conséquence des propriétés (4.10) et (4.11), on a que pour tout n ,

$$\left\{ f_{\frac{\varepsilon}{4}, y}(X_{k/n}^{(n)}) + \frac{k}{n} C_{\frac{\varepsilon}{4}} \right\}_{k \in \mathbb{N}} \text{ est une sous-martingale.}$$

On notera $\tau_1 \equiv \tau_1^{(n)}$. On remarque que τ_1 est un multiple (aléatoire) de $\frac{1}{n}$ et donc pour $\tilde{\eta} = \frac{1}{n} \lfloor n\eta \rfloor$, on a $X_{\tau_1 \wedge \tilde{\eta}}^{(n)} = X_{\tau_1 \wedge \eta}^{(n)}$. Ainsi, d'après le théorème d'arrêt,

$$\begin{aligned} \mathbb{E}_y \left(f_{y, \frac{\varepsilon}{4}}(X_{\tau_1 \wedge \eta}^{(n)}) + (\tau_1 \wedge \eta) C_{\frac{\varepsilon}{4}} \right) &= \mathbb{E}_y \left(f_{y, \frac{\varepsilon}{4}}(X_{\tau_1 \wedge \tilde{\eta}}^{(n)}) + (\tau_1 \wedge \eta) C_{\frac{\varepsilon}{4}} \right) \\ &\geq \mathbb{E}_y \left(f_{y, \frac{\varepsilon}{4}}(X_{\tau_1 \wedge \tilde{\eta}}^{(n)}) + (\tau_1 \wedge \tilde{\eta}) C_{\frac{\varepsilon}{4}} \right) \\ &\geq \mathbb{E}_y \left(f_{y, \frac{\varepsilon}{4}}(X_0^{(n)}) + \frac{0}{n} C_{\frac{\varepsilon}{4}} \right) = 1. \end{aligned}$$

On en déduit d'une part que

$$\mathbb{E}_y \left(1 - f_{y, \frac{\varepsilon}{4}}(X_{\tau_1 \wedge \eta}^{(n)}) \right) \leq \mathbb{E}_y \left((\tau_1 \wedge \eta) C_{\frac{\varepsilon}{4}} \right) \leq \eta C_{\frac{\varepsilon}{4}}.$$

D'autre part, on remarque que si $\tau_1 \leq \eta$ alors $|X_{\tau_1 \wedge \eta}^{(n)} - X_0^{(n)}| = |X_{\tau_1}^{(n)} - X_0^{(n)}| \geq \frac{\varepsilon}{4}$ et donc sous $\mathbb{P}_y$, $f_{y, \frac{\varepsilon}{4}}(X_{\tau_1 \wedge \eta}^{(n)}) = 0$. Ainsi

$$\mathbb{P}_y (\tau_1 \leq \eta) = \mathbb{E}_y \left(\mathbb{1}_{\{\tau_1 \leq \eta\}} (1 - f_{y, \frac{\varepsilon}{4}}(X_{\tau_1 \wedge \eta}^{(n)})) \right) \leq \eta C_{\frac{\varepsilon}{4}}. \quad (4.12)$$

Revenons à l'équation (4.9) : il nous reste à contrôler $\mathbb{P}_x(N^{(n)} > k)$. En utilisant l'inégalité de Markov, il vient

$$\mathbb{P}_x(N^{(n)} > k) = \mathbb{P}_x(\tau_k^{(n)} \leq 1) \leq e^{-1} \mathbb{E}_x(e^{-\tau_k^{(n)}}). \quad (4.13)$$

On a

$$\begin{aligned} \mathbb{E}_x(e^{-\tau_1}) &\leq \mathbb{P}_x(\tau_1 \leq \eta) + e^{-\eta} \mathbb{P}_x(\tau_1 > \eta) \\ &\leq \eta C_{\frac{\varepsilon}{4}} + e^{-\eta} (1 - \eta C_{\frac{\varepsilon}{4}}) =: p_\eta < 1. \end{aligned}$$

Remarquons particulièrement que p_η ne dépend pas de x . Les temps $\tau_1^{(n)}, \dots, \tau_k^{(n)}$ étant des temps d'arrêt (qu'on note ci-dessous $\tau_1, \dots, \tau_k$), on a en utilisant la propriété de Markov forte que

$$\begin{aligned} \mathbb{E}_x(e^{-\tau_k}) &= \mathbb{E}_x(\mathbb{E}_x[e^{-\tau_k} | \mathcal{F}_{\tau_{k-1}}]) = \mathbb{E}_x(e^{-\tau_{k-1}} \mathbb{E}_x[e^{-(\tau_k - \tau_{k-1})} | \mathcal{F}_{\tau_{k-1}}]) \\ &= \mathbb{E}_x(e^{-\tau_{k-1}} \mathbb{E}_{X_{\tau_{k-1}}}[e^{-\tau_1}]) \\ &\leq p_\eta \mathbb{E}_x(e^{-\tau_{k-1}}), \end{aligned}$$

et donc en itérant, on obtient

$$\mathbb{E}_x(e^{-\tau_k}) \leq p_\eta^k.$$

En utilisant l'inégalité précédente dans (4.13), on obtient donc

$$\mathbb{P}_x(N^{(n)} > k) \leq e^{-1} p_\eta^k. \quad (4.14)$$

Ainsi, en utilisant (4.12) et (4.14) dans l'équation (4.9), on a

$$\mathbb{P}_x(\Delta\tau^{(n)} \leq \delta) \leq C_{\frac{\varepsilon}{4}} k \delta + e^{-1} p_\eta^k,$$

($\eta \in (0, 1)$ est quelconque, indépendant de δ). Ainsi en choisissant k suffisamment grand de sorte que $e^{-1} p_\eta^k \leq \frac{\varepsilon}{4}$ et ensuite δ suffisamment petit de sorte que $C_{\frac{\varepsilon}{4}} k \delta \leq \frac{\varepsilon}{4}$, on obtient bien le résultat souhaité (4.7).

Étape 2 : Unicité de la limite pour les sous-suites convergentes.

Montrons désormais que toute sous-suite qui converge, converge vers la solution du problème de martingale $\mathbf{M}(a, b)$ (qui est unique par hypothèse). On va faire appel pour cela au générateur infinitésimal $\mathcal{L}$ de l'EDS (4.2). On sait (voir exercice 4.8) que $\mathcal{L}$ est l'opérateur différentiel associé à $\mathbf{M}(a, b)$ donné pour toute $f \in \mathcal{C}_0^2(\mathbb{R}^d)$ par

$$\mathcal{L}f(x) = \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(x) \partial_i \partial_j f(x) + \sum_{i=1}^d b_i(x) \partial_i f(x), \quad \forall x \in \mathbb{R}^d.$$

On a la convergence suivante (**exercice**, on utilise ici les hypothèses (ii'), (iii') et (iv')) : pour toute fonction $f \in \mathcal{C}_0^2(\mathbb{R}^d)$,

$$\|\mathcal{L}^{(n)} f - \mathcal{L}f\|_\infty \xrightarrow{n \rightarrow +\infty} 0. \quad (4.15)$$

Quitte à renommer la suite, on suppose que $X^{(n)}$ converge en loi dans $\mathcal{D}(\mathbb{R}_+; \mathbb{R}^d)$ vers un processus qu'on note X .

