



# **C**onvergence de mesures, **G**randes déviations, **P**ercolation.

Thierry Lévy, Sorbonne université, LPSM

Cours de Master 2 – Automne 2022

Version du 5 septembre 2022

# Table des matières

---

<b>I</b>	<b>Convergence en loi des processus continus</b>	<b>4</b>
<b>1</b>	<b>Mesures sur les espaces polonais</b>	<b>6</b>
1.1	Espaces polonais . . . . .	6
1.2	Mesures boréliennes . . . . .	10
1.3	Convergence faible des mesures de probabilité . . . . .	16
1.4	L'espace topologique $\mathcal{M}(E)$ . . . . .	19
1.5	Le théorème de Prokhorov . . . . .	28
<b>2</b>	<b>Convergence en loi des processus continus</b>	<b>33</b>
2.1	L'espace des fonctions continues à valeurs dans un espace polonais . . . . .	33
2.2	Compacité dans $\mathcal{W}(E)$ . . . . .	36
2.3	Convergence en loi dans $\mathcal{W}(E)$ . . . . .	38
2.4	Critères de tension dans $\mathcal{M}(\mathcal{C}([0,1], E))$ . . . . .	41
2.5	Le théorème de Donsker . . . . .	47
<b>II</b>	<b>Introduction aux grandes déviations</b>	<b>51</b>
<b>3</b>	<b>Principes de grandes déviations</b>	<b>53</b>
3.1	Des événements rares . . . . .	53
3.2	Une concentration exponentielle . . . . .	55
3.3	Principes de grandes déviations . . . . .	58
<b>4</b>	<b>Grandes déviations pour des lois empiriques</b>	<b>62</b>
4.1	Entropie relative de deux mesures . . . . .	62
4.2	Résultats précis . . . . .	65
4.3	Le théorème de Sanov . . . . .	66
<b>5</b>	<b>Grandes déviations pour des moyennes empiriques</b>	<b>69</b>
5.1	Transformée de Legendre et fonction de taux . . . . .	69
5.2	Le théorème de Cramér . . . . .	71

---

<b>III</b>	<b>Introduction à la percolation</b>	<b>74</b>
<b>6</b>	<b>Une transition de phase</b>	<b>76</b>
6.1	Le modèle de la percolation par arêtes . . . . .	76
6.2	Existence d'une transition de phase . . . . .	81
<b>7</b>	<b>Phases sous-critique et sur-critique</b>	<b>88</b>
7.1	Ergodicité . . . . .	88
7.2	Nombre d'agrégats infinis . . . . .	90
7.3	Régularité de $\theta_d$ . . . . .	97
7.4	Deux inégalités classiques . . . . .	98
7.5	Petitesse exponentielle des agrégats sous-critiques . . . . .	100
<b>8</b>	<b>Le théorème de Kesten</b>	<b>107</b>
8.1	Une inégalité... . . . . .	107
8.2	...et l'autre . . . . .	109
<b>IV</b>	<b>Compléments</b>	<b>111</b>
<b>A</b>	<b>Complété d'un espace métrique</b>	<b>112</b>
<b>B</b>	<b>Limite inférieure, limite supérieure</b>	<b>116</b>
<b>C</b>	<b>Fonctions semi-continues</b>	<b>118</b>
<b>D</b>	<b>Théorème de représentation de Riesz</b>	<b>122</b>
<b>E</b>	<b>La mesure de Wiener</b>	<b>130</b>
<b>F</b>	<b>Transformée de Legendre</b>	<b>134</b>

**Première partie**

**Convergence en loi des processus  
continus**

L'objectif de cette première partie du cours est de démontrer un théorème de Donsker qui affirme qu'une grande classe de marches aléatoires, convenablement renormalisées, converge vers le mouvement brownien. L'article original de Donsker [**Donsker**] s'appelle « Un principe d'invariance ... » et ce mot d'*invariance* signifie que l'objet limite, le mouvement brownien, est le même pour une grande classe de marches aléatoires, en l'occurrence celles dont les pas sont centrés et de variance finie.

Pour parvenir à démontrer ce théorème, nous allons devoir étudier de manière assez générale les mesures boréliennes de probabilités sur les espaces métriques, et la convergence faible de suites de telles mesures. Deux jalons importants de cette étude sont le théorème de Portmanteau (théorème 1.3.2) et le théorème de Prokhorov (théorème 1.5.3), qui illustrent les interactions possibles des structures métrique et mesurable.

Après avoir mis en évidence le rôle crucial de la notion de tension dans la convergence des processus (voir la définition 1.5.1 et le théorème 2.3.5), nous démontrerons deux critères de tension qui permettent d'étudier des convergences de processus continus (voir la proposition 2.4.3 et le théorème 2.4.4). Nous appliquerons enfin le premier de ces critères à la démonstration du théorème de Donsker (théorème 2.5.1).

Une référence classique pour le contenu de ce chapitre est le livre de Billingsley [**Billingsley**]. L'ouvrage de Kallenberg [**Kallenberg**] peut aussi être une référence utile. J'ai aussi énormément utilisé, pour ce chapitre, les notes de cours de Grégory Miermont [**Miermont**] qui sont disponibles sur sa page web à l'adresse suivante : <http://perso.ens-lyon.fr/gregory.miermont/thlim.pdf>.

# 1

## Mesures sur les espaces polonais

---

Dans ce premier chapitre, nous allons étudier la convergence faible des mesures sur des espaces plus généraux que  $\mathbb{R}$  ou  $\mathbb{R}^n$ . Pour cela, il faut identifier le bon niveau de généralité, et l'usage conduit à choisir le cadre des espaces *polonais*.

### 1.1 Espaces polonais

Rappelons la définition d'un espace topologique.

**Définition 1.1.1** (Espace topologique). Soit  $E$  un ensemble. On appelle topologie sur  $E$  une classe  $\mathcal{O}$  de partie de  $E$ , appelées parties ouvertes de  $E$ , qui vérifient les propriétés suivantes.

1.  $\emptyset \in \mathcal{O}$  et  $E \in \mathcal{O}$ .
2. Pour tous  $U, V \in \mathcal{O}$ , on a  $U \cap V \in \mathcal{O}$ .
3. Pour toute famille  $(U_i)_{i \in I}$  d'éléments de  $\mathcal{O}$ , on a  $\bigcup_{i \in I} U_i \in \mathcal{O}$ .

Si  $\mathcal{O}$  est une topologie sur  $E$ , on appelle la paire  $(E, \mathcal{O})$  un espace topologique.

Voici une liste de quelques notions et faits fondamentaux de topologie dont nous aurons besoin.

- On appelle *fermée* une partie dont la partie complémentaire est ouverte.
- Une union finie de fermés, une intersection quelconque de fermés, sont encore des parties fermées.
- L'intersection de toutes les parties fermées contenant une partie  $A$  de  $E$  est une partie fermée de  $E$ , la plus petite partie fermée contenant  $A$ . On l'appelle l'*adhérence* de  $A$  et on la note  $\overline{A}$ .
- L'union de toutes les parties ouvertes incluses dans une partie  $A$  de  $E$  est une partie ouverte de  $E$ , la plus grande partie ouverte incluse dans  $A$ . On l'appelle l'*intérieur* de  $A$  et on la note  $\overset{\circ}{A}$ .
- Une partie  $A$  de  $E$  telle que  $\overline{A} = E$  est dite *dense* dans  $E$ .
- On dit qu'un espace topologique  $(E, \mathcal{O})$  est *séparé* si pour tous  $x, y$  points distincts de  $E$  il existe des ouverts disjoints  $U$  et  $V$  tels que  $x \in U$  et  $y \in V$ .
- On dit qu'un espace topologique  $(E, \mathcal{O})$  est *compact* s'il est séparé (voir la note 5) et si de tout recouvrement de  $E$  par des ouverts on peut extraire un sous-recouvrement fini.
- On dit qu'une application  $f : (E, \mathcal{O}_E) \rightarrow (F, \mathcal{O}_F)$  entre deux espaces topologiques est *continue* si l'image réciproque de tout ouvert de  $F$  est un ouvert de  $E$ .

— Une application continue d'un espace topologique compact dans  $\mathbb{R}$  muni de sa topologie usuelle est bornée, et atteint ses bornes.

Beaucoup des espaces topologiques qu'on utilise couramment sont construits à partir d'un espace métrique.

**Définition 1.1.2** (Espace métrique). Soit  $E$  un ensemble. On appelle distance sur  $E$  une fonction  $d : E \times E \rightarrow \mathbb{R}_+$  telle que les propriétés suivantes soient vérifiées.

1. Pour tous  $x, y \in E$ ,  $d(x, y) = d(y, x)$ .
2. Pour tous  $x, y \in E$ ,  $d(x, y) = 0$  entraîne  $x = y$ .
3. Pour tous  $x, y, z \in E$ , on a  $d(x, z) \leq d(x, y) + d(y, z)$ .

Si  $d$  est une distance sur  $E$ , on appelle la paire  $(E, d)$  un espace métrique.

Si  $(E, d)$  est un espace métrique,  $x$  un point de  $E$  et  $r$  un réel positif ou nul, on note respectivement

$$B(x, r) = \{y \in E : d(x, y) < r\} \quad \text{et} \quad \bar{B}(x, r) = \{y \in E : d(x, y) \leq r\}$$

la boule ouverte et la boule fermée de centre  $x$  et de rayon  $r$ .

Sur un espace métrique, on peut construire une topologie, de la manière suivante.

**Proposition 1.1.3.** Soit  $(E, d)$  un espace métrique. La classe

$$\mathcal{O} = \{U \subset E : \forall x \in U, \exists r > 0, B(x, r) \subset U\}$$

est une topologie sur  $E$ , dite induite par la distance  $d$ .

La démonstration de cette proposition est laissée en exercice. La topologie usuelle de  $\mathbb{R}$  est la topologie induite par la distance usuelle. Sur  $\mathbb{R}^n$ , la topologie usuelle est également la topologie induite par la distance  $d(x, y) = \|x - y\|$ , où  $\|\cdot\|$  est une norme sur  $\mathbb{R}^n$ . Il semblerait que la topologie ainsi définie dépende de la norme qu'on a choisie, mais il n'en est rien. En effet, si la distance  $d$  dépend bien de la norme choisie, le fait que sur un espace vectoriel réel de dimension finie deux normes quelconques sont équivalentes entraîne que les topologies induites par les distances associées à ces deux normes sont égales.

Il n'est bien sûr pas question ici de faire un cours de topologie générale. Nous nous contenterons de définir les quelques notions dont nous aurons besoin.

**Définition 1.1.4** (Espace à base dénombrable). On dit qu'un espace topologique  $(E, \mathcal{O})$  est à base dénombrable s'il existe une famille dénombrable  $(O_n)_{n \in \mathbb{N}}$  d'ouverts telle que tout ouvert s'écrive comme union d'une sous-famille de cette famille :

$$\forall U \in \mathcal{O}, \exists J \subseteq \mathbb{N}, U = \bigcup_{n \in J} O_n.$$

Une telle famille d'ouverts s'appelle une base dénombrable de la topologie.

Par exemple, sur  $\mathbb{R}$ , la famille des intervalles ouverts à extrémités rationnelles est dénombrable et il est vrai que tout ouvert de  $\mathbb{R}$  peut être réalisé comme l'union d'une famille de tels intervalles. Ainsi, la topologie usuelle de  $\mathbb{R}$  est à base dénombrable.

**Exercice 1.1.** Vérifier que pour tout  $n \geq 1$ , la topologie usuelle de  $\mathbb{R}^n$  est à base dénombrable.

Pour les espaces métriques, le fait que la topologie soit à base dénombrable est équivalent à une propriété parfois plus facile à vérifier. Commençons par une définition.