Soient $s < t \in \mathbb{R}_+$, et notons $k_n = \lfloor sn \rfloor + 1$ et $l_n = \lfloor tn \rfloor + 1$. Soit $F : \mathcal{D}(\mathbb{R}_+; \mathbb{R}^d) \rightarrow \mathbb{R}$ une application bornée et mesurable par rapport à la tribu $\mathbf{F}_s := \sigma(\pi_u^{-1}(A); A \in \mathcal{B}(\mathbb{R}^d), u \leq s)$. En utilisant la propriété de martingale (discrète) énoncée en (4.10), on a pour $f \in \mathcal{C}_0^2(\mathbb{R}^d)$,

$$\mathbb{E} \left[F(X^{(n)}) \left(f(X_{l_n/n}^{(n)}) - f(X_{k_n/n}^{(n)}) - \frac{1}{n} \sum_{j=k_n}^{l_n-1} \mathcal{L}^{(n)} f(X_{j/n}^{(n)}) \right) \right] = 0. \quad (4.16)$$

Utilisons maintenant le théorème de représentation de Skorokhod (théorème 2.33) : il existe un espace de probabilité et des variables aléatoires $Z^{(n)}$ définies sur cet espace et à valeurs dans $\mathcal{D}(\mathbb{R}_+; \mathbb{R}^d)$ telles que pour tout n , $Z^{(n)} \stackrel{(d)}{=} X^{(n)}$ et $Z^{(n)} \rightarrow X$ p.s. Alors par le théorème de convergence dominée, on déduit de (4.16) appliqué à $Z^{(n)}$, de la convergence uniforme de $\mathcal{L}^{(n)} f$ (éq. (4.15)) et de la convergence des sommes de Riemann (remarquez que $\mathcal{L}^{(n)} f$ est une suite de fonctions continues bornées) que

$$\mathbb{E} \left[F(X) \left(f(X_t) - f(X_s) - \int_s^t \mathcal{L} f(X_u) du \right) \right] = 0.$$

Puisque cette égalité est valable pour toute fonction F qui est $\mathbf{F}_s$ -mesurable bornée, on en déduit que $\{f(X_t) - \int_0^t \mathcal{L} f(X_u) du\}_{t \in \mathbb{R}_+}$ est une martingale. Il ne reste plus qu'à appliquer cette propriété aux fonctions $f_1(x) = x_i$ et $f_2(x) = x_i x_j$ pour obtenir que X est solution du problème de martingale $\mathbf{M}(a, b)$. $\square$

Exercice 4.15. Montrer l'inégalité (4.11) et la convergence (4.15).

Exercice 4.16. A la fin de la preuve, on a que $\{f(X_t) - \int_0^t \mathcal{L} f(X_u) du\}_{t \in \mathbb{R}_+}$ est une martingale pour tout $f \in \mathcal{C}_0^2(\mathbb{R}^d)$. Or on applique ce résultat aux fonctions f_1 et f_2 qui ne tendent pas vers 0 à l'infini. En approchant f_1 et f_2 par des fonctions qui tendent vers 0 à l'infini, justifier que $\{f_1(X_t) - \int_0^t \mathcal{L} f_1(X_u) du\}_{t \in \mathbb{R}_+}$ et $\{f_2(X_t) - \int_0^t \mathcal{L} f_2(X_u) du\}_{t \in \mathbb{R}_+}$ sont des martingales locales.

4.5 L'exemple de l'urne d'Ehrenfest

Reprenons l'exemple de l'urne d'Ehrenfest présenté au début de ce cours dans la section 1.1.2. Pour tout $n \in \mathbb{N}^*$, on dispose d'une chaîne de Markov homogène $(Y_k^n)_{k \in \mathbb{N}}$ à valeurs dans $\{0, \dots, n\}$ qui donne le nombre de molécules dans l'urne A au temps (discret) k . On définit alors la chaîne renormalisée $Y_k^{(n)} = \frac{1}{\sqrt{n/2}}(Y_k^n - \frac{n}{2})$ dont la matrice de transition est donnée, pour tout $x \in \{\frac{j}{\sqrt{2n}}; -n \leq j \leq n\}$, par

$$Q_n^0 \left(x, x + \frac{1}{\sqrt{n/2}} \right) = \frac{n - x\sqrt{2n}}{2n} \quad \text{et} \quad Q_n^0 \left(x, x - \frac{1}{\sqrt{n/2}} \right) = \frac{n + x\sqrt{2n}}{2n}.$$

A nouveau, on définit le processus càdlàg $X^{(n)}$ par $X_t^{(n)} = Y_{\lfloor nt \rfloor}^{(n)}$, pour tout $t \geq 0$.

Théorème 4.17. La suite de processus $(X^{(n)})_{n \in \mathbb{N}}$ converge en loi dans $\mathcal{D}(\mathbb{R}_+; \mathbb{R})$ vers un processus d'Ornstein-Uhlenbeck, solution de l'EDS :

$$dX_t = -X_t dt + dW_t.$$

Démonstration. Remarquons que $\Delta_\varepsilon^{(n)}(x) = n Q_n^0(x, B(x, \varepsilon)^c) = 0$ dès que $n > \varepsilon^{-2}$. Ensuite, on définit les coefficients de drift et variance infinitésimaux (non tronqués) :

$$\begin{aligned}\tilde{a}^{(n)}(x) &= \int_{\mathbb{R}} (x-y)^2 n Q_n^0(x, dy) = \frac{n}{2} \left(\frac{2}{n} \frac{n-x\sqrt{2n}}{2n} + \frac{2}{n} \frac{n+x\sqrt{2n}}{2n} \right) = 1 \\ \tilde{b}^{(n)}(x) &= \int_{\mathbb{R}} (x-y) n Q_n^0(x, dy) = \frac{n}{2} \left(\frac{1}{\sqrt{n/2}} \frac{n-x\sqrt{2n}}{2n} - \frac{1}{\sqrt{n/2}} \frac{n+x\sqrt{2n}}{2n} \right) = -x,\end{aligned}$$

et on vérifie (exercice) que les coefficients tronqués a^n et b^n convergent uniformément sur les compacts vers, respectivement, $a(x) = 1$ et $b(x) = -x$. Ces deux fonctions sont Lipschitz, donc le problème de martingale $\mathbf{M}(a, b)$ est bien posé.

On conclut donc grâce au théorème 4.13. □

Chapitre 5

Particules diffusives en interaction et propagation du chaos

Ce dernier chapitre présente quelques propriétés des systèmes d'EDS en interaction, couramment appelés systèmes de *particules* en interaction. Notons qu'il s'agit bien ici de particules stochastiques, le pendant déterministe faisant également l'objet d'intenses recherches. La plupart des notions exposées dans les pages qui suivent (limite champ moyen, propagation du chaos) sont développées dans le cours de recherche de Sznitman [19]. On trouvera également d'intéressants développements dans les notes de Méléard [12] pour des particules admettant des sauts, avec des applications à l'équation de Boltzmann.

5.1 Couplage entre le système de particules et l'équation de McKean-Vlasov

Considérons le système de particules suivant :

$$dX_t^i = \frac{1}{N} \sum_{j=1}^N b(X_t^i, X_t^j) dt + dW_t^i, \quad i = 1, \dots, N, \quad t > 0, \quad (5.1)$$

où les $(W^i)_{i \in \{1, \dots, N\}}$ sont des mouvements browniens standard à valeurs dans $\mathbb{R}^d$ qui sont indépendants, et les particules ont pour condition initiale des variables aléatoires $(X_0^i)_{i \in \{1, \dots, N\}}$ supposées pour l'instant quelconques.

Formulons quelques remarques :

- ▷ Hormis dans la section 5.3, on supposera tout le temps que $b : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ est une fonction Lipschitz bornée. Cette fonction modélise à la fois les interactions entre particules et une possible dérive de X^i en fonction de la position des autres particules. Par exemple, $b(x, y) = V(x) - (x - y) \wedge 1$ avec V une fonction Lipschitz bornée (ici $(x - y) \wedge 1$ est le terme d'interaction).
- ▷ On ne traite pas du cas où le coefficient b dépend de i (peut mener à quelques complications intéressantes) ni de coefficients de diffusion non-constants pour simplifier les calculs. Dans le cas de particules interagissant via la diffusion, les résultats qui suivent restent globalement vrais. On pourra se référer à [12] pour un traitement plus global.

Avant d'énoncer un théorème de convergence, introduisons la particule “typique” qui sera la candidate à la limite en loi de chaque particule du système.

Théorème 5.1. Soit $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ un espace de probabilité filtré et $(W_t)_{t \in \mathbb{R}_+}$ un $\mathcal{F}_t$ -mouvement brownien. Soit ξ une v.a. $\mathcal{F}_0$ -mesurable. Alors l'équation :

$$\begin{aligned} dX_t &= \left(\int_{\mathbb{R}^d} b(X_t, y) \mu_t(dy) \right) dt + dW_t \\ X_0 &= \xi, \quad X_t \stackrel{(d)}{=} \mu_t, \end{aligned} \tag{5.2}$$

admet une solution qui est unique trajectoriellement et en loi.

Avant de commencer la preuve de ce théorème, rappelons la définition de la distance de Wasserstein, qui fait l'objet d'une étude dans le DM. Pour tout $T > 0$, on utilise ici la distance de Wasserstein-1 sur l'espace métrique complet et séparable $\mathcal{C}_T \equiv \mathcal{C}([0, T]; \mathbb{R}^d)$ muni de la topologie de la convergence uniforme. Plutôt que la distance issue de la norme uniforme, on travaillera avec la distance topologiquement équivalente suivante :

$$d_T(x, y) = \left(\sup_{s \in [0, T]} \|x(s) - y(s)\| \right) \wedge 1.$$

La distance de Wasserstein-1 sur $\mathcal{C}_T$ est alors donnée par

$$\mathcal{W}_T(\mu, \nu) = \inf_{\pi \in \Pi_T(\mu, \nu)} \int_{\mathcal{C}_T \times \mathcal{C}_T} d_T(x, y) \pi(dx, dy),$$

où $\Pi_T(\mu, \nu)$ est l'ensemble des couplages entre $\mu \in \mathcal{P}(\mathcal{C}_T)$ et $\nu \in \mathcal{P}(\mathcal{C}_T)$, i.e. $\pi \in \Pi_T(\mu, \nu)$ si la première marginale de π est μ et la seconde est ν . Les résultats du DM tiennent toujours pour $\mathcal{W}_T$, à savoir que $\mathcal{W}_T$ est une distance sur $\mathcal{P}(\mathcal{C}_T)$ qui rend cet espace complet et séparable, et qui implique la convergence étroite.

Démonstration. On va introduire une transformation Φ sur l'espace $\mathcal{P}(\mathcal{C}_T)$ qui est Lipschitz continue et dont les itérées sont contractantes. Grâce à Φ , on montrera l'existence d'une solution à l'équation (5.2) par un argument de point fixe/contraction, et l'unicité grâce au lemme de Grönwall.

Soit $T > 0$. Pour toute mesure $\nu \in \mathcal{P}(\mathcal{C}_T)$ (on rappelle que ν_t est la marginale au temps t), on considère l'équation

$$X_t^\nu = X_0 + \int_0^t \left(\int_{\mathbb{R}^d} b(X_s^\nu, y) \nu_s(dy) \right) ds + W_t. \tag{5.3}$$

Par un argument de point fixe (b étant Lipschitz), on montre (exercice) que cette équation admet une solution forte, et par une application du lemme de Grönwall, cette solution est trajectoriellement unique. Il y a donc aussi unicité en loi¹, et on note $\mathbb{P}^\nu$ la loi de cette solution sur $\mathcal{C}_T$. Considérons l'application

$$\Phi : \begin{array}{ccc} \mathcal{P}(\mathcal{C}_T) & \longrightarrow & \mathcal{P}(\mathcal{C}_T) \\ \nu & \longmapsto & \mathbb{P}^\nu \end{array}.$$

Remarquons que si X est solution de (5.2), alors sa loi est un point fixe de Φ . Réciproquement, si ν est un point fixe de Φ , la solution X^ν de (5.3) est solution de (5.2).

1. par un résultat classique mais non trivial de Yamada et Watanabe, voir e.g. [15, Th. 1.7, p.368].

Montrons que Φ est Lipschitz pour la distance de Wasserstein. Considérons donc deux mesures ν^1, ν^2 , et X^{ν^1}, X^{ν^2} les deux processus associés solutions de l'équation (5.3) avec le même brownien W . Alors pour tout $t \in [0, T]$, on a

$$\begin{aligned} \sup_{s \leq t} |X_s^{\nu^1} - X_s^{\nu^2}| &\leq \int_0^t \left| \int_{\mathbb{R}^d} b(X_u^{\nu^1}, y) \nu_u^1(dy) - \int_{\mathbb{R}^d} b(X_u^{\nu^2}, z) \nu_u^2(dz) \right| du \\ &= \int_0^t \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} (b(X_u^{\nu^1}, y) - b(X_u^{\nu^2}, z)) \nu_u(dy, dz) \right| du \end{aligned}$$

où ν est un couplage quelconque de ν^1, ν^2 . Par hypothèse sur b , il existe M tel que

$$|b(x_1, y) - b(x_2, z)| \leq M(1 \wedge |x_1 - x_2| + 1 \wedge |y - z|).$$

Ainsi,

$$\sup_{s \leq t} |X_s^{\nu^1} - X_s^{\nu^2}| \leq M \int_0^t \left(|X_u^{\nu^1} - X_u^{\nu^2}| \wedge 1 + \int_{\mathbb{R}^d} (|y - z| \wedge 1) \nu_u(dy, dz) \right) du,$$

et donc

$$\sup_{s \leq t} |X_s^{\nu^1} - X_s^{\nu^2}| \wedge 1 \leq M \int_0^t \left(\sup_{s \leq u} |X_s^{\nu^1} - X_s^{\nu^2}| \wedge 1 \right) du + M \int_0^t \mathcal{W}_u(\nu^1, \nu^2) du.$$

Donc par le lemme de Grönwall,

$$\sup_{s \leq t} |X_s^{\nu^1} - X_s^{\nu^2}| \wedge 1 \leq M e^{tM} \int_0^t \mathcal{W}_u(\nu^1, \nu^2) du.$$

Or $\mathcal{W}_t(\Phi(\nu^1), \Phi(\nu^2)) \leq \mathbb{E}[\sup_{s \leq t} |X_s^{\nu^1} - X_s^{\nu^2}| \wedge 1]$, puisque la loi du couple (X^{ν^1}, X^{ν^2}) fournit un couplage (dit “parallèle”) entre ν^1 et ν^2 . On obtient donc finalement, pour tout $t \in [0, T]$:

$$\mathcal{W}_t(\Phi(\nu^1), \Phi(\nu^2)) \leq M e^{MT} \int_0^t \mathcal{W}_u(\nu^1, \nu^2) du, \quad \forall \nu^1, \nu^2 \in \mathcal{P}(\mathcal{C}_T). \quad (5.4)$$

Notons $\Phi^k, k \in \mathbb{N}^*$, les itérées de Φ . Pour une mesure $\nu \in \mathcal{P}(\mathcal{C}_T)$, montrons que la suite $(\Phi^k(\nu))_{k \in \mathbb{N}^*}$ est de Cauchy pour la distance de Wasserstein. D'après (5.4),

$$\begin{aligned} \mathcal{W}_T(\Phi^{k+1}(\nu), \Phi^k(\nu)) &\leq M e^{MT} \int_0^T \mathcal{W}_{u_1}(\Phi^k(\nu), \Phi^{k-1}(\nu)) du_1 \\ &\leq M^k e^{kMT} \int_0^T \int_0^{u_1} \dots \int_0^{u_{k-1}} \mathcal{W}_{u_k}(\Phi(\nu), \nu) du_k \dots du_1 \\ &\leq \frac{(MT)^k e^{kMT}}{k!} \mathcal{W}_T(\Phi(\nu), \nu). \end{aligned}$$

Donc $(\Phi^k(\nu))_{k \in \mathbb{N}^*}$ est bien de Cauchy, et converge puisque $\mathcal{P}(\mathcal{C}_T)$ est complet. La limite fournit une mesure qui, placée dans l'équation (5.2), donne une solution forte à l'équation (5.2). L'unicité trajectorielle (et donc en loi) s'obtient par un calcul direct (**exercice**) grâce au caractère Lipschitz de b .

Notons $\mu^{(T)}$ la loi de la solution de (5.2) construite sur $[0, T]$. Si $t \in (0, T)$, alors la solution construite sur $[0, T]$ donne également la solution sur $[0, t]$, donc $\mu^{(t)}$ est la restriction de $\mu^{(T)}$ à $[0, t]$, et la famille de mesure $(\mu^{(T)})_{T \in \mathbb{R}_+}$ est consistante. Il existe donc une mesure μ sur $\mathcal{C}(\mathbb{R}_+; \mathbb{R}^d)$ qui fournit la solution de (5.2) sur tout $\mathbb{R}_+$. $\square$

Exercice 5.2. Montrer que les marginales $(\mu_t)_{t \in \mathbb{R}_+}$ de la solution de l'équation de McKean-Vlasov (5.2) satisfont au sens faible l'EDP non-linéaire suivante :

$$\begin{aligned} \partial_t \mu_t &= \frac{1}{2} \Delta \mu_t - \nabla \cdot \left(\int_{\mathbb{R}^d} b(\cdot, y) \mu_t(dy) \mu_t \right), \quad t > 0, \\ \mu_0 &= \mathcal{L}(\xi_0). \end{aligned} \quad (5.5)$$