**Définition 1.1.5** (Espace séparable). *On dit qu'un espace topologique  $(E, \mathcal{O})$  est séparable s'il admet une partie qui est dénombrable et dense.*

Le lien entre le fait d'être à base dénombrable et le fait d'être séparable est le suivant.

**Proposition 1.1.6.** 1. *Un espace topologique à base dénombrable est séparable.*

2. *Un espace métrique séparable est à base dénombrable.*

Il est utile avant de démontrer cette proposition de se rappeler, ou de vérifier, qu'une partie d'un espace topologique  $(E, \mathcal{O})$  est dense si et seulement si elle rencontre toute partie ouverte non vide : pour tout  $A \subseteq E$ ,

$$\overline{A} = E \iff \forall U \in \mathcal{O} \setminus \{\emptyset\}, A \cap U \neq \emptyset.$$

*Démonstration.* 1. Soit  $(E, \mathcal{O})$  un espace topologique à base dénombrable. Soit  $(O_n)_{n \in \mathbb{N}}$  une base dénombrable de sa topologie. Pour tout  $n \in \mathbb{N}$  tel que  $O_n \neq \emptyset$ , choisissons un point dans  $O_n$ . L'ensemble  $D$  des points ainsi choisis est dénombrable et rencontre, par construction, tout ouvert non vide. Il est donc dense, et la topologie  $\mathcal{O}$  est séparable.

2. Soit  $D$  une partie dénombrable et dense d'un espace métrique séparable  $(E, d)$ . L'ensemble des boules ouvertes de rayon rationnel et de centre appartenant à  $D$  est une base dénombrable de la topologie induite sur  $E$  par  $d$ . (La vérification de ce fait est laissée en exercice.)  $\square$

L'existence, dans  $\mathbb{R}^n$ , de la partie dénombrable et dense  $\mathbb{Q}^n$  nous donne une nouvelle preuve du fait que la topologie de  $\mathbb{R}^n$  est à base dénombrable.

Nous allons maintenant parler d'espaces métriques *complets*, et il faut pour cela parler de suites. Introduisons une notation : dans un espace topologique  $(E, \mathcal{O})$ , on dit qu'une partie  $V$  de  $E$  est un *voisinage* d'un point  $x$  de  $E$  si  $x$  appartient à l'intérieur de  $V$ , c'est-à-dire s'il existe une partie ouverte  $U$  telle que  $x \in U \subseteq V$ . On note  $\mathcal{V}(x)$  l'ensemble des voisinages de  $x$ .

**Définition 1.1.7** (Suites convergentes). *Soit  $(E, \mathcal{O})$  un espace topologique. Soit  $(x_n)_{n \geq 0}$  une suite d'éléments de  $E$ . Soit  $\ell$  un élément de  $E$ . On dit que la suite  $(x_n)_{n \geq 0}$  converge vers  $\ell$  si*

$$\forall V \in \mathcal{V}(\ell), \exists N \geq 0, \forall n \geq N, x_n \in V.$$

Si la suite  $(x_n)_{n \geq 0}$  converge vers  $\ell$ , on écrit

$$\lim_{n \rightarrow \infty} x_n = \ell \quad \text{ou} \quad x_n \xrightarrow[n \rightarrow \infty]{} \ell.$$

Dans un espace métrique  $(E, d)$ , une suite  $(x_n)_{n \geq 0}$  converge vers une limite  $\ell$  au sens de la définition précédente et relativement à la topologie induite par  $d$  si et seulement si

$$\forall \varepsilon > 0, \exists N \geq 0, \forall n \geq N, d(\ell, x_n) < \varepsilon.$$



**Définition 1.1.8** (Espace complet). Soit  $(E, d)$  un espace métrique. On dit qu'une suite  $(x_n)_{n \geq 0}$  d'éléments de  $E$  est de Cauchy si

$$\forall \varepsilon > 0, \exists N \geq 0, \forall m, n \geq N, d(x_m, x_n) < \varepsilon.$$

On dit que l'espace métrique  $(E, d)$  est complet si toute suite de Cauchy d'éléments de  $E$  est convergente.

Il n'est pas difficile de vérifier que dans un espace métrique, toute suite convergente est de Cauchy. Un espace où la réciproque est vraie est, par définition, un espace complet. Un fait fondamental de l'analyse réelle est que  $\mathbb{R}$  est complet.

**Exercice 1.2.** (Complété d'un espace métrique).<sup>1</sup> Soit  $(E, d)$  un espace métrique. On note  $\mathcal{S}$  l'ensemble des suites de Cauchy de  $(E, d)$ . On dit que deux suites de Cauchy  $(x_n)_{n \geq 0}$  et  $(y_n)_{n \geq 0}$  sont *confluentes* si  $\lim_{n \rightarrow \infty} d(x_n, y_n) = 0$ .

La relation de confluence est une relation d'équivalence sur  $\mathcal{S}$  et on note  $\hat{E}$  l'ensemble quotient. On note  $[(x_n)_{n \geq 0}]$  la classe de confluence d'une suite  $(x_n)_{n \geq 0}$ , c'est-à-dire l'ensemble des suites qui lui sont confluentes.

Étant donné deux éléments de  $\mathcal{S}$ , c'est-à-dire deux suites  $(x_n)_{n \geq 0}$  et  $(y_n)_{n \geq 0}$  de Cauchy sur  $E$ , confluentes ou non, la suite de réels  $(d(x_n, y_n))_{n \geq 0}$  admet une limite finie. De plus, cette limite ne dépend que de la classe de confluence des suites  $(x_n)_{n \geq 0}$  et  $(y_n)_{n \geq 0}$ . On peut donc poser sans ambiguïté

$$\hat{d}([(x_n)_{n \geq 0}], [(y_n)_{n \geq 0}]) = \lim_{n \rightarrow \infty} d(x_n, y_n).$$

La fonction  $\hat{d} : \hat{E} \times \hat{E} \rightarrow \mathbb{R}_+$  est une distance sur  $\hat{E}$ .

L'application de  $i : E \rightarrow \hat{E}$  qui envoie un élément  $x$  de  $E$  sur la classe de la suite constante égale à  $x$ <sup>2</sup> est isométrique, au sens où pour tous  $x, y \in E$ , on a  $\hat{d}(i(x), i(y)) = d(x, y)$ . Elle est en particulier injective et l'espace métrique  $(E, d)$  s'identifie au sous-espace métrique  $(i(E), \hat{d}|_{i(E)})$  de  $(\hat{E}, \hat{d})$ .

L'espace métrique  $(\hat{E}, \hat{d})$  est complet. De plus, la partie  $i(E)$  de  $\hat{E}$  est dense.

On appelle  $(\hat{E}, \hat{d})$  l'espace métrique complété de  $(E, d)$ . L'application  $i$  est surjective si et seulement si  $(E, d)$  est complet.

[ Considérons maintenant un espace métrique complet  $(F, \delta)$ . Soit  $f : E \rightarrow F$  une application continue de  $E$  dans  $F$ . Alors il existe une unique application continue  $\hat{f} : \hat{E} \rightarrow F$  telle que  $\hat{f} \circ i = f$ .

Enfin, soit  $(\hat{E}', \hat{d}')$  un espace métrique complet muni d'une application isométrique  $i' : E \rightarrow \hat{E}'$ . Supposons que pour tout espace métrique complet  $(F, \delta)$  et toute application continue  $f : E \rightarrow F$ , il existe une unique application continue  $\hat{f}' : \hat{E}' \rightarrow F$  telle que  $\hat{f}' \circ i' = f$ . Alors il existe une bijection isométrique  $g : \hat{E} \rightarrow \hat{E}'$  telle que  $g \circ i = i'$ . ]

La notion d'espace complet est d'essence métrique : la structure d'espace topologique ne permet pas de définir une suite de Cauchy. En effet, la structure topologique permet de donner un sens à l'idée d'être "proche d'un point donné", grâce à la notion de voisinage. Une suite  $(x_n)_{n \geq 0}$  converge vers une limite  $\ell$  si quelle que soit l'exigence avec laquelle on définit une

1. Cet exercice consiste en un texte dans lequel des affirmations sont faites mais non démontrées. Votre travail est de trouver quelles assertions demandent à être démontrées, et à les démontrer. La partie entre crochets et en caractères plus petits peut être laissée de côté dans un premier temps.

2. Cette classe n'est autre que l'ensemble des suites qui convergent vers  $x$ .

notion de proximité de  $\ell$  (ce qu'on exprime en disant "pour tout voisinage de  $\ell$ "), les termes de la suite finissent par être proches de  $\ell$  au sens de la notion qu'on a choisie. En revanche, une suite est de Cauchy si pour toute notion de proximité *entre deux points*, aussi exigeante soit-elle (pour tout  $\varepsilon > 0$ ), deux termes de rang assez grand de la suite finissent par satisfaire cette notion de proximité. Pour définir une suite de Cauchy, il faut donc une notion de proximité de deux points, qui soit en quelque sorte uniforme vis-à-vis de ces deux points. La manière la plus simple de définir cette notion est de disposer d'une distance.<sup>3</sup>

Nous pouvons maintenant définir les espaces dans lesquels nous nous placerons tout au long de ce cours.

**Définition 1.1.9** (Espace polonais). *On dit qu'un espace topologique  $(E, \mathcal{O})$  est polonais s'il est séparable et s'il existe une distance sur  $E$  qui induit la topologie  $\mathcal{O}$  et qui en fait un espace complet.*

Il faut prendre garde au fait que sur un espace topologique, la topologie peut être induite par plusieurs distances et que certaines peuvent être complètes alors que d'autres ne le sont pas. Ainsi, au sens strict, un espace métrique  $(E, d)$  peut être polonais sans être complet : c'est le cas par exemple de l'intervalle  $]0, 1[$  muni de la distance usuelle. En effet, cet espace métrique n'est pas complet, mais il est homéomorphe à  $\mathbb{R}$  muni de sa topologie usuelle, par exemple par l'application  $t \mapsto \log \frac{t}{1-t}$ . La distance

$$d'(s, t) = \left| \log \frac{t}{1-t} - \log \frac{s}{1-s} \right|$$

sur  $]0, 1[$  induit donc sa topologie usuelle et en fait un espace complet.

Ceci étant, l'usage veut que lorsqu'on dit sans autre précision d'un espace métrique  $(E, d)$  qu'il est polonais, cela signifie en général qu'il est séparable et complet, avec la distance  $d$ .<sup>4</sup>

**Exercice 1.3.** Montrer que sur  $\mathbb{R}$ , la fonction

$$\delta : (x, y) \mapsto \delta(x, y) = |\arctan(x) - \arctan(y)|$$

est une distance, qui induit la topologie usuelle de  $\mathbb{R}$ , mais que  $(\mathbb{R}, \delta)$  n'est pas complet.

## 1.2 Mesures boréliennes

Nous allons étudier la convergence faible de suites de mesures de probabilité sur un espace polonais  $(E, d)$ . Ces mesures sont naturellement définies sur la tribu borélienne de cet espace.

**Définition 1.2.1** (Tribu borélienne). *Soit  $(E, \mathcal{O})$  un espace topologique. On appelle tribu borélienne de  $(E, \mathcal{O})$  la tribu sur  $E$  engendrée par  $\mathcal{O}$ , c'est-à-dire la plus petite tribu sur  $E$  qui contient toutes les parties ouvertes.*

3. Si vous avez des penchants bourbakistes, vous pourrez toutefois avec profit vous renseigner sur la notion d'espace uniforme. Cette notion est, incidemment, apparue dans les travaux d'André Weil sur l'intégration dans les groupes topologiques.

4. On pourrait décider d'appeler *topologiquement polonais* un espace métrique dont la topologie est polonaise, sans exiger que sa distance en fasse un espace complet.