Sur l'espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ muni d'une famille de $\mathcal{F}_t$ -mouvements browniens indépendants $(W^i)_{i \in \mathbb{N}}$, on considère pour tout $N \in \mathbb{N}^*$ l'unique solution forte $(X^i)_{1 \leq i \leq N}$ du système de particules (5.1) (**exercice : justifier cette assertion concernant $(X^i)_{1 \leq i \leq N}$**). Sur le même espace, on considère également les processus $(\bar{X}^i)_{1 \leq i \leq N}$ définis par

$$\begin{aligned} \bar{X}_t^i &= X_0^i + \int_0^t \left(\int_{\mathbb{R}^d} b(\bar{X}_s^i, y) \mu_s(dy) \right) ds + W_t^i, \\ \bar{X}_t^i &\stackrel{(d)}{=} \mu_t, \end{aligned}$$

dont l'existence et l'unicité sont justifiées par le théorème 5.1. On souhaite comparer les particules $(X^{i,N})_{i \leq N}$ aux particules McKean-Vlasov $(\bar{X}^i)_{i \leq N}$. Pour ce faire, il semble raisonnable, au moins pour une comparaison trajectorielle, de faire partir $X^{i,N}$ et $\bar{X}^i$ du même point de départ X_0^i et de les doter du même brownien W^i . On observe que si les variables $(X_0^i)_{i \leq N}$ sont indépendantes, alors les processus $(\bar{X}^i)_{i \leq N}$ sont indépendants, mais pas les particules $(X^{i,N})_{i \leq N}$. En fait sous ces conditions, ces dernières sont échangeables, dans le sens donné par la définition suivante.

Définition 5.3. On dit que les variables aléatoires $(X_0^i)_{i \leq N}$ sont échangeables si pour toute permutation $\sigma : \{1, \dots, N\} \rightarrow \{1, \dots, N\}$ on a

$$(X_0^1, \dots, X_0^N) \stackrel{(d)}{=} (X_0^{\sigma(1)}, \dots, X_0^{\sigma(N)}).$$

Par exemple, si $(X_0^i)_{i \leq N}$ est une famille de v.a. indépendantes et identiquement distribuées, alors les $(X_0^i)_{i \leq N}$ sont échangeables.

Exercice 5.4. En supposant que les particules données par le système (5.1) sont initialement échangeables, montrer qu'elles le restent en tout temps $t > 0$.

Ainsi, cette propriété d'échangeabilité est préservée par la dynamique du système de particules, contrairement à la propriété d'indépendance.

Théorème 5.5. On suppose que les hypothèses formulées précédemment sont vérifiées, en particulier que pour tout $N \in \mathbb{N}^*$, $(X_0^i)_{i \leq N}$ est une famille échangeable et que les particules du système (5.1) et celles données par l'équation de McKean-Vlasov sont dirigées par les mêmes browniens et ont les mêmes CI $(X_0^i)_{i \leq N}$.

Pour tout $T > 0$, il existe $C_T \in (0, +\infty)$ tel que

$$\mathbb{E} \left(\sup_{s \in [0, T]} |X_s^{i,N} - \bar{X}_s^i| \right) \leq \frac{C_T}{\sqrt{N}}, \quad \forall N \in \mathbb{N}^*, \quad \forall i \in \{1, \dots, N\}.$$

Démonstration.

$$\begin{aligned}
X_t^i - \bar{X}_t^i &= \int_0^t \left(\frac{1}{N} \sum_{j=1}^N b(X_s^i, X_s^j) - \int_{\mathbb{R}^d} b(\bar{X}_s^i, y) \mu_s^i(dy) \right) ds \\
&= \int_0^t \frac{1}{N} \sum_{j=1}^N \left(b(X_s^i, X_s^j) - b(\bar{X}_s^i, X_s^j) \right. \\
&\quad \left. + b(\bar{X}_s^i, X_s^j) - b(\bar{X}_s^i, \bar{X}_s^j) \right. \\
&\quad \left. + b(\bar{X}_s^i, \bar{X}_s^j) - \int_{\mathbb{R}^d} b(\bar{X}_s^i, y) \mu_s^i(dy) \right) ds.
\end{aligned}$$

On adopte la notation $b_s(x, y) = b(x, y) - \int_{\mathbb{R}^d} b(x, z) \mu_s(dz)$, de sorte qu'avec b Lipschitz et la propriété d'échangeabilité $(X_s^i, \bar{X}_s^i) \stackrel{(d)}{=} (X_s^j, \bar{X}_s^j)$, on obtient

$$\begin{aligned}
\mathbb{E} \left(\sup_{s \in [0, T]} |X_s^i - \bar{X}_s^i| \right) &\leq M \int_0^T \left(\mathbb{E}(|X_s^i - \bar{X}_s^i|) + \frac{1}{N} \sum_{j=1}^N \mathbb{E}(|X_s^j - \bar{X}_s^j|) + \mathbb{E} \left(\left| \frac{1}{N} \sum_{j=1}^N b_s(\bar{X}_s^i, \bar{X}_s^j) \right| \right) \right) ds \\
&\leq M \int_0^T \left(2\mathbb{E}(|X_s^i - \bar{X}_s^i|) + \mathbb{E} \left(\left| \frac{1}{N} \sum_{j=1}^N b_s(\bar{X}_s^i, \bar{X}_s^j) \right| \right) \right) ds.
\end{aligned}$$

Donc en appliquant le lemme de Grönwall, on trouve :

$$\mathbb{E} \left(\sup_{s \in [0, T]} |X_s^i - \bar{X}_s^i| \right) \leq 2Me^{2MT} \int_0^T \mathbb{E} \left(\left| \frac{1}{N} \sum_{j=1}^N b_s(\bar{X}_s^i, \bar{X}_s^j) \right| \right) ds.$$

Par l'inégalité de Cauchy-Schwarz,

$$\begin{aligned}
\left(\mathbb{E} \left| \frac{1}{N} \sum_{j=1}^N b_s(\bar{X}_s^i, \bar{X}_s^j) \right| \right)^2 &\leq \mathbb{E} \left(\left| \frac{1}{N} \sum_{j=1}^N b_s(\bar{X}_s^i, \bar{X}_s^j) \right|^2 \right) \\
&= \frac{1}{N^2} \sum_{j,k=1}^N \mathbb{E} \left(b_s(\bar{X}_s^i, \bar{X}_s^j) \cdot b_s(\bar{X}_s^i, \bar{X}_s^k) \right).
\end{aligned}$$

A partir d'ici, on suppose que les $(X_0^i)_{i \leq N}$ sont iid (voir exercice 5.6 pour les hypothèses générales). On calcule alors, pour $j \neq k$,

$$\begin{aligned}
\mathbb{E} \left(b_s(\bar{X}_s^i, \bar{X}_s^j) \cdot b_s(\bar{X}_s^i, \bar{X}_s^k) \right) &= \mathbb{E} \left(b(\bar{X}_s^i, \bar{X}_s^j) \cdot b(\bar{X}_s^i, \bar{X}_s^k) \right) - \int_{\mathbb{R}^d} \mathbb{E} \left(b(\bar{X}_s^i, z) \cdot b(\bar{X}_s^i, \bar{X}_s^k) \right) \mu_s(dz) \\
&\quad - \int_{\mathbb{R}^d} \mathbb{E} \left(b(\bar{X}_s^i, z) \cdot b(\bar{X}_s^i, \bar{X}_s^j) \right) \mu_s(dz) \\
&\quad + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{E} \left(b(\bar{X}_s^i, z) \cdot b(\bar{X}_s^i, z') \right) \mu_s(dz) \mu_s(dz'). \tag{5.6}
\end{aligned}$$

Or $\mathbb{E} \left(b(\bar{X}_s^i, \bar{X}_s^j) \cdot b(\bar{X}_s^i, \bar{X}_s^k) \right) = \int_{(\mathbb{R}^d)^3} b(z, z') \cdot b(z, z'') \mu_s^{\otimes 3}(dz, dz', dz'')$ par indépendance de $\bar{X}_s^i, \bar{X}_s^j$ et $\bar{X}_s^k$. En procédant ainsi pour tous les termes ci-dessus, on trouve

$$\mathbb{E} \left(b_s(\bar{X}_s^i, \bar{X}_s^j) \cdot b_s(\bar{X}_s^i, \bar{X}_s^k) \right) = 0.$$

On a finalement

$$\begin{aligned}\mathbb{E} \left(\sup_{s \in [0, T]} |X_s^i - \bar{X}_s^i| \right) &\leq 2Me^{2MT} \int_0^T \frac{1}{N} \left(\sum_{j=1}^N \mathbb{E} \left(|b_s(\bar{X}_s^i, \bar{X}_s^j)|^2 \right) \right)^{\frac{1}{2}} ds \\ &= \frac{1}{\sqrt{N}} 2Me^{2MT} \int_0^T \left(\int_{(\mathbb{R}^d)^2} |b(z, z')|^2 \mu_s^{\otimes 2}(dz, dz') \right)^{\frac{1}{2}} ds,\end{aligned}$$

d'où le résultat. $\square$

Exercice 5.6. Reprendre la preuve précédente à partir de l'équation (5.6), lorsque nous avons supposé que les CI étaient iid. Faire la démonstration dans le cas général de l'énoncé, à savoir lorsque les CI sont échangeables et non iid. On pourra observer que dans le terme de gauche de (5.6), conditionnellement à X_0^i, X_0^j, X_0^k , les variables $\bar{X}_s^i, \bar{X}_s^j$ et $\bar{X}_s^k$ sont indépendantes.

Le théorème précédent décrit le comportement d'une seule particule lorsque $N \rightarrow +\infty$, mais ne décrit pas le comportement global du système de particules. Pour cela, on peut par exemple s'intéresser à la loi $\mathcal{L}^N$ de $(X^1, \dots, X^N)$ sur l'espace $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^N)$. Notons $\mathcal{L}_k^N$ la loi de k des N particules. En supposant que les particules sont échangeables², on a

$$\begin{aligned}\mathcal{W}_T(\mathcal{L}_k^N, \mu^{\otimes k}) &\leq \mathbb{E} \left(\sup_{s \in [0, T]} \left| (X_s^1, \dots, X_s^k) - (\bar{X}_s^1, \dots, \bar{X}_s^k) \right| \right) \\ &\leq k \max_{1 \leq i \leq k} \mathbb{E} \left(\sup_{s \in [0, T]} |X_s^{i, N} - \bar{X}_s^i| \right) \leq \frac{kC_T}{\sqrt{N}}.\end{aligned}$$

Remarquons que le facteur k ci-dessus n'est pas optimal (en travaillant en Wasserstein-2 et par extension du théorème 5.5, on aurait plutôt $\sqrt{k}$). On retiendra néanmoins que la méthode qui consiste à passer par le théorème 5.5 est propre aux interactions lipschitziennes et ne s'étend pas au cadre non-lipschitzien.

Une alternative au calcul ci-dessus consiste à considérer la mesure empirique du système de particules. On verra dans la section suivante que ce problème est équivalent au précédent (convergence de $\mathcal{L}_k^N$). Profitons-en pour faire une remarque importante concernant la convergence des mesures aléatoires. Par le théorème Portmanteau, montrer la convergence en loi de μ^N vers μ ³ revient à montrer que pour toute fonction Lipschitz bornée $\Phi : (\mathcal{P}(\mathcal{C}_T), d_{LP}) \rightarrow \mathbb{R}$, on a

$$\mathbb{E}\Phi(\mu^N) \xrightarrow{N \rightarrow +\infty} \mathbb{E}\Phi(\mu).$$

Le lemme suivant va nous permettre de simplifier considérablement cette caractérisation de la convergence en loi des mesures aléatoires, *lorsque la limite μ est déterministe*.

Lemme 5.7. Soit E un espace polonais, soit $\mu \in \mathcal{P}(E)$ une mesure déterministe et soit $(\mu^N)_{N \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathcal{P}(E)$. Si pour tout $\varphi \in \mathcal{C}_b(E)$, on a

$$\mathbb{E} |\langle \mu^N, \varphi \rangle - \langle \mu, \varphi \rangle| \xrightarrow{N \rightarrow +\infty} 0,$$

alors $\mu^N \xrightarrow{(d)} \mu$.

2. ce qui est le cas dès qu'on choisit une loi initiale symétrique, cf Exercice 5.4.

3. c'est-à-dire la convergence étroite de la loi de μ^N dans l'espace $\mathcal{P}(\mathcal{C}_T)$.

Démonstration. Soit $\Phi \in \mathcal{C}_b(\mathcal{P}(E))$ et notons $M = \sup_{\nu \in \mathcal{P}(E)} |\Phi(\nu)|$. Sans perte de généralité, supposons que $\Phi(\mu) = 0$. On veut donc montrer que $\mathbb{E}\Phi(\mu^N) \rightarrow 0$.

Soit $\varepsilon > 0$. Par continuité de Φ , il existe un ouvert U de $\mathcal{P}(E)$ qui contient μ tel que pour tout $\nu \in U$, $|\Phi(\nu)| < \varepsilon$. On rappelle que la convergence étroite des mesures est un cas particulier de la convergence faible sur $\mathcal{C}_b(E)^*$. Or par définition de la topologie faible, il existe un nombre fini de fonctions $f_1, \dots, f_n \in \mathcal{C}_b(E)$ et un $\delta > 0$ tels que si

$$\max_{1 \leq i \leq n} \left| \int_E f_i d\nu - \int_E f_i d\mu \right| < \delta,$$

alors $\nu \in U$.

Or par l'inégalité de Markov et l'hypothèse du lemme, on a qu'il existe $N_0 \in \mathbb{N}$ tel que pour tout $N \geq N_0$,

$$\mathbb{P}(|\langle \mu^N, f_i \rangle - \langle \mu, f_i \rangle| \geq \delta) \leq \frac{\varepsilon}{n},$$

de sorte que

$$\mathbb{P}\left(\bigcup_{i=1}^n \{|\langle \mu^N, f_i \rangle - \langle \mu, f_i \rangle| \geq \delta\}\right) \leq \varepsilon.$$

Autrement dit, $\mathbb{P}(\mu^N \notin U) \leq \varepsilon$, pour tout $N \geq N_0$. Par conséquent, puisque Φ est borné par ε sur U et que Φ est borné globalement par M , on obtient

$$\begin{aligned} |\mathbb{E}\Phi(\mu^N)| &\leq |\mathbb{E}(\Phi(\mu^N)\mathbb{1}_{\{\mu^N \in U\}})| + |\mathbb{E}(\Phi(\mu^N)\mathbb{1}_{\{\mu^N \notin U\}})| \\ &\leq \varepsilon + M\mathbb{P}(\mu^N \notin U) \\ &\leq \varepsilon(1 + M), \end{aligned}$$

d'où le résultat. □

Remarque 5.8. *Ce résultat nécessite que la limite μ soit déterministe. Où cette propriété est-elle utilisée dans la preuve ?*

5.2 Propagation du chaos

L'objectif de cette section est de présenter quelques propriétés fondamentales liées aux systèmes chaotiques, propriétés que nous appliquerons au système de particules à interaction lipschitzienne rencontré dans la section précédente.

5.2.1 Propriétés générales

Définition 5.9. *Soit E un espace polonais. On dit qu'une mesure $\mathcal{L}^N \in \mathcal{P}(E^N)$ est symétrique si pour toute permutation π de $\{1, \dots, N\}$, on a pour tous $A_1, \dots, A_N \in \mathcal{B}(E)$:*

$$\mathcal{L}^N(A_1 \times \dots \times A_N) = \mathcal{L}^N(A_{\pi(1)} \times \dots \times A_{\pi(N)}).$$

Si $\mathcal{L}^N$ est la loi de $(X_1, \dots, X_N)$, on dit que ces v.a. sont échangeables⁴.

4. cette définition est cohérente avec celle vue précédemment, cf. définition 5.3.

Définition 5.10. Soit E un espace polonais. Soit $\mathcal{L} \in \mathcal{P}(E)$ et soit $(\mathcal{L}^N)_{N \in \mathbb{N}}$ une suite de mesures où pour tout $N \geq 1$, $\mathcal{L}^N \in \mathcal{P}(E^N)$ est symétrique. On dit que $(\mathcal{L}^N)_{N \in \mathbb{N}}$ est $\mathcal{L}$ -chaotique si pour tout $k \geq 1$, pour tout $\varphi_1, \dots, \varphi_k \in \mathcal{C}_b(E)$,

$$\int_{E^N} \varphi_1(x_1) \dots \varphi_k(x_k) \mathcal{L}^N(dx_1, \dots, dx_k, \dots, dx_N) \xrightarrow{N \rightarrow +\infty} \prod_{i=1}^k \int_E \varphi_i d\mathcal{L}. \quad (5.7)$$

Dans le cas d'un système de particules échangeables (i.e. la loi du système est symétrique), on dit qu'il y a *propagation du chaos* si la suite de lois du système $(\mathcal{L}^N)$ est $\mathcal{L}$ -chaotique pour une certaine mesure $\mathcal{L} \in \mathcal{P}(E)$. Comme le montre le résultat suivant, la propagation du chaos pour un système de particules échangeables est équivalente à la convergence en loi de la mesure empirique de ce système vers la mesure (déterministe) $\mathcal{L}$. C'est d'une certaine manière une loi des grands nombres.