Nous noterons  $\mathcal{B}(\mathcal{O})$ ,  $\mathcal{B}_E$  ou plus simplement  $\mathcal{B}$  la tribu borélienne de  $(E, \mathcal{O})$ . Le choix de la tribu borélienne va nous permettre de lier des considérations de théorie de la mesure et de topologie. Un exemple élémentaire de ce lien est donné par la proposition suivante, dont nous donnons une démonstration au titre de familiarisation avec la topologie générale.

**Proposition 1.2.2.** Soient  $(E, \mathcal{O}_E)$  et  $(F, \mathcal{O}_F)$  des espaces topologiques. On munit  $E$  et  $F$  de leurs tribus boréliennes respectives  $\mathcal{B}_E$  et  $\mathcal{B}_F$ . Soit  $f : (E, \mathcal{O}_E) \rightarrow (F, \mathcal{O}_F)$  une fonction continue. Alors  $f$  est mesurable par rapport aux tribus  $\mathcal{B}_E$  et  $\mathcal{B}_F$ .

*Démonstration.* La démonstration de cette proposition est l'occasion de rappeler que par définition, une fonction  $f : (E, \mathcal{O}_E) \rightarrow (F, \mathcal{O}_F)$  est continue si pour tout ouvert  $V$  de  $F$ , l'image réciproque de  $V$  par  $f$  est un ouvert de  $E$ .

Ainsi, la tribu image de  $\mathcal{B}_E$  par  $f$ , qui est définie (attention!) par

$$f_*\mathcal{B}_E = \{B \subseteq F : f^{-1}(B) \in \mathcal{B}_E\}$$

contient  $\mathcal{O}_F$ . Or  $f_*\mathcal{B}_E$  est une tribu (vérifiez-le) — c'est la plus grande tribu dont on puisse munir  $F$  de telle sorte que  $f$  soit mesurable sur  $(E, \mathcal{B}_E)$ . Et cette tribu contient  $\mathcal{O}_F$ , donc elle contient  $\mathcal{B}_F$ .  $\square$

**Exercice 1.4.** Soient  $(E, \mathcal{O}_E)$  et  $(F, \mathcal{O}_F)$  des espaces topologiques. Montrer qu'une application  $f : E \rightarrow F$  est continue si et seulement si pour toute partie  $A$  de  $E$ , on a

$$f(\overline{A}) \subseteq \overline{f(A)}.$$

Dans tout ce qui suit,  $\mathbb{R}$  sera muni de sa distance, de sa topologie, et de sa tribu borélienne usuelles. Soit  $E$  un espace topologique. Nous noterons  $\mathcal{C}(E)$  l'ensemble de toutes les fonctions continues à valeurs réelles sur  $E$ , et  $\mathcal{C}_b(E)$  le sous-ensemble de  $\mathcal{C}(E)$  formé des fonctions continues et bornées.

L'ensemble  $\mathcal{C}(E)$  est un espace vectoriel réel, dont  $\mathcal{C}_b(E)$  est un sous-espace vectoriel. Considéré comme un espace vectoriel à part entière,  $\mathcal{C}_b(E)$  peut naturellement être muni d'une norme : on pose, pour toute fonction continue bornée  $f : E \rightarrow \mathbb{R}$ ,

$$\|f\| = \sup\{|f(x)| : x \in E\}.$$

La fonction  $\|\cdot\|$  est bien une norme et muni de cette norme,  $\mathcal{C}_b(E)$  est un espace de complet : on dit que c'est un espace de Banach.

**Exercice 1.5.** Montrer que  $(\mathcal{C}_b(E), \|\cdot\|)$  est un espace complet. Autrement dit, montrer que pour toute suite  $(f_n)_{n \geq 0}$  de fonctions continues et bornées sur  $E$  telle que

$$\lim_{n \rightarrow \infty} \left( \sup_{m \geq n} \|f_n - f_m\| \right) = 0$$

(ce qui est la définition d'une suite de Cauchy), il existe une fonction continue bornée  $f$  telle que

$$\lim_{n \rightarrow \infty} \|f_n - f\| = 0,$$

ce qui est la définition du fait que la suite  $(f_n)_{n \geq 0}$  converge vers  $f$ .

**Notation 1.2.3.** Soit  $(E, \mathcal{O})$  un espace topologique. Nous noterons  $\mathcal{M}(E)$  l'ensemble des mesures de probabilité boréliennes sur  $E$ , c'est-à-dire l'ensemble des mesures positives de masse totale 1 sur l'espace mesurable  $(E, \mathcal{B}_E)$ .

L'objet de ce premier chapitre est de munir  $\mathcal{M}(E)$  d'une topologie, et de l'étudier lorsque  $E$  est polonais. Nous n'allons toutefois pas faire cette hypothèse tout de suite.

Une première remarque importante est que si  $f : E \rightarrow \mathbb{R}$  est une fonction continue et bornée, et si  $\mu$  est un élément de  $\mathcal{M}(E)$ , alors  $f$  est mesurable par rapport à la tribu borélienne de  $E$ , et intégrable par rapport à  $\mu$ , puisque

$$\int_E |f| d\mu \leq \|f\| < \infty.$$

Il y a donc une application  $\mathcal{M}(E) \times \mathcal{C}_b(E) \rightarrow \mathbb{R}$  qui à un couple  $(\mu, f)$  associe  $\mu(f) = \int_E f d\mu$ . Pour toute mesure  $\mu \in \mathcal{M}(E)$ , l'application  $f \mapsto \mu(f)$  est linéaire en  $f$ , et continue (vérifiez qu'elle est de norme 1, c'est-à-dire qu'elle est 1-lipschitzienne). On a donc une application  $\mathcal{M}(E) \rightarrow \mathcal{C}_b(E)'$ , à valeurs dans l'espace dual de  $\mathcal{C}_b(E)$  et nous aimerions qu'elle soit injective, pour identifier  $\mathcal{M}(E)$  à une partie de  $\mathcal{C}_b(E)'$ . Il y a toutefois peu d'espoir que cette application soit injective sans autre hypothèse sur  $E$  que le fait qu'il est un espace topologique. Nous allons montrer qu'elle l'est si  $E$  est un espace métrique.

Nous allons saisir cette occasion pour démontrer un lemme qui nous sera souvent utile.

**Définition 1.2.4** (Distance à une partie). Soit  $(E, d)$  un espace métrique. Soit  $A$  une partie de  $E$ . On note, pour tout  $x \in E$ ,

$$d(x, A) = \inf\{d(x, y) : y \in A\}.$$

**Exercice 1.6.** Soit  $A$  une partie d'un espace métrique  $(E, d)$ . Montrer que la fonction  $x \mapsto d(x, A)$  est 1-lipschitzienne sur  $E$  et que pour tout  $x \in E$ , on a l'équivalence

$$d(x, A) = 0 \iff x \in \overline{A}.$$

La démonstration du lemme suivant est laissée en exercice.

**Lemme 1.2.5** (Approximation d'indicatrices). Soit  $(E, d)$  un espace métrique.

1. Soit  $F$  une partie fermée de  $E$ . Pour tout  $n \geq 1$ , on définit  $f_n \in \mathcal{C}_b(E)$  en posant, pour tout  $x \in E$ ,

$$f_n(x) = \max(1 - n d(x, F), 0).$$

La suite  $(f_n)_{n \geq 0}$  est décroissante et converge simplement vers l'indicatrice de  $F$ .

2. Soit  $G$  une partie ouverte de  $E$ . Pour tout  $n \geq 1$ , on définit  $g_n \in \mathcal{C}_b(E)$  en posant, pour tout  $x \in E$ ,

$$g_n(x) = \min(n d(x, G^c), 1).$$

La suite  $(g_n)_{n \geq 0}$  est croissante et converge simplement vers l'indicatrice de  $G$ .

De ce lemme, nous allons déduire la proposition suivante.

**Proposition 1.2.6.** *Soit  $(E, d)$  un espace métrique. Si deux mesures de probabilité  $\mu$  et  $\nu$  sont telles que pour tout  $f \in \mathcal{C}_b(E)$  on ait  $\mu(f) = \nu(f)$ , alors  $\mu = \nu$ .*

*En d'autres termes, l'application  $\mathcal{M}(E) \rightarrow \mathcal{C}_b(E)'$  qui à  $\mu$  associe la forme linéaire  $f \mapsto \mu(f)$  est injective.*

*Démonstration.* Soit  $G$  un ouvert de  $E$  et  $(g_n)_{n \geq 0}$  la suite de fonctions construite au lemme 1.2.5. Par hypothèse, on a pour tout  $n \geq 0$  l'égalité  $\mu(g_n) = \nu(g_n)$ . Lorsque  $n$  tend vers l'infini, le théorème de convergence monotone nous permet d'affirmer que  $\mu(g_n)$  tend vers  $\mu(\mathbb{1}_G) = \mu(G)$ , et que  $\nu(g_n)$  tend vers  $\nu(\mathbb{1}_G) = \nu(G)$ . Ainsi,  $\mu(G) = \nu(G)$ . La classe

$$\mathcal{E} = \{B \in \mathcal{B}_E : \mu(B) = \nu(B)\}$$

contient donc la classe  $\mathcal{O}$  de toutes les parties ouvertes de  $E$ . C'est par ailleurs une classe monotone (ou  $\lambda$ -système). Puisque la classe  $\mathcal{O}$  des parties ouvertes de  $E$  est un  $\pi$ -système, le théorème de la classe monotone assure que  $\mathcal{E}$  contient la tribu engendrée par  $\mathcal{O}$ , c'est-à-dire que  $\mathcal{E}$  contient  $\mathcal{B}_E$ . Ainsi,  $\mu$  et  $\nu$  coïncident sur  $\mathcal{B}_E$ , et sont donc égales.  $\square$

Cette proposition affirme qu'un élément  $\mu \in \mathcal{M}(E)$  est déterminé par les intégrales  $\int_E f \, d\mu$ , où  $f$  décrit l'ensemble des fonctions continues bornées sur  $E$ . Puisque l'argument de la démonstration ne fait intervenir que des fonctions lipschitziennes, on voit qu'il suffit, pour caractériser  $\mu$ , de connaître  $\int_E f \, d\mu$  pour toute fonction  $f$  lipschitzienne sur  $E$ .

D'un point de vue plus abstrait, cette proposition nous permet, pour un espace métrique  $(E, d)$ , d'identifier  $\mathcal{M}(E)$  avec une partie de l'espace de Banach  $\mathcal{C}_b(E)'$ . Dans la section suivante, nous allons le munir de la trace de la topologie faible-\*, bien que les probabilistes préfèrent parler de topologie faible, ou de convergence étroite.

Avant cela, nous allons démontrer une autre propriété fondamentale, dite de régularité, qui illustre la manière dont interagissent les structures métrique et mesurable.

Commençons par un point de terminologie. Soit  $(E, d)$  un espace métrique muni de sa tribu borélienne  $\mathcal{B}$ . Soit  $\mu$  une mesure sur  $(E, \mathcal{B})$ . On s'intéresse parfois, et nous nous intéresserons, aux trois assertions suivantes :

$$\forall B \in \mathcal{B}, \mu(B) = \inf\{\mu(O) : O \text{ partie ouverte de } E \text{ contenant } B\} \quad (R_1)$$

$$\forall B \in \mathcal{B}, \mu(B) = \sup\{\mu(F) : F \text{ partie fermée de } E \text{ incluse dans } B\} \quad (R_2)$$

$$\forall B \in \mathcal{B}, \mu(B) = \sup\{\mu(K) : K \text{ partie compacte de } E \text{ incluse dans } B\} \quad (R_3)$$

Puisque toute partie compacte d'un espace métrique est fermée<sup>5</sup>,  $(R_3)$  implique  $(R_2)$ .

5. Cette assertion est moins simple qu'elle n'en a l'air. Tout d'abord, dire qu'une partie, disons  $A$ , de  $E$  est fermée signifie que  $A$  appartient à un certain sous-ensemble de  $\mathcal{P}(E)$ , en l'occurrence à l'ensemble des parties fermées de  $E$ . Dire que  $A$  est compacte signifie que l'espace métrique  $(A, d|_{A \times A})$  est un espace métrique compact. Ces assertions ne sont donc pas exactement de la même nature.