Théorème 5.11. Soit E un espace polonais et pour tout $N \in \mathbb{N}^*$, soit $\mathcal{L}^N$ une mesure symétrique sur E^N . Pour tout $N \geq 1$, on peut introduire une variable aléatoire $(X^{1,N}, \dots, X^{N,N})$ de loi $\mathcal{L}^N$, dont on note alors $\mu^N = \frac{1}{N} \sum_{k=1}^N \delta_{X^{k,N}}$ la mesure empirique. Soit $\mu \in \mathcal{P}(E)$. Les assertions suivantes sont équivalentes :

- (i) $(\mathcal{L}^N)_{N \in \mathbb{N}^*}$ est μ -chaotique ;
- (ii) $(\mu^N)_{N \in \mathbb{N}^*}$ converge en loi vers μ ;
- (iii) la convergence (5.7) est vérifiée pour $k = 2$.

Le prochain résultat donnera un critère de tension pour la mesure empirique. On retrouve ainsi le schéma de preuve général présenté dans ce cours pour démontrer la convergence d'une suite de variables aléatoires : 1) établir la tension ; 2) identifier la limite.

Remarquons qu'en général pour les systèmes de particules, le deuxième point consistera à montrer que toute valeur d'adhérence de $(\mu^N)_{N \in \mathbb{N}^*}$ est solution d'un problème de martingale non-linéaire. Par exemple dans le cas du système de particules (5.1), on peut montrer que si X est le processus canonique sur $\mathcal{C}$, alors pour tout $\varphi \in \mathcal{C}_b^2$,

$$\varphi(X_t) - \varphi(X_0) - \int_0^t \left(\frac{1}{2} \Delta \varphi + \int b(\cdot, y) \mu_s(dy) \cdot \nabla \varphi \right) (X_s) ds$$

est une μ -martingale. En appliquant l'espérance, on en déduit alors que μ est solution faible de l'EDP (5.5).

Démonstration. Dans toute la preuve, on notera X^i pour $X^{i,N}$, mais gardez bien en tête que la loi de X^i dépend implicitement de N .

Montrons que (iii) implique (ii). Soit $\varphi \in \mathcal{C}_b(E)$, alors

$$\begin{aligned} \mathbb{E} \left[(\langle \mu^N, \varphi \rangle - \langle \mu, \varphi \rangle)^2 \right] &= \frac{1}{N^2} \sum_{i,j=1}^N \mathbb{E} (\varphi(X^i) \varphi(X^j)) - \frac{2}{N} \langle \mu, \varphi \rangle \sum_{i=1}^N \mathbb{E} \varphi(X^i) + \langle \mu, \varphi \rangle^2 \\ &= \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} (\varphi(X^i)^2) + \frac{1}{N^2} \sum_{i \neq j} \mathbb{E} (\varphi(X^i) \varphi(X^j)) - \frac{2}{N} \langle \mu, \varphi \rangle \sum_{i=1}^N \mathbb{E} \varphi(X^i) + \langle \mu, \varphi \rangle^2, \end{aligned}$$

ce qui donne, en utilisant l'échangeabilité et la propriété de convergence (5.7) avec $k = 1$ et $k = 2$:

$$\begin{aligned} \mathbb{E} \left[(\langle \mu^N, \varphi \rangle - \langle \mu, \varphi \rangle)^2 \right] &= \frac{1}{N} \mathbb{E} (\varphi(X^1)^2) + \frac{N-1}{N} \mathbb{E} (\varphi(X^1) \varphi(X^2)) - 2 \langle \mu, \varphi \rangle \mathbb{E} \varphi(X^1) + \langle \mu, \varphi \rangle^2 \\ &\xrightarrow{N \rightarrow +\infty} 0 + \langle \mu, \varphi \rangle^2 - 2 \langle \mu, \varphi \rangle^2 + \langle \mu, \varphi \rangle^2 = 0. \end{aligned}$$

D'après le lemme 5.7, on en déduit que $\mu^N \xrightarrow{(d)} \mu$.

Montrons maintenant que (ii) $\Rightarrow$ (i). Soient $k \geq 1$ et $\varphi_1, \dots, \varphi_k \in \mathcal{C}_b(E)$, alors pour $N \geq k$,

$$\left| \int_{E^N} \varphi_1(x_1) \dots \varphi_k(x_k) \mathcal{L}^N(dx_1, \dots, dx_N) - \prod_{i=1}^k \int_E \varphi_i d\mathcal{L} \right| \leq \left| \int_{E^N} \varphi_1(x_1) \dots \varphi_k(x_k) \mathcal{L}^N(dx_1, \dots, dx_N) - \mathbb{E} \left(\prod_{i=1}^k \langle \mu^N, \varphi_i \rangle \right) \right| \quad (5.8)$$

$$+ \left| \mathbb{E} \left(\prod_{i=1}^k \langle \mu^N, \varphi_i \rangle \right) - \prod_{i=1}^k \int_E \varphi_i d\mathcal{L} \right|. \quad (5.9)$$

Par hypothèse ($\mu^N \xrightarrow{(d)} \mu$), le terme (5.9) tend vers 0. Par symétrie, le premier terme (5.8) vaut quant à lui

$$\left| \mathbb{E} \left(\prod_{i=1}^k \varphi_i(X^i) \right) - \mathbb{E} \left(\prod_{i=1}^k \langle \mu^N, \varphi_i \rangle \right) \right| = \left| \mathbb{E} \left(\frac{1}{N!} \sum_{\pi \in S_N} \prod_{i=1}^k \varphi_i(X^{\pi(i)}) - \prod_{i=1}^k \frac{1}{N} \sum_{j=1}^N \varphi_i(X^j) \right) \right|,$$

où on a noté S_N le groupe des permutations de taille N . On va majorer ponctuellement le contenu de cette dernière espérance, c'est-à-dire

$$\left| \mathbb{E} \left(\prod_{i=1}^k \varphi_i(X^i) \right) - \mathbb{E} \left(\prod_{i=1}^k \langle \mu^N, \varphi_i \rangle \right) \right| \leq \sup_{(x_1, \dots, x_N) \in E^N} \left| \frac{1}{N!} \sum_{\pi \in S_N} \prod_{i=1}^k \varphi_i(x_{\pi(i)}) - \prod_{i=1}^k \frac{1}{N} \sum_{j=1}^N \varphi_i(x_j) \right|.$$

On remarque que le deuxième produit se décompose de la façon suivante :

$$\begin{aligned} \prod_{i=1}^k \frac{1}{N} \sum_{j=1}^N \varphi_i(x_j) &= \frac{1}{N^k} \sum_{j_1, \dots, j_k=1}^N \prod_{i=1}^k \varphi_i(x_{j_i}) \\ &= \frac{1}{N^k} \sum_{J \in \mathcal{I}_{k,N}} \prod_{i=1}^k \varphi_i(x_{J(i)}) + \frac{1}{N^k} \sum_{J \notin \mathcal{I}_{k,N}} \prod_{i=1}^k \varphi_i(x_{J(i)}), \end{aligned}$$

où $\mathcal{I}_{k,N}$ est l'ensemble des injections de $\{1, \dots, k\}$ vers $\{1, \dots, N\}$. Par ailleurs, pour une injection $J \in \mathcal{I}_{k,N}$ donnée, on retrouve le terme $\prod_{i=1}^k \varphi_i(x_{J(i)})$ $(N-k)!$ -fois dans la somme $\sum_{\pi \in S_N} \prod_{i=1}^k \varphi_i(x_{\pi(i)})$, donc en regroupant ces termes, on obtient :

$$\begin{aligned} &\left| \frac{1}{N!} \sum_{\pi \in S_N} \prod_{i=1}^k \varphi_i(x_{\pi(i)}) - \prod_{i=1}^k \frac{1}{N} \sum_{j=1}^N \varphi_i(x_j) \right| \\ &= \left| \frac{(N-k)!}{N!} \sum_{J \in \mathcal{I}_{k,N}} \prod_{i=1}^k \varphi_i(x_{J(i)}) - \frac{1}{N^k} \sum_{J \in \mathcal{I}_{k,N}} \prod_{i=1}^k \varphi_i(x_{J(i)}) - \frac{1}{N^k} \sum_{J \notin \mathcal{I}_{k,N}} \prod_{i=1}^k \varphi_i(x_{J(i)}) \right| \\ &= \left| \left(\frac{(N-k)!}{N!} - \frac{1}{N^k} \right) \sum_{J \in \mathcal{I}_{k,N}} \prod_{i=1}^k \varphi_i(x_{J(i)}) - \frac{1}{N^k} \sum_{J \notin \mathcal{I}_{k,N}} \prod_{i=1}^k \varphi_i(x_{J(i)}) \right|. \end{aligned}$$