Par ailleurs, c'est peut-être une bonne occasion de souligner que les traditions française et anglo-saxonne diffèrent sur le sens du mot *compact*. En français, un espace topologique est compact s'il vérifie la propriété de Borel-Lebesgue (c'est-à-dire que de tout recouvrement par des ouverts on peut extraire un sous-recouvrement fini) et s'il est *séparé* (au sens de Hausdorff, c'est-à-dire au sens où deux points distincts quelconques admettent deux voisinages respectifs disjoints). En anglais, on n'impose pas la propriété de séparation.

Puisque tout espace métrique est séparé, cette distinction ne doit pas nous inquiéter. Profitons tout de même de

Si  $\mu$  est une mesure finie, alors en passant au complémentaire, il n'est pas difficile de voir que  $(R_1)$  et  $(R_2)$  sont équivalentes. Toujours si  $\mu$  est finie, il n'est pas difficile de voir que  $(R_1)$  et  $(R_2)$  sont équivalentes à

$$\forall B \in \mathcal{B}, \forall \varepsilon > 0, \exists F \text{ fermé}, \exists O \text{ ouvert}, F \subseteq B \subseteq O \text{ et } \mu(O \setminus F) < \varepsilon. \quad (R_{12})$$

**Proposition 1.2.7.** *Soit  $(E, d)$  un espace métrique. Soit  $\mu \in \mathcal{M}(E)$  une mesure borélienne de probabilité sur  $E$ . Alors pour tout borélien  $B$ , on a*

$$\mu(B) = \inf\{\mu(O) : O \text{ partie ouverte de } E \text{ contenant } B\}.$$

On dit que  $\mu$  est extérieurement régulière. De plus,

$$\mu(B) = \sup\{\mu(F) : F \text{ partie fermée de } E \text{ incluse dans } B\}.$$

La propriété  $(R_3)$ , sur laquelle nous reviendrons, s'appelle parfois la régularité intérieure. La proposition ci-dessus n'affirme pas que toute mesure de probabilité sur un espace métrique soit régulière intérieurement. Les contre-exemples à cette assertion semblent toutefois nécessiter des constructions un peu exotiques de théorie des ensembles. Nous démontrerons de toute façon que sur un espace polonais, toute mesure de probabilité est régulière extérieurement et intérieurement.

*Démonstration.* Commençons par observer que les deux assertions, d'approximation extérieure par des ouverts et intérieure par des fermés, sont équivalentes. En effet, la régularité extérieure de la mesure peut s'exprimer en disant que pour tout borélien  $B$  et tout  $\varepsilon > 0$ , il existe un ouvert  $O$  contenant  $B$  et tel que  $\mu(O) < \mu(B) + \varepsilon$ . Si cette propriété a lieu, en l'appliquant au complémentaire de  $B$ , on voit qu'il existe aussi un fermé  $F$  inclus dans  $B$  tel que  $\mu(F) > \mu(B) - \varepsilon$ , ce qui donne la deuxième assertion. Une manière symétrique d'exprimer la régularité extérieure est la suivante : pour tout borélien  $B$  et tout  $\varepsilon > 0$ , il existe un ouvert  $O$  et un fermé  $F$  tels que<sup>6</sup>

$$F \subseteq B \subseteq O \text{ et } \mu(O) - \mu(F) < \varepsilon.$$

Considérons la classe

$$\mathcal{A} = \{B \in \mathcal{B}_E : \forall \varepsilon > 0, \exists F \text{ fermé}, \exists O \text{ ouvert}, F \subseteq B \subseteq O \text{ et } \mu(O) - \mu(F) < \varepsilon\}.$$

cette note pour donner deux exemples d'énoncés un peu surprenants.

En France, l'énoncé « l'image d'un espace topologique compact par une application continue est un compact » est faux. En effet, s'il est vrai que cette image a nécessairement la propriété de Borel–Lebesgue, il n'est pas vrai en général qu'elle soit séparée. Par exemple, considérons une droite  $D$  de pente irrationnelle dans  $\mathbb{R}^2$ . On peut former le groupe additif  $\mathbb{R}^2/\mathbb{Z}^2$ , qui est géométriquement un tore, et considérer l'image de l'application composée  $D \rightarrow \mathbb{R}^2 \rightarrow \mathbb{R}^2/\mathbb{Z}^2$ . Cette application est injective, et son image, que nous notons encore  $D$ , est un sous-groupe de notre tore ; c'est géométriquement une droite qui s'y enroule sans jamais se refermer, et qui y est dense. L'application quotient  $\mathbb{R}^2/\mathbb{Z}^2 \rightarrow (\mathbb{R}^2/\mathbb{Z}^2)/D$  est continue pour la topologie quotient sur l'espace d'arrivée ; mais cette topologie quotient n'est pas séparée, car c'est la topologie grossière sur un espace qui a plus qu'un point.

Par ailleurs, dans le monde anglo-saxon, l'énoncé « a compact subset of a topological space is closed » est faux. Considérons par exemple  $\mathbb{R}$  muni de la topologie de Zariski, dont les parties fermées autres que  $\mathbb{R}$  sont exactement les parties finies. Cet espace a la propriété de Borel–Lebesgue. En fait, toute partie de  $\mathbb{R}$  a la propriété de Borel–Lebesgue, par exemple  $\mathbb{Z}$ , mais  $\mathbb{Z}$  n'est pas une partie fermée de  $\mathbb{R}$ .

6. Ainsi considérée, cette régularité n'a plus grand-chose d'extérieur. La terminologie dans ce domaine n'est pas parfaite, ni complètement universelle.

Nous allons montrer que c'est une tribu qui contient tous les ouverts. Commençons par montrer que tout ouvert appartient à  $\mathcal{A}$ .

Soit  $O$  un ouvert de  $E$ . Donnons-nous  $\varepsilon > 0$ . Pour tout  $n \geq 1$ , considérons

$$F_n = \{x \in E : d(x, E \setminus O) \geq \frac{1}{n}\}.$$

Il s'agit d'un fermé inclus dans  $O$ . La suite  $(F_n)_{n \geq 1}$  est croissante pour l'inclusion, et on a

$$O = \bigcup_{n=1}^{\infty} F_n,$$

si bien que la suite  $(\mu(F_n))_{n \geq 1}$  tend en croissant vers  $\mu(O)$  et qu'il existe donc un entier  $N \geq 1$  tel que  $\mu(F_N) > \mu(O) - \varepsilon$ . Ainsi,  $O$  appartient à  $\mathcal{A}$ .

Montrons maintenant que  $\mathcal{A}$  est une tribu. Le fait que  $\mathcal{A}$  contienne  $E$  et soit stable par passage au complémentaire découlent immédiatement de sa définition. Montrons qu'elle est stable par union dénombrable.

Considérons donc une suite  $(B_n)_{n \geq 1}$  d'éléments de  $\mathcal{A}$  et notons  $B = \bigcup_{n=1}^{\infty} B_n$ . Donnons-nous  $\varepsilon > 0$ . Pour tout  $n \geq 1$ , il existe, et nous choisissons, un fermé  $F_n$  et un ouvert  $O_n$  tels que  $F_n \subseteq B_n \subseteq O_n$  et  $\mu(O_n) - \mu(F_n) < \frac{\varepsilon}{2^{n+1}}$ . Notons  $O = \bigcup_{n=1}^{\infty} O_n$  et  $F = \bigcup_{n=1}^{\infty} F_n$ . On a

$$F \subseteq B \subseteq O \text{ et } \mu(O) - \mu(F) = \mu(O \setminus F) \leq \sum_{n=1}^{\infty} \mu(O_n \setminus F_n) < \frac{\varepsilon}{2}.$$

De plus,  $O$  est un ouvert. Cependant, il nous faut prendre garde au fait que  $F$ , qui est une réunion dénombrable de fermés, n'est pas nécessairement fermé. Nous observons alors que la suite d'ensembles  $(\bigcup_{n=1}^p F_n)_{p \geq 1}$  est croissante et d'union égale à  $F$ , si bien qu'il en existe un terme dont la mesure diffère de celle de  $F$  de moins de  $\frac{\varepsilon}{2}$ . Il existe donc un entier  $N \geq 1$  tel qu'en posant  $F' = \bigcup_{n=1}^N F_n$ , on ait bien  $F'$  fermé,  $O$  ouvert, et

$$F' \subseteq B \subseteq O \text{ et } \mu(O) - \mu(F') < \varepsilon.$$

Ceci montre que  $\mathcal{A}$  est une tribu, donc la tribu borélienne de  $E$  tout entière.  $\square$

On peut se demander pourquoi on cherche à approcher un borélien par l'extérieur par un ouvert, et par l'intérieur par un fermé, et non le contraire, qui pourrait sembler plus naturel vu l'importance des notions d'adhérence et d'intérieur. Mais précisément, si l'on cherchait à approcher un borélien par un ouvert à l'intérieur et par un fermé à l'extérieur, on saurait tout de suite quel est le meilleur choix possible : ce serait  $\overset{\circ}{B} \subseteq B \subseteq \overline{B}$ . Or, par exemple, le cas de  $B = [0, 1] \cap \mathbb{Q}$  dans l'espace métrique  $[0, 1]$  muni de la distance usuelle et de la mesure de Lebesgue montre qu'on n'a pas du tout, en général,  $\mu(\overline{B}) = \mu(\overset{\circ}{B})$ .

**Exercice 1.7.** Montrer par un exemple qu'on peut avoir  $\mu(\overset{\circ}{B}) < \mu(B) < \mu(\overline{B})$ .

Remarquons enfin que la distance n'intervient dans cet énoncé que pour donner une topologie à  $E$ . Le résultat reste vrai sur tout espace topologique métrisable, c'est-à-dire dont la topologie est telle qu'il existe une distance sur  $E$  qui l'induit.<sup>7</sup>

7. Un théorème d'Urysohn donne une condition nécessaire et suffisante pour la métrisabilité d'une topologie.

### 1.3 Convergence faible des mesures de probabilité

**Définition 1.3.1** (Convergence faible). Soit  $(E, d)$  un espace métrique. Soit  $(\mu_n)_{n \geq 0}$  une suite d'éléments de  $\mathcal{M}(E)$ , c'est-à-dire une suite de mesures de probabilités boréliennes sur  $E$ . Soit  $\mu \in \mathcal{M}(E)$ . On dit que la suite  $(\mu_n)_{n \geq 0}$  converge étroitement, ou faiblement vers  $\mu$ , si

$$\forall f \in C_b(E), \lim_{n \rightarrow \infty} \int_E f \, d\mu_n = \int_E f \, d\mu.$$

Lorsque cette convergence a lieu, on écrit

$$\mu_n \xrightarrow[n \rightarrow \infty]{} \mu.$$

Remarquons que la limite faible d'une suite de mesures de probabilités sur  $E$ , si elle existe, est unique : en effet, deux limites faibles d'une même suite donnent la même intégrale à toute fonction continue bornée sur  $E$ , et sont donc, d'après la proposition 1.2.6, égales.

Notre premier résultat va consister à donner plusieurs caractérisations de la convergence faible d'une suite de mesures de probabilité. Ce résultat est vrai sur un espace métrique quelconque.

Rappelons qu'on dit qu'une fonction  $f$  à valeurs réelles sur un espace métrique  $(E, d)$  est *uniformément continue* si

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

Nous noterons  $\mathcal{U}_b(E)$  le sous-espace vectoriel de  $C_b(E)$  formé des fonctions uniformément continues et bornées.

**Exercice 1.8.** Vérifier que  $\mathcal{U}_b(E)$  est un sous-espace vectoriel fermé de  $C_b(E)$ . Autrement dit, une limite uniforme de fonctions uniformément continues est uniformément continue.

Rappelons que si  $(E, d)$  est compact, alors toute fonction continue sur  $E$  est bornée et uniformément continue, si bien que les espaces  $\mathcal{C}(E)$ ,  $C_b(E)$  et  $\mathcal{U}_b(E)$  coïncident.

Introduisons une dernière notation. Pour toute partie  $A$  de  $E$ , on note

$$\partial A = \overline{A} \setminus \overset{\circ}{A}$$

et on appelle cette partie la *frontière* de  $A$ .

**Exercice 1.9.** Vérifier que  $\partial A = \overline{A} \cap \overline{E \setminus A}$  et que c'est une partie fermée de  $E$ . Montrer que si  $A$  est une partie fermée, alors  $\partial A$  est d'intérieur vide. Montrer que si  $A$  est fermée et d'intérieur vide, alors  $\partial A = A$ . En déduire que pour toute partie  $A$  on a l'égalité  $\partial \partial \partial A = \partial \partial A$ , mais montrer par un contre-exemple qu'on n'a pas toujours  $\partial \partial A = \partial A$ .

**Théorème 1.3.2** (Théorème de Portmanteau<sup>8</sup>). Soit  $(E, d)$  un espace métrique. Soient respectivement  $(\mu_n)_{n \geq 0}$  et  $\mu$  une suite d'éléments et un élément de  $\mathcal{M}(E)$ . Les assertions suivantes sont équivalentes.

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

---

8. Ce nom étrange, dont l'origine est un canular d'Alexandrov ou de Billingsley, est entré dans l'usage.



2. Pour tout  $f \in \mathcal{U}_b(E)$ ,  $\lim_{n \rightarrow \infty} \int_E f \, d\mu_n = \int_E f \, d\mu$ .
3. Pour toute partie ouverte  $G$  de  $E$ , on a  $\mu(G) \leq \underline{\lim} \mu_n(G)$ .
4. Pour toute partie fermée  $F$  de  $E$ , on a  $\mu(F) \geq \overline{\lim} \mu_n(F)$ .
5. Pour toute partie borélienne  $A$  telle que  $\mu(\partial A) = 0$ , on a  $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ .

La démonstration de ce théorème n'est pas très difficile. Le problème principal avec cet énoncé, comme par exemple avec le lemme de Fatou (voir Théorème B.1 et, d'une manière générale, l'annexe B si vous n'êtes pas trop à l'aise avec les notions de limite inférieure et supérieure), est de se le rappeler, en particulier des assertions 3 et 4. Comme pour le lemme de Fatou, une bonne manière de limiter le risque d'erreur est de retenir un exemple simple où la suite  $(\mu_n(A))_{n \geq 1}$  ne converge pas vers  $\mu(A)$ . On peut prendre, sur  $\mathbb{R}$ , l'ouvert  $G = \mathbb{R}_+$ , le fermé  $F = \mathbb{R}_-$  et, pour tout  $n \geq 1$ ,  $\mu_n = \delta_{\frac{1}{n}}$ .

*Démonstration.* L'implication 1  $\Rightarrow$  2 est une conséquence de la définition de la convergence faible et de l'inclusion  $\mathcal{U}_b(E) \subseteq \mathcal{C}_b(E)$ .

L'implication 2  $\Rightarrow$  3 est une conséquence de la deuxième assertion du lemme 1.2.5. En effet, la suite de fonctions  $(g_k)_{k \geq 0}$  construite dans ce lemme est une suite de fonctions uniformément continues : pour chaque  $k \geq 0$ , la fonction  $g_k$  est  $k$ -lipschitzienne. Or nous pouvons écrire, pour tout  $k \geq 0$ , et en style un peu télégraphique,

$$\underline{\lim}_{n \rightarrow \infty} \mu_n(G) \geq \underline{\lim}_{n \rightarrow \infty} \int_E g_k \, d\mu_n = \int_E g_k \, d\mu \xrightarrow[k \rightarrow \infty]{} \mu(G).$$

On peut déduire 4 de 2 de la même manière que nous avons déduit 3 de 2, en utilisant le lemme 1.2.5. On peut aussi, et c'est utile, observer que les assertions 3 et 4 sont équivalentes en passant au complémentaire.

Montrons que (3 et 4)  $\Rightarrow$  5. Soit  $A$  un borélien de  $E$  tel que  $\mu(\partial A) = 0$ . Les inclusions  $\mathring{A} \subseteq A \subseteq \overline{A}$  entraînent, en toute généralité, les égalités  $\mu(\mathring{A}) \leq \mu(A) \leq \mu(\overline{A})$ . L'hypothèse  $\mu(\partial A) = 0$  s'écrit

$$0 = \mu(\partial A) = \mu(\overline{A} \setminus \mathring{A}) = \mu(\overline{A}) - \mu(\mathring{A})$$

et on en déduit que  $\mu(\overline{A}) = \mu(A) = \mu(\mathring{A})$ .

On a par ailleurs, pour tout  $n \geq 0$ ,  $\mu_n(A) \leq \mu_n(\overline{A})$  et  $\mu_n(\mathring{A}) \leq \mu_n(A)$ . En passant aux limites supérieures dans la première inégalité et aux limites inférieures dans la seconde, et en appliquant les assertions 3 et 4, nous trouvons

$$\overline{\lim}_{n \rightarrow \infty} \mu_n(A) \leq \overline{\lim}_{A \subseteq \overline{A}} \mu_n(\overline{A}) \leq \mu(\overline{A}) \stackrel{\mu(\partial A)=0}{=} \mu(\mathring{A}) \leq \underline{\lim}_{A \subseteq \mathring{A}} \mu_n(\mathring{A}) \leq \underline{\lim}_{A \subseteq A} \mu_n(A).$$

La suite  $(\mu_n(A))_{n \geq 0}$  a une limite supérieure inférieure ou égale à sa limite inférieure. Les deux sont donc égales. La suite converge donc vers leur valeur commune, qui est  $\mu(\overline{A}) = \mu(\mathring{A})$ , et qui est aussi  $\mu(A)$ .

Montrons enfin l'implication 5  $\Rightarrow$  1. Soit  $f$  une fonction continue et bornée sur  $E$ . Quitte à composer  $f$  par une transformation affine, c'est-à-dire quitte à remplacer  $f$  par  $af + b$  avec des constantes  $a$  et  $b$  bien choisies, ce qui ne changera rien à la convergence de la suite  $(\mu_n(f))_{n \geq 0}$ , nous pouvons supposer que  $f$  prend ses valeurs dans l'intervalle ouvert  $]0, 1[$ .

Soit  $\nu$  un élément arbitraire de  $\mathcal{M}(E)$ . Le théorème de Fubini nous permet d'écrire

$$\int_E f(x) \, d\nu(x) = \int_{E \times [0,1]} \mathbb{1}_{t \leq f(x)} \, dt \, d\nu(x) = \int_0^1 \nu(\{f \geq t\}) \, dt.$$

Ainsi, pour tout  $n \geq 0$ ,

$$\mu_n(f) = \int_0^1 \mu_n(\{f \geq t\}) \, dt.$$

Nous allons maintenant démontrer que la fonction  $t \mapsto \mu_n(\{f \geq t\})$  converge simplement, ou du moins pour presque tout  $t \in [0, 1]$ , en étant dominée par la constante 1, vers la fonction  $t \mapsto \mu(\{f \geq t\})$  et appliquer le théorème de convergence dominée.

L'assertion 5 assure la convergence de  $\mu_n(\{f \geq t\})$  vers  $\mu(\{f \geq t\})$  pour tout  $t \in [0, 1]$  tel que  $\mu(\partial\{f \geq t\}) = 0$ . Il suffit donc de montrer que l'ensemble des  $t \in [0, 1]$  tels que  $\mu(\partial\{f \geq t\}) > 0$  est négligeable. Nous allons montrer qu'il est dénombrable. En effet, pour tout  $t \in [0, 1]$ , nous avons l'inclusion

$$\partial\{f \geq t\} \subseteq \{f = t\}.$$

En effet, un point de  $\partial\{f \geq t\}$  est à la fois limite d'une suite de points où  $f$  prend des valeurs supérieures à  $t$ , et limite d'une suite de points où  $f$  prend des valeurs strictement inférieures à  $t$ . Les ensembles  $\{\partial\{f \geq t\}, t \in [0, 1]\}$  sont donc des boréliens deux à deux disjoints de  $E$ . La mesure  $\mu$ , qui est finie, ne peut donner une mesure strictement positive qu'à un sous-ensemble dénombrable de l'ensemble de ces ensembles :

$$\{t \in [0, 1] : \mu(\{f = t\}) > 0\} \text{ est au plus dénombrable.}$$

Le résultat s'ensuit, comme nous l'avons annoncé, par une application du théorème de convergence dominée.  $\square$

Revenons sur le dernier point de la démonstration.

**Exercice 1.10.** Soit  $(B_i)_{i \in I}$  une collection de boréliens deux à deux disjoints de  $E$ . Soit  $\mu$  un élément de  $\mathcal{M}(E)$ . Montrer que

$$\{i \in I : \mu(B_i) > 0\}$$

est au plus dénombrable. On pourra commencer par réfléchir au fait que  $\{i \in I : \mu(B_i) \geq \frac{1}{10}\}$  a au plus 10 éléments.

Avant de réfléchir à l'exercice suivant, il peut être utile d'étudier l'annexe C.

**Exercice 1.11.** Montrer que dans le théorème de Portmanteau, on peut ajouter les caractérisations suivantes de la convergence faible.

2.5. Pour tout fonction  $f$  s.c.i. et bornée sur  $E$ , on a  $\mu(f) \leq \underline{\lim}_{n \rightarrow \infty} \mu_n(f)$ .

3.5. Pour tout fonction  $f$  s.c.s. et bornée sur  $E$ , on a  $\mu(f) \geq \overline{\lim}_{n \rightarrow \infty} \mu_n(f)$ .

## 1.4 L'espace topologique $\mathcal{M}(E)$

Nous supposons désormais que  $(E, d)$  est un espace métrique. Nous avons mentionné, à la suite de la proposition 1.2.6, que  $\mathcal{M}(E)$  pouvait être identifié à un sous-ensemble de l'espace dual de l'espace de Banach  $(\mathcal{C}_b(E), \|\cdot\|)$ , et que nous allons le munir de la trace de la topologie faible-\* sur cet espace dual.

Cependant, nous n'avons pas défini sur  $\mathcal{M}(E)$  une topologie, mais une notion de convergence de suites. Nous allons voir que cette notion de convergence est bien celle associée à la topologie faible-\*, discuter certaines propriétés de cette topologie, en particulier sa séparabilité et sa compacité, et indiquer qu'elle peut être induite par une distance sur  $\mathcal{M}(E)$ .

### 1.4.1 La topologie faible-\*

Rappelons ce qu'est la topologie faible-\* sur le dual d'un espace de Banach. Soit donc  $B$  un espace de Banach et  $B'$  l'espace dual des formes linéaires continues sur  $B$ . La norme

$$\|\phi\| = \sup\{\phi(x) : x \in B, \|x\| \leq 1\}$$

fait de  $B'$  un espace de Banach. La topologie induite par cette norme sur  $B'$  est toutefois trop fine pour beaucoup d'usages : la notion de convergence de suites qui y est associée est trop exigeante. Il existe sur  $B'$  des topologies moins fines et très utiles, en particulier la topologie faible-\*, qui est la topologie la plus grossière qui rende continue toutes les applications  $\phi \mapsto \phi(x)$  où  $x$  parcourt  $B$ . On peut décrire cette topologie assez explicitement, comme suit. Pour tous  $x \in B$ ,  $a \in \mathbb{R}$  et  $\varepsilon \in \mathbb{R}_+$ , posons

$$U(x, a, \varepsilon) = \{\phi \in B' : |\phi(x) - a| < \varepsilon\}. \quad (1.1)$$

Cette partie de  $B'$ , qui est l'image réciproque de l'ouvert  $]a - \varepsilon, a + \varepsilon[$  par l'application  $\phi \mapsto \phi(x)$ , est une partie ouverte pour toute topologie qui rend cette application continue. C'est donc un ouvert faible-\*.