On utilise maintenant le fait que les φ_i sont bornés (on note $M = \max \|\varphi_i\|_\infty$) et que le cardinal de $\mathcal{I}_{k,N}$ vaut $\frac{N!}{(N-k)!}$:

$$\begin{aligned} & \left| \frac{1}{N!} \sum_{\pi \in S_N} \prod_{i=1}^k \varphi_i(x_{\pi(i)}) - \prod_{i=1}^k \frac{1}{N} \sum_{j=1}^N \varphi_i(x_j) \right| \\ & \leq M^k \left(\left(\frac{(N-k)!}{N!} - \frac{1}{N^k} \right) \frac{N!}{(N-k)!} + \frac{1}{N^k} \left(N^k - \frac{N!}{(N-k)!} \right) \right) \\ & = 2M^k \left(1 - \frac{N!}{N^k(N-k)!} \right). \end{aligned}$$

Ainsi

$$\left| \mathbb{E} \left(\prod_{i=1}^k \varphi_i(X^i) \right) - \mathbb{E} \left(\prod_{i=1}^k \langle \mu^N, \varphi_i \rangle \right) \right| \leq 2M^k \left(1 - \frac{N!}{N^k(N-k)!} \right),$$

et on vérifie que le terme de droite tend vers 0 lorsque $N \rightarrow +\infty$. On a donc montré que le terme (5.8) tend vers 0, d'où la propagation du chaos.

Enfin, on a bien sûr que $(i) \Rightarrow (iii)$, d'où l'équivalence des trois propositions. $\square$

Enonçons maintenant un critère de tension.

Théorème 5.12. *On reprend les mêmes hypothèses et notations qu'au théorème 5.11. On note $\mathcal{L}_1^N$ la première marginale de $\mathcal{L}^N$, c'est-à-dire la loi de X^1 (et par symétrie de n'importe quel X^k) dans le système de N particules. Alors $(\mu^N)_{N \in \mathbb{N}}$ est tendue⁵ si et seulement si la suite $(\mathcal{L}_1^N)_{N \in \mathbb{N}}$ est tendue.*

Remarque 5.13. *Attention, il ne faut pas confondre avec le résultat que vous deviez démontrer en DM. Dans le DM, on dispose d'une suite de variables $(X_k)_{k \in \mathbb{N}}$ et on utilise la tension de la loi de X_1 seule pour démontrer que $(\frac{1}{N} \sum_{k=1}^N \delta_{X_k})_{N \in \mathbb{N}^*}$ est tendue. Alors qu'ici on a pour tout $N \in \mathbb{N}^*$, des variables $(X^1, \dots, X^N)$ où la loi de X^1 dépend de N .*

Démonstration. On commence par introduire la notion d'intensité d'une mesure : pour $\nu \in \mathcal{P}(\mathcal{P}(E))$ ⁶, l'intensité de ν , notée $I(\nu)$ est une mesure de probabilité sur E définie par

$$\langle I(\nu), f \rangle = \int_{\mathcal{P}(E)} \langle m, f \rangle \nu(dm), \quad \forall f \in L^\infty(E).$$

On peut en fait démontrer que si (ν_n) est une famille de mesures de $\mathcal{P}(\mathcal{P}(E))$, (ν_n) est tendue si et seulement si la famille $(I(\nu_n))$ est tendue dans $\mathcal{P}(E)$ (voir preuve de la Proposition 2.2 (ii) de [19]).

Ainsi, on remarque que par symétrie, la mesure d'intensité de la loi de μ^N n'est autre que la loi $\mathcal{L}_1^N$ (**exercice**), d'où le résultat souhaité. $\square$

5. on rappelle que cela signifie que pour tout $\varepsilon > 0$, il existe un compact K de $\mathcal{P}(\mathcal{P}(E))$ tel que $\inf_N \mathbb{P}(\mu^N \in K) \geq 1 - \varepsilon$.

6. on rappelle que grâce au théorème 2.32, $\mathcal{P}(\mathcal{P}(E))$ est un espace polonais.

5.2.2 Applications aux systèmes de particules à coefficients Lipschitz

On reprend ici les notations du paragraphe 5.1 et les hypothèses du théorème 5.5, en particulier on suppose que pour tout $N \in \mathbb{N}^*$, $(X_0^i)_{i \leq N}$ est une famille échangeable et que les particules du système (5.1) et celles données par l'équation de McKean-Vlasov sont dirigées par les mêmes browniens et ont les mêmes CI $(X_0^i)_{i \leq N}$.

En utilisant le lemme 5.7, on va montrer la convergence en loi de la mesure empirique sur l'espace des trajectoires : pour $\varphi \in \mathcal{C}_b(\mathcal{C}_T)$, en utilisant l'échangeabilité des particules

$$\begin{aligned} |\mathbb{E}(\langle \mu^N, \varphi \rangle - \langle \mu, \varphi \rangle)| &\leq \left| \mathbb{E} \left(\langle \mu^N, \varphi \rangle - \frac{1}{N} \sum_{j=1}^N \varphi(\bar{X}^j) \right) \right| + \mathbb{E} \left(\left| \frac{1}{N} \sum_{j=1}^N \varphi(\bar{X}^j) - \langle \mu, \varphi \rangle \right| \right) \\ &\leq |\mathbb{E}(\varphi(X^{1,N}) - \varphi(\bar{X}^1))| + \mathbb{E} \left(\left| \frac{1}{N} \sum_{j=1}^N \varphi(\bar{X}^j) - \langle \mu, \varphi \rangle \right| \right) \end{aligned}$$

Pour le dernier terme, on a l'identité générale suivante : si $(Y_k)_{k \in \mathbb{N}}$ est une famille de variables aléatoires iid, centrées et de second moment fini, alors

$$\mathbb{E} \left(\left(\frac{1}{N} \sum_{k=1}^N Y_k \right)^2 \right) = \frac{1}{N} \mathbb{E}((Y_1)^2).$$

Ainsi

$$\mathbb{E} \left(\left| \frac{1}{N} \sum_{j=1}^N \varphi(\bar{X}^j) - \langle \mu, \varphi \rangle \right| \right) \leq \frac{(\mathbb{E}(\varphi(\bar{X}^1) - \langle \mu, \varphi \rangle)^2)^{1/2}}{\sqrt{N}} \xrightarrow[N \rightarrow \infty]{} 0. \quad (5.10)$$

Pour le terme $\mathbb{E}(\varphi(X^{1,N}) - \varphi(\bar{X}^1))$, on a simplement qu'il tend vers 0. En effet, cela est équivalent à la convergence en loi de X^1 vers $\bar{X}^1$ dans $\mathcal{C}_T$. Par le théorème Portmanteau, il suffit de montrer que $\mathbb{E}(\psi(X^{1,N}) - \psi(\bar{X}^1))$ tend vers 0 pour toute fonction Lipschitz $\psi : \mathcal{C}_T \rightarrow \mathbb{R}$. Or pour une telle fonction, on sait d'après le théorème 5.5 que

$$|\mathbb{E}(\psi(X^{1,N}) - \psi(\bar{X}^1))| \leq \frac{C_T}{\sqrt{N}} \|\psi\|_{\text{Lip}}. \quad (5.11)$$

Ainsi $\mathbb{E}(\varphi(X^{1,N}) - \varphi(\bar{X}^1))$ tend vers 0, pour tout $\varphi \in \mathcal{C}_b(\mathcal{C}_T)$, ce qui combiné avec (5.10) implique que

$$|\mathbb{E}(\langle \mu^N, \varphi \rangle - \langle \mu, \varphi \rangle)| \xrightarrow[N \rightarrow \infty]{} 0.$$

D'après le lemme 5.7, on a donc que $(\mu^N)_{N \in \mathbb{N}}$ converge en loi vers μ dans l'espace $\mathcal{P}(\mathcal{C}_T)$.