Considérons la classe des parties de  $B'$  de la forme  $U(x, a, \varepsilon)$  et rendons-la stable par intersections finies, en lui ajoutant toutes les intersections finies. On obtient la classe, que nous noterons  $\mathcal{U}$ , des parties de la forme

$$U(x_1, a_1, \varepsilon_1) \cap \dots \cap U(x_p, a_p, \varepsilon_p)$$

que nous appellerons des ouverts basiques. Formons maintenant la classe  $\mathcal{T}$  des unions quelconques d'ouverts basiques.

**Lemme 1.4.1.** *La classe  $\mathcal{T}$  est une topologie sur  $B'$ .*

*Démonstration.* Notons que  $\emptyset = U(0, 1, 1)$  et  $B' = U(0, 0, 1)$  sont des ouverts basiques. Par construction, la classe  $\mathcal{T}$  est stable par unions quelconques. Il faut montrer qu'elle est stable par intersections finies. Considérons deux éléments  $\bigcup_{i \in I} U_i$  et  $\bigcup_{j \in J} V_j$  deux réunions arbitraires d'ouverts basiques. L'égalité

$$\bigcup_{i \in I} U_i \cap \bigcup_{j \in J} V_j = \bigcup_{(i,j) \in I \times J} U_i \cap V_j$$

montre que leur intersection est encore une réunion d'ouverts basiques. Ainsi,  $\mathcal{T}$  est une topologie sur  $B'$ .  $\square$

Notons que par construction, la classe  $\mathcal{U}$  des ouverts basiques est une base de la topologie  $\mathcal{T}$ . Nous avons observé que toute topologie qui rendait les applications  $\phi \mapsto \phi(x)$  continues contenait les parties de la forme  $U(x, a, \varepsilon)$ , donc les ouverts basiques, donc la topologie  $\mathcal{T}$ . La topologie  $\mathcal{T}$  est donc moins fine que (c'est-à-dire incluse dans) la topologie faible-\*.

Remarquons maintenant que puisque tout ouvert de  $\mathbb{R}$  est une réunion d'intervalles ouverts, l'image réciproque d'un tel ouvert par l'application  $\phi \mapsto \phi(x)$  (où  $x \in B$  est donné) est une union de parties de la forme  $U(x, a, \varepsilon)$  pour des  $a$  et  $\varepsilon$  bien choisis, et en particulier un élément de  $\mathcal{T}$ . Ainsi, cette application  $\phi \mapsto \phi(x)$  est continue par rapport à  $\mathcal{T}$ . Finalement,  $\mathcal{T}$  est bien la topologie la moins fine sur  $B'$  qui rende continues toutes les applications  $\phi \mapsto \phi(x)$ ,  $x \in B$ . C'est la topologie faible-\*.

**Lemme 1.4.2.** *Une suite  $(\phi_n)_{n \geq 0}$  d'éléments de  $B'$  converge vers  $\phi \in B'$  pour la topologie faible-\* si et seulement si pour tout  $x \in B$  la suite  $(\phi_n(x))_{n \geq 0}$  converge vers  $\phi(x)$ .*

*Démonstration.* Soit  $(\phi_n)_{n \geq 0}$  une suite d'éléments de  $B'$  et  $\phi \in B'$ . Supposons tout d'abord que la suite  $(\phi_n)_{n \geq 0}$  converge vers  $\phi$  relativement à la topologie  $\mathcal{T}$ . Considérons  $x \in B$  et posons  $a = \phi(x)$ . Pour tout  $\varepsilon > 0$ , il existe un  $n_0$  tel que pour tout  $n \geq n_0$ , la forme linéaire  $\phi_n$  appartienne au voisinage  $U(x, a, \varepsilon)$  de  $\phi$ , ce qui signifie que  $|\phi_n(x) - \phi(x)| < \varepsilon$ . Ceci montre que la suite  $(\phi_n(x))_{n \geq 0}$  converge vers  $\phi(x)$ .<sup>9</sup>

Réciproquement, supposons que pour tout  $x \in B$ , la suite  $(\phi_n(x))_{n \geq 0}$  converge vers  $\phi(x)$ . Soit  $V$  un voisinage de  $\phi$  pour la topologie faible-\*. Il existe un ouvert basique contenant  $\phi$  et inclus dans  $V$ , et cet ouvert basique est de la forme  $U = U(x_1, a_1, \varepsilon_1) \cap \dots \cap U(x_p, a_p, \varepsilon_p)$ . Soit  $\varepsilon > 0$  tel que pour tout  $i \in \{1, \dots, p\}$ , on ait l'inclusion

$$] \phi(x_i) - \varepsilon, \phi(x_i) + \varepsilon [ \subseteq ] a_i - \varepsilon_i, a_i + \varepsilon_i [.$$

Soit  $n_0$  tel que pour tout  $n \geq n_0$  et tout  $i \in \{1, \dots, p\}$  on ait  $|\phi_n(x_i) - \phi(x_i)| < \varepsilon$ . Alors pour  $n \geq n_0$ , la forme linéaire  $\phi_n$  appartient à  $U$ , donc à  $V$ . Ainsi, la suite  $(\phi_n)_{n \geq 0}$  converge vers  $\phi$  au sens de la topologie faible-\*.  $\square$

Nous venons de décrire une situation très générale. Le cas qui nous intéresse est celui d'un espace métrique  $(E, d)$  et de l'espace  $B = \mathcal{C}_b(E)$ . Dans ce cas, il découle de la proposition 1.2.6 que l'application  $\mu \mapsto (f \mapsto \int_E f d\mu)$  identifie  $\mathcal{M}(E)$  à une partie de  $B'$ . Nous munissons  $\mathcal{M}(E)$  de la topologie induite par la topologie faible-\*, c'est-à-dire de la topologie dont les ouverts sont les intersections des ouverts faible-\* avec  $\mathcal{M}(E)$ . Il découle de ce qui précède que la notion de convergence de suites pour cette topologie est exactement la convergence faible.<sup>10</sup>

9. Nous aurions aussi pu dire que puisque l'évaluation en  $x$  est continue par rapport à la topologie faible-\*, la suite  $(\phi_n(x))_{n \geq 0}$  converge vers  $\phi(x)$ .

10. Une topologie n'est pas, en général, déterminée par l'ensemble de ses suites convergentes. Il est toutefois vrai que si deux topologies à bases dénombrables de voisinages sur un même ensemble ont les mêmes suites convergentes, alors elles sont égales. Être à bases dénombrables de voisinages signifie pour une topologie que pour tout point, il existe une suite de voisinages de ce point telle que tout voisinage de ce point contienne un terme de cette suite.

Avant de nous lancer dans une étude plus détaillée de l'espace topologique  $\mathcal{M}(E)$  (qui sera toujours, dans ce qui suit, muni de la topologie que nous venons de décrire, c'est-à-dire la trace de la topologie faible- $*$ ), nous pouvons noter que c'est un espace topologique séparé. En effet, si  $\mu$  et  $\nu$  sont deux mesures distinctes, il existe, d'après la proposition 1.2.6, une fonction continue bornée  $f$  sur  $E$  telle que  $\int_E f \, d\mu \neq \int_E f \, d\nu$ . Notons respectivement  $a$  et  $b$  ces deux intégrales. Alors les ouverts

$$\left\{ \pi \in \mathcal{M}(E) : \left| \int_E f \, d\pi - a \right| < \frac{|b-a|}{2} \right\} \text{ et } \left\{ \pi \in \mathcal{M}(E) : \left| \int_E f \, d\pi - b \right| < \frac{|b-a|}{2} \right\}$$

sont disjoints et contiennent respectivement  $\mu$  et  $\nu$ .

Voici maintenant quelques questions que nous pourrions nous poser à propos de  $\mathcal{M}(E)$ .

1. La topologie de  $\mathcal{M}(E)$  est-elle métrisable? Si oui, il s'ensuivra par exemple que c'est la seule topologie qui ait pour suites convergentes les suites faiblement convergentes.
2. Comment les propriétés de l'espace métrique  $(E, d)$  déterminent-elles celles de  $\mathcal{M}(E)$ ? Si  $E$  est précompact (nous y reviendrons), ou compact, ou connexe, ou complet, quelles conséquences cela a-t-il sur  $\mathcal{M}(E)$ ?<sup>11</sup>
3. Réciproquement, comment les propriétés de  $\mathcal{M}(E)$  nous renseignent-elles sur  $E$ ?

En général, dans un espace topologique, on peut définir, entre une partie  $A$  et son adhérence  $\bar{A}$ , une partie intermédiaire qui est l'adhérence séquentielle de  $A$ , et qui est l'ensemble des limites des suites convergentes d'éléments de  $A$ . On peut la noter  $\bar{A}$  et on vérifie qu'on a toujours les inclusions  $A \subseteq \bar{A} \subseteq \bar{A}$ . Pour une topologie à bases dénombrables de voisinages, l'adhérence séquentielle coïncide avec l'adhérence. Ainsi, si deux topologies à bases dénombrables de voisinages ont les mêmes suites convergentes, elles ont la même notion d'adhérence séquentielle, donc la même notion d'adhérence, donc les mêmes parties fermées (celles qui sont égales à leur adhérence), donc elles sont égales.

Un exemple simple de deux topologies distinctes ayant les mêmes suites convergentes est le suivant : sur un ensemble non dénombrable, par exemple  $\mathbb{R}$ , on peut considérer d'une part la topologie discrète (pour qui toutes les parties sont ouvertes) et la topologie co-dénombrable, dont les ouverts non vides sont les parties de complémentaire au plus dénombrable.

Un exemple plus naturel, et où les topologies sont de plus séparées, mais aussi un exemple un peu moins simple à étudier (et dû à Schur) est celui des topologies normique et faible sur l'espace  $\ell^1$  des suites absolument sommables.

Dans la veine du premier exemple, et à titre d'amusement, on peut se demander ce que sont les suites convergentes de la topologie co-finie de  $\mathbb{R}$  (auss appelé topologie de Zariski), dont les ouverts non vides sont les parties de complémentaire fini. La réponse semble être la suivante : sont convergentes les suites qui prennent au plus une valeur une infinité de fois. Plus précisément, une suite qui ne prend aucune valeur une infinité de fois, c'est-à-dire une suite  $(x_n)_{n \geq 0}$  telle que pour tout  $a \in \mathbb{R}$ , l'ensemble  $\{n \geq 0 : x_n = a\}$  soit fini, est convergente, et admet chaque réel comme limite (la topologie n'est pas séparée!). Une suite pour laquelle il existe un unique réel  $a$  tel que l'ensemble  $\{n \geq 0 : x_n = a\}$  soit infini converge vers ce réel  $a$  et vers aucun autre. Les autres suites n'admettent pas de limite.

On notera que la topologie de Zariski n'est pas à bases de voisinages dénombrables, et que pourtant, pour cette topologie, l'adhérence séquentielle et l'adhérence coïncident.

La topologie de Zariski, encore elle, fournit un contre-exemple à la proposition 1.2.6 lorsque  $E$  est un espace topologique. En effet, sur  $\mathbb{R}$  muni de la topologie de Zariski, la tribu borélienne est la tribu des parties dénombrables ou co-dénombrables. Par ailleurs, les seules fonctions continues sont les fonctions constantes. Ainsi, par exemple,  $\delta_0$  et  $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$  sont deux mesures de probabilité boréliennes différentes, puisqu'elles ne donnent pas la même masse à la partie dénombrable  $\{0\}$ , mais qui intègrent de la même façon les fonctions continues.

11. Une idée qui peut être un guide intuitif dans ces questions est que  $\mathcal{M}(E)$  est une sorte d'enveloppe convexe de  $E$ , ou pour mieux le dire un convexe dont les points extrémaux sont ceux de  $E$ . En effet, sur un espace raisonnable,

Pour étudier ces questions, il semble que nous ne puissions pas faire l'économie d'une étude abstraite un peu plus poussée de certains aspects des espaces métriques. Commençons par tirer au clair les notions de précompacité et, tant que nous y sommes, de relative compacité, ce qui nous sera utile par la suite.

### 1.4.2 Précompacité, compacité, relative compacité

**Définition 1.4.3.** *On dit qu'un espace métrique  $(E, d)$  est précompact si pour tout  $\varepsilon > 0$  il est possible de recouvrir  $E$  par un nombre fini de boules de rayon  $\varepsilon$ .*

Tout espace métrique compact est précompact, puisque du recouvrement par la famille de toutes les boules ouvertes de rayon  $\varepsilon$ , on peut extraire un sous-recouvrement fini. Par contre, l'intervalle  $]0, 1[$ , par exemple, est précompact sans être compact.

Un espace métrique précompact est nécessairement séparable (c'est un bon exercice de le démontrer) et borné (au sens où son diamètre est fini). Mais par exemple,  $\mathbb{R}$  est séparable sans être précompact, et un ensemble infini dénombrable muni de la distance  $d(x, y) = \mathbb{1}_{x \neq y}$  est séparable et borné sans être précompact.

La démonstration des trois énoncés suivants est laissée en exercice.

**Lemme 1.4.4.** *Un espace métrique est précompact si et seulement si pour tout  $\varepsilon > 0$  il est possible de le recouvrir par un nombre fini de parties de diamètre inférieur à  $\varepsilon$ .*

**Proposition 1.4.5.** *Un espace métrique est précompact si et seulement si de toute suite on peut extraire une sous-suite qui soit de Cauchy.*

**Proposition 1.4.6.** *Un espace métrique est compact si et seulement s'il est précompact et complet.*

On pourra aussi vérifier et noter la jolie symétrie des trois assertions suivantes.

- Sur un espace métrique compact, toute fonction continue est bornée.
- Sur un espace métrique précompact, toute fonction uniformément continue est bornée.
- Sur un espace métrique borné, toute fonction lipschitzienne est bornée.

Une propriété de la précompacité dont nous aurons besoin est la suivante.

**Lemme 1.4.7.** *Soit  $(E, d)$  un espace métrique. Soit  $A$  une partie de  $E$ . Alors  $A$  est précompacte si et seulement si  $\overline{A}$  est précompacte.*

*Démonstration.* Supposons  $\overline{A}$  précompacte et donnons-nous  $\varepsilon > 0$ . Un recouvrement de  $\overline{A}$  par des parties de diamètre inférieur à  $\varepsilon$  recouvre en particulier  $A$ , qui est donc précompact.

Supposons maintenant  $A$  précompact et donnons-nous  $\varepsilon > 0$ . Soient  $x_1, \dots, x_n$  les centres de boules de rayon  $\varepsilon/2$  qui recouvrent  $A$ . Alors, puisque l'adhérence d'une union finie est

toute mesure de probabilités est en un certain sens un barycentre de masses de Dirac, par exemple par la formule

$$\mu = \int_E \delta_x \, d\mu(x)$$

à laquelle nous ne cherchons pas ici à donner un sens très précis. Or nous pouvons penser à l'application  $x \mapsto \delta_x$  comme à un plongement (toujours en un sens vague) de  $E$  dans  $\mathcal{C}_b(E)'$ . Et dans cet espace,  $\mathcal{M}(E)$  est l'enveloppe convexe de l'ensemble des masses de Dirac, donc en un certain sens, une enveloppe convexe de  $E$  lui-même.

l'union des adhérences, et puisque l'adhérence d'une boule est incluse dans toute boule de même centre et de rayon strictement supérieur, on a

$$\overline{A} = \bigcup_{i=1}^n \overline{B}(x_i, \varepsilon/2) \subseteq \bigcup_{i=1}^n B(x_i, \varepsilon),$$

d'où il découle que  $\overline{A}$  est précompact.  $\square$

Abordons maintenant la notion de relative compacité, dont nous aurons besoin plus tard.

**Définition 1.4.8.** On dit qu'une partie  $A$  d'un espace métrique  $(E, d)$  est relativement compacte si son adhérence est compacte.

De ce qui précède, on déduit la caractérisation suivante des parties relativement compactes d'un espace complet.

**Proposition 1.4.9.** Une partie d'un espace métrique complet est relativement compacte si et seulement si elle est précompacte.

*Démonstration.* Soit  $A$  une partie d'un espace métrique complet  $(E, d)$ . Supposons  $A$  relativement compacte. Alors  $\overline{A}$  est compacte, donc précompacte, donc d'après le lemme 1.4.7,  $A$  est précompacte.

Supposons maintenant  $A$  précompacte. Alors, toujours d'après le lemme 1.4.7,  $\overline{A}$  est précompacte. Par ailleurs,  $\overline{A}$  est fermée dans  $E$  qui est complet, donc elle est complète. Ainsi,  $\overline{A}$  est précompacte et complète, donc compacte d'après la proposition 1.4.6.  $\square$

### 1.4.3 Séparabilité et précompacité

Nous avons déjà observé qu'alors qu'un espace métrique précompact était nécessairement séparable, un espace métrique séparable pouvait ne pas être précompact. Nous allons voir que quitte à changer sa distance, mais sans changer sa topologie, on peut rendre un espace séparable précompact. Par exemple, sur  $\mathbb{R}$ , la topologie usuelle est induite par la distance  $\delta(x, y) = |\arctan x - \arctan y|$  qui en fait un espace isométrique à  $] -\frac{\pi}{2}, \frac{\pi}{2}[$ , donc précompact.

**Proposition 1.4.10.** Soit  $(E, d)$  un espace métrique séparable. Il existe sur  $E$  une distance  $\delta$  qui induit la même topologie que  $d$  et telle que  $(E, \delta)$  est précompact.

*Démonstration.* Nous allons montrer que  $E$  est homéomorphe à une partie d'un espace métrique compact. En l'occurrence, nous allons choisir l'espace  $[0, 1]^{\mathbb{N}}$  muni de la topologie produit, qui d'après le théorème de Tychonoff, ou plus concrètement d'après le procédé d'extraction diagonale, est compact.<sup>12</sup>

12. Rappelons de quoi il s'agit. Soit  $((x_n^m)_{n \geq 0})_{m \geq 0}$  une suite d'éléments de  $[0, 1]^{\mathbb{N}}$ , c'est-à-dire une suite (indexée par  $m$ ) de suites (indexées par  $n$ ) de réels entre 0 et 1. De la suite  $(x_0^m)_{m \geq 0}$  des premiers termes des suites considérées, on peut, par compacité de  $[0, 1]$ , extraire une sous-suite convergente  $(x_0^{\psi_0(m)})_{m \geq 0}$ , où  $\psi_0 : \mathbb{N} \rightarrow \mathbb{N}$  est une injection croissante. Notons  $x_0$  la limite de cette sous-suite. De la suite  $(x_1^{\psi_0(m)})_{m \geq 0}$ , on peut maintenant extraire une sous-suite  $(x_1^{\psi_0(\psi_1(m))})_{m \geq 0}$ , qui converge vers  $x_1$ . On construit, par récurrence, une suite  $(\psi_n)_{n \geq 0}$  d'injections croissantes



Pour commencer, considérons dans  $E$  une partie dénombrable dense  $Z = \{z_p : p \geq 0\}$ . Définissons une application  $h : E \rightarrow [0, 1]^{\mathbb{N}}$  en posant, pour tout  $x \in E$ ,

$$h(x) = \left( \frac{d(x, z_p)}{1 + d(x, z_p)} \right)_{p \geq 0}.$$

Nous munissons  $[0, 1]^{\mathbb{N}}$  de la distance

$$D((a_p)_{p \geq 0}, (b_p)_{p \geq 0}) = \sum_{p=0}^{\infty} 2^{-p} |a_p - b_p|.$$

Pour cette distance, une suite de suites converge si et seulement si elle converge terme à terme. Autrement dit, en voyant  $[0, 1]^{\mathbb{N}}$  comme l'ensemble des fonctions de  $\mathbb{N}$  dans  $[0, 1]$ , une suite d'éléments de  $[0, 1]^{\mathbb{N}}$  converge pour la distance  $D$  si et seulement si elle converge simplement.<sup>13</sup> Comme nous l'avons dit plus haut, le théorème de Tychonoff, ou le procédé d'extraction diagonale, entraînent que  $([0, 1]^{\mathbb{N}}, D)$  est un espace métrique compact.

L'application  $h$  est continue : en effet, si  $(x_n)_{n \geq 0}$  est une suite de  $E$  qui converge vers  $x$ , alors pour tout  $p \geq 0$ , la suite  $\left( \frac{d(x_n, z_p)}{1 + d(x_n, z_p)} \right)_{n \geq 0}$  converge vers  $\frac{d(x, z_p)}{1 + d(x, z_p)}$ .<sup>14</sup>

L'application  $h$  est injective : en effet, si  $h(x) = h(y)$ , alors  $x$  et  $y$  sont à la même distance de chaque point de  $Z$ . Or il existe une suite d'éléments de  $Z$  qui converge vers  $x$ , et cette suite converge donc également vers  $y$ , qui est donc égal à  $x$ .

L'application  $h$  admet donc une application réciproque sur son image  $h(E)$ . Montrons que cette réciproque est continue. Supposons pour cela qu'une suite  $(x_n)_{n \geq 0}$  soit telle que la suite de ses images  $(h(x_n))_{n \geq 0}$  soit convergente, vers  $h(y)$ . Alors pour tout  $p \geq 0$ , la suite  $(d(x_n, z_p))_{n \geq 0}$  converge vers  $d(y, z_p)$ . Montrons que la suite  $(x_n)_{n \geq 0}$  converge vers  $y$ . Donnons-nous donc  $\varepsilon > 0$ . Soit  $p \in \mathbb{N}$  tel que  $d(y, z_p) < \varepsilon/2$ . Pour  $n$  assez grand, on a aussi  $d(x_n, z_p) < \varepsilon/2$ , si bien que  $d(x_n, y) < \varepsilon$ .

Finalement,  $h : E \rightarrow h(E) \subseteq [0, 1]^{\mathbb{N}}$  est un homéomorphisme de  $(E, d)$  sur  $(h(E), D)$ .

Posons maintenant, pour tous  $x, y \in E$ ,

$$\delta(x, y) = D(h(x), h(y)).$$

Alors  $h$  est une isométrie de  $(E, \delta)$  sur  $(h(E), D)$ .

Nous concluons en disant d'une part que  $h(E)$ , qui est une partie d'un espace métrique compact, est précompacte, si bien que  $(E, \delta)$  est précompact; et d'autre part que  $(E, d)$  est homéomorphe à  $(h(E), D)$  donc à  $(E, \delta)$ , ce qui signifie que  $d$  et  $\delta$  induisent la même topologie sur  $E$ .  $\square$

---

de  $\mathbb{N}$  dans  $\mathbb{N}$  telles que pour tout  $n \geq 0$ , la suite  $(x_n^{\psi_0 \circ \dots \circ \psi_n(m)})_{m \geq 0}$  converge, vers  $x_n$ . Alors la suite de suites  $((x_n^{\psi_0 \circ \dots \circ \psi_m(m)})_{n \geq 0})_{m \geq 0}$  converge dans  $[0, 1]^{\mathbb{N}}$ , c'est-à-dire qu'elle converge terme à terme, vers la suite  $(x_n)_{n \geq 0}$ . En effet, pour tout  $n \geq 0$ , la suite  $(x_n^{\psi_0 \circ \dots \circ \psi_m(m)})_{m \geq 0}$  est, à partir du rang  $m = n$ , extraite de la suite  $(x_n^{\psi_0 \circ \dots \circ \psi_n(m)})_{m \geq 0}$  qui converge vers  $x_n$ . Le même argument mot à mot permet en fait de montrer que le produit cartésien d'une famille dénombrable d'espaces métriques compacts muni de la topologie produit est compact.