Grâce au théorème 5.11, on a donc propagation du chaos pour ce système de particules.

Même si cela ne caractérise pas la convergence en loi, on peut noter grâce à (5.10) et (5.11) que pour tout $\varphi \in \text{Lip}(\mathcal{C}_T)$,

$$|\mathbb{E}(\langle \mu^N, \varphi \rangle - \langle \mu, \varphi \rangle)| \leq \frac{C}{\sqrt{N}}.$$

5.3 Considérations diverses sur les interactions singulières

▷ Equation de Burgers, en dimension 1, modèle de mécanique des fluides, $b(x, y) = c\delta_0(x - y)$:

$$\partial_t u = \frac{1}{2} \partial_{xx}^2 u - c \partial_x u^2.$$

- ▷ Noyau de Biot-Savart dans $\mathbb{R}^2$, $b(x, y) = b_0(x - y)$ avec $b_0(x) = \frac{1}{\pi} \frac{x^\perp}{|x|^2}$, alors si u est la vortécité de v ($u = \text{rot } v$), solution de l'équation de Navier-Stokes en $2d$, u est solution de l'équation

$$\partial_t u = \frac{1}{2} \Delta u - \nabla (u b_0 * u).$$

- ▷ Keller-Segel (parabolique-elliptique), modèle chimiotaxie. Dans $\mathbb{R}^2$, $b(x, y) = b_0(x - y)$ avec $b_0(x) = -\chi \frac{x}{|x|^2}$, $\chi > 0$.
- ▷ Noyaux de Riesz pour les plasmas (...).
- ▷ Interactions de Lennard-Jones (noyau attractif-répulsif, dynamique moléculaire).

On voit qu'il s'agit d'EDP non-linéaires dans tous les cas.

Annexe A

Résultats divers

A.1 Topologie et analyse fonctionnelle

Lemme A.1. Soit (E, d) un espace métrique. Alors E est compact si et seulement si il est précompact et complet.

Démonstration. Exercice. □

Lemme A.2. Soit (E, d) un espace métrique séparable. Alors il existe une distance $\tilde{d}$ équivalente à d (i.e. qui induit la même topologie) et telle que $(E, \tilde{d})$ est précompact.

Démonstration. Voir Lemme 1.2 de [14]. □

Complété d'un espace métrique. Soit (E, d) un espace métrique. Soit $\widehat{E}$ l'ensemble des suites de Cauchy sur E . On considère la relation d'équivalence suivante : pour $u, v \in \widehat{E}$, $u \sim v$ si et seulement si $d(u_n, v_n) \rightarrow 0$ lorsque $n \rightarrow \infty$. On définit alors le complété de E comme étant l'ensemble $\widetilde{E}$ donné par

$$\widetilde{E} = \widehat{E} / \sim = \{[u] : u \in \widehat{E}\},$$

où $[u]$ désigne la classe d'équivalence de u pour $\sim$. On considérera toujours que E est un sous-ensemble de $\widetilde{E}$ grâce à l'injection canonique

$$\begin{array}{ccc} E & \rightarrow & \widetilde{E} \\ x & \mapsto & [x] \end{array},$$

où $[x]$ désigne la classe d'équivalence de la suite constante égale à x (qui est bien de Cauchy). Par ailleurs, on définit une distance sur $\widetilde{E}$, toujours notée d par simplicité, par

$$d([u], [v]) = \lim_{n \rightarrow +\infty} d(u_n, v_n), \quad \forall [u], [v] \in \widetilde{E}.$$

(On vérifiera que le membre de droite ne dépend pas du choix du représentant, que la limite existe et qu'ils s'agit bien d'une distance sur $\widetilde{E}$). Cette distance étend naturellement la distance d de E , à nouveau en considérant les suites constantes : $d(x, y) = d([x], [y])$, $\forall x, y \in E$. $(\widetilde{E}, d)$ est donc un espace métrique.

Montrons que E est dense dans $\widetilde{E}$, ce qui nous sera utile ensuite pour montrer que $\widetilde{E}$ est complet. Soit $[x] \in \widetilde{E}$ et $\varepsilon > 0$. Soit $(x_n)_{n \in \mathbb{N}}$ une suite de Cauchy de E dans la classe

d'équivalence de $[x]$. Il existe $N \in \mathbb{N}$ tel que pour tous $m, n \geq N$, $d(x_m, x_n) < \varepsilon/2$. On a alors

$$d([x_N], [x]) = \lim_{n \rightarrow +\infty} d(x_N, x_n) \leq \varepsilon/2 < \varepsilon.$$

D'où la densité de E dans $\tilde{E}$.

Montrons que toute suite de Cauchy de (E, d) converge dans $(\tilde{E}, d)$. Soit $(x_n)_{n \in \mathbb{N}}$ une suite de Cauchy de E et notons $x = [(x_n)_{n \in \mathbb{N}}]$. On rappelle que $[x_n] \in \tilde{E}$ désigne la classe d'équivalence de la suite constante égale à x_n . Alors pour tout $m \in \mathbb{N}$,

$$d([x_m], [x]) = \lim_{n \rightarrow +\infty} d(x_m, x_n),$$

et comme (x_n) est de Cauchy, la limite précédente est aussi petite qu'on veut pour m suffisamment grand. Donc $\lim_{m \rightarrow +\infty} d([x_m], [x]) = 0$ et la propriété est démontrée.

Puisque E est dense dans $\tilde{E}$, on pourra vérifier que le fait que toute suite de Cauchy de (E, d) converge dans $(\tilde{E}, d)$ implique que $(\tilde{E}, d)$ est complet.

Lemme A.3. *Soit E et F deux espaces métriques avec F complet. Soit E_0 une partie dense de E et f une application uniformément continue de E_0 vers F . Alors f se prolonge de manière unique en une fonction uniformément continue de E vers F .*

Bibliographie

- [1] Nathanaël Berestycki. Stochastic calculus. <https://homepage.univie.ac.at/nathanael.berestycki/teach/StoCal/sc3.pdf>, 2010. Cambridge, Part III Lecture notes.
- [2] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [3] V. I. Bogachev. *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007.
- [4] W. Braun and K. Hepp. The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Comm. Math. Phys.*, 56(2) :101–113, 1977.
- [5] R. L. Dobrušin. Vlasov equations. *Funktsional. Anal. i Prilozhen.*, 13(2) :48–58, 96, 1979.
- [6] Richard Durrett. *Stochastic calculus*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996. A practical introduction.
- [7] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.
- [8] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197. University of California Press, Berkeley and Los Angeles, 1956.
- [9] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, 1997.
- [10] Claude Kipnis and Claudio Landim. *Scaling limits of interacting particle systems*, volume 320 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.
- [11] Jean-François Le Gall. Intégration, probabilités et processus aléatoires. <https://www.imo.universite-paris-saclay.fr/~jflegall/IPPA2.pdf>, 2006. Notes de cours, ENS.
- [12] Sylvie Méléard. Asymptotic behaviour of some interacting particle systems ; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, Berlin, 1996.

- [13] Pierre-Loïc Méliot. Convergence de mesures, processus de poisson et processus de lévy. <https://www.imo.universite-paris-saclay.fr/~meliot/master/poisson.pdf>, 2016. Notes de cours de M2, Orsay.
- [14] Grégory Miermont. Théorèmes limites et processus de poisson. <http://perso.ens-lyon.fr/gregory.miermont/thlim.pdf>, 2011-2012. Notes de cours de M2, Orsay.
- [15] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [16] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.
- [17] Sylvia Serfaty. Mean field limit for Coulomb-type flows. *Duke Math. J.*, 169(15) : 2887–2935, 2020.
- [18] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006.
- [19] Alain-Sol Sznitman. Topics in propagation of chaos. In *École d’Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.
- [20] Cédric Villani. Limite de champ moyen. https://cedricvillani.org/sites/dev/files/old_images/2013/03/champ-moyen.pdf, 2001-2002. Notes de cours de DEA de l’ENS Lyon.