13. On peut aussi dire que la topologie induite par  $D$  est la topologie produit.

14. On peut aussi montrer que  $h$  est 2-lipschitzienne.



#### 1.4.4 Métrisabilité de $\mathcal{M}(E)$ lorsque $E$ est séparable

Pour étudier la topologie de  $\mathcal{M}(E)$ , même sans vouloir entrer dans des raffinements extrêmes, nous avons besoin de savoir que cette topologie est *métrisable*. En effet, le savoir nous permettra de démontrer, sous de bonnes hypothèses, sa compacité, en utilisant la caractérisation séquentielle de la compacité, ce qui est plus facile que d'utiliser la caractérisation de Borel–Lebesgue.

Nous allons montrer que si  $E$  est séparable, alors  $\mathcal{M}(E)$  est métrisable. Pour cela, nous allons utiliser un argument de séparabilité d'un bon espace de fonctions sur  $E$ . La proposition suivante énonce ce dont nous aurons besoin.

Rappelons que nous notons  $\mathcal{U}(E)$  l'espace des fonctions uniformément continues sur un espace métrique, et  $\mathcal{U}_b(E)$  l'espace des fonctions uniformément continues bornées. Nous avons déjà observé que si  $E$  est précompact, alors  $\mathcal{U}_b(E)$  coïncide avec  $\mathcal{U}(E)$ . Dans ce cas, l'espace  $(\mathcal{U}(E), \|\cdot\|)$  est un espace de Banach (voir l'exercice 1.8).

**Proposition 1.4.11.** *Soit  $(E, d)$  un espace métrique précompact. Alors l'espace  $\mathcal{U}(E) = \mathcal{U}_b(E)$  muni de la distance uniforme est séparable.*

*Démonstration.* Soit  $(\hat{E}, \hat{d})$  le complété de l'espace  $(E, d)$  (voir l'exercice 1.2). Puisque  $E$ , qui est précompact, est dense dans  $\hat{E}$ , ce dernier est aussi précompact. Puisque  $\hat{E}$  est précompact et complet, il est compact.

Toute fonction continue sur  $\hat{E}$  y est uniformément continue et détermine donc, par restriction, une fonction uniformément continue sur  $E$ . Réciproquement, toute fonction uniformément continue sur  $E$ , qui est une partie dense de  $\hat{E}$ , se prolonge en une fonction continue sur  $\hat{E}$ . Ainsi, la restriction de  $\hat{E}$  à  $E$  détermine une bijection  $\mathcal{C}(\hat{E}) \rightarrow \mathcal{U}(E)$ . Cette bijection préserve de plus la norme (du supremum de la valeur absolue), elle est donc une isométrie d'espaces de Banach.

Soit maintenant  $\{z_p : p \geq 0\}$  une partie dénombrable dense de  $E$ . Alors l'ensemble des polynômes à coefficients rationnels en les fonctions  $\hat{d}(\cdot, z_p)$  sur  $\hat{E}$  est une algèbre de fonctions continues qui sépare les points. Le théorème de Stone–Weierstraß nous assure que cette algèbre est dense dans  $\mathcal{C}(\hat{E})$ . Les restrictions à  $E$  des éléments de cette algèbre forment donc une partie dense de  $\mathcal{U}(E)$ .  $\square$

Notons que si notre espace  $(E, d)$  est compact, la proposition 1.4.10 est sans objet, et le fait que l'espace  $\mathcal{U}(E) = \mathcal{C}(E) = \mathcal{C}_b(E)$  soit séparable peut se démontrer directement dans  $(E, d)$ , sans utiliser de complété (qui serait de toute façon isométrique à  $(E, d)$ ), en utilisant les quatre dernières lignes de la preuve de la proposition 1.4.11.

**Exercice 1.12.** Soit  $(E, d)$  un espace métrique. Soit  $A \subseteq E$  une partie de  $E$ . Soit  $f : A \rightarrow \mathbb{R}$  une fonction uniformément continue (vis-à-vis de la restriction à  $A$  de la distance  $d$ ). Montrer qu'il existe une unique application continue  $g : \overline{A} \rightarrow \mathbb{R}$  dont la restriction à  $A$  coïncide avec  $f$ . Montrer que  $g$  est uniformément continue et que  $\sup\{|g(x)| : x \in \overline{A}\} = \sup\{|f(x)| : x \in A\}$ .

**Théorème 1.4.12.** *Soit  $(E, d)$  un espace métrique séparable. Alors la topologie de  $\mathcal{M}(E)$  est métrisable.*

*Démonstration.* Soit  $\delta$  une distance sur  $E$ , donnée par la proposition 1.4.10, qui induit la même topologie que  $d$  et pour laquelle  $(E, \delta)$  est précompact. D'après la proposition 1.4.11, l'espace

$\mathcal{U}^\delta(E)$  des fonctions sur  $E$  uniformément continues par rapport à la distance  $\delta$  est séparable pour la norme uniforme. Soit donc  $\{g_p : p \geq 0\}$  une partie dénombrable dense de  $\mathcal{U}^\delta(E)$ . Définissons sur  $\mathcal{M}(E) \times \mathcal{M}(E)$  la fonction

$$d(\mu, \nu) = \sum_{p=0}^{\infty} 2^{-p} \min \left( 1, \left| \int_E g_p \, d\mu - \int_E g_p \, d\nu \right| \right).$$

On vérifie que c'est une distance sur  $\mathcal{M}(E)$ , et une suite  $(\mu_n)_{n \geq 0}$  converge vers  $\mu$  relativement à cette distance si et seulement si pour tout  $p \geq 0$ , la suite  $(\int_E g_p \, d\mu_n)_{n \geq 0}$  converge vers  $\int_E g_p \, d\mu$ .

En particulier, si  $\mu_n \xrightarrow[n \rightarrow \infty]{} \mu$ , alors  $d(\mu_n, \mu) \xrightarrow[n \rightarrow \infty]{} 0$ . Réciproquement, supposons que  $d(\mu_n, \mu)$  tende vers 0. Soit  $f$  une fonction uniformément continue sur  $(E, \delta)$ . Donnons-nous  $\varepsilon > 0$ . Choisissons un entier  $p$  tel que  $\|f - g_p\| < \varepsilon/2$ . Alors

$$\overline{\lim}_{n \rightarrow \infty} \left| \int_E f \, d\mu_n - \int_E f \, d\mu \right| \leq \varepsilon + \overline{\lim}_{n \rightarrow \infty} \left| \int_E g_p \, d\mu_n - \int_E g_p \, d\mu \right| = \varepsilon,$$

si bien que  $\int_E f \, d\mu_n \xrightarrow[n \rightarrow \infty]{} \int_E f \, d\mu$ . Ceci montre que  $\mu_n \xrightarrow[n \rightarrow \infty]{} \mu$ .

Toutefois, pour montrer que la topologie de  $\mathcal{M}(E)$  est induite par la distance  $d$ , il ne suffit pas de montrer que les suites de mesures faiblement convergentes sont exactement les suites convergentes pour la distance  $d$  : ce serait supposer que la topologie  $\mathcal{M}(E)$  est caractérisée par ses suites convergentes, ce que nous ne savons pas encore. Procédons donc prudemment.

Pour toute mesure  $\mu \in \mathcal{M}(E)$ , la fonction  $d(\mu, \cdot)$  est la somme d'une série normalement convergente de fonctions continues par rapport à la topologie faible-\* : c'est donc une fonction faible-\* continue. Ainsi, pour tout  $r > 0$ , la boule  $\{v \in \mathcal{M}(E) : d(\mu, v) < r\}$  est un ouvert faible-\*. Tout ouvert de la topologie induite par  $d$  est une réunion de boules ouvertes, donc un ouvert faible-\*.

Réciproquement, soit  $V$  un ouvert faible-\* et soit  $\mu$  un point de  $V$ . Supposons que  $V$  ne soit pas un voisinage de  $\mu$  pour la topologie induite par  $d$ . Ceci signifie qu'aucune  $d$ -boule ouverte centrée en  $\mu$  n'est incluse dans  $V$ , si bien qu'il existe une suite  $(\mu_n)_{n \geq 0}$  de mesures qui n'appartiennent pas à  $V$  et telle que  $d(\mu_n, \mu) \xrightarrow[n \rightarrow \infty]{} 0$ . Or une telle suite vérifie  $\mu_n \xrightarrow[n \rightarrow \infty]{} \mu$ , et appartient donc à  $V$  à partir d'un certain rang : c'est une contradiction.  $\square$

**Exercice 1.13.** Sur un ensemble  $X$ , soit  $\mathcal{T}$  une topologie,  $d$  une distance, et soit  $\mathcal{D}$  la topologie induite par  $d$ . Montrer que si pour tout  $x \in X$ , la fonction  $d(x, \cdot)$  est continue par rapport à  $\mathcal{T}$ , alors  $\mathcal{D} \subseteq \mathcal{T}$ . Montrer que si  $\mathcal{T}$  et  $\mathcal{D}$  ont les mêmes suites convergentes, alors  $\mathcal{T} \subseteq \mathcal{D}$ .

### 1.4.5 Compacité de $\mathcal{M}(E)$ lorsque $E$ est compact

Nous pouvons maintenant démontrer le premier résultat d'intérêt autonome pour des applications de ce que nous sommes en train d'étudier.

**Théorème 1.4.13.** Soit  $(E, d)$  un espace métrique compact. L'espace  $\mathcal{M}(E)$  est compact.

Dans la démonstration de ce théorème, nous allons utiliser le théorème de Riesz, dont l'énoncé et une démonstration sont donnés à l'appendice D.

*Démonstration.* Puisque  $E$  est métrique compact, il est séparable, donc la topologie de  $\mathcal{M}(E)$  est métrisable. Pour démontrer que  $\mathcal{M}(E)$  est compact, il suffit donc de montrer que de toute suite de mesures de probabilité sur  $E$  on peut extraire une sous-suite faiblement convergente.

Donnons-nous une suite  $(\mu_n)_{n \geq 0}$  d'éléments de  $\mathcal{M}(E)$ . La proposition 1.4.11 nous assure que nous pouvons trouver une suite  $(g_p)_{p \geq 0}$  dense dans  $\mathcal{C}(E) = \mathcal{U}(E)$ . Pour tous  $n, p \geq 0$ , l'intégrale  $\int_E g_p d\mu_n$  appartient au segment  $[-\|g_p\|, \|g_p\|]$ . Le théorème de Tychonoff ou un argument d'extraction diagonale (revoir au besoin la note 12) permettent d'affirmer qu'il existe une sous-suite  $(\mu_{\psi(n)})_{n \geq 0}$  de la suite de mesures que nous nous sommes donnée qui est telle que pour tout  $p \geq 0$ , la suite  $(\int_E g_p d\mu_{\psi(n)})_{n \geq 0}$  converge, lorsque  $n$  tend vers l'infini, vers un nombre que nous notons  $\Lambda(g_p)$ . Quitte à remplacer la suite  $(\mu_n)_{n \geq 0}$  par cette sous-suite, nous supposons donc désormais que pour tout  $p \geq 0$ ,

$$\int_E g_p d\mu_n \xrightarrow{n \rightarrow \infty} \Lambda(g_p) \in [-\|g_p\|, \|g_p\|].$$

Nous supposons aussi, quitte à ajouter un terme à notre suite, que  $g_0 = 1$ .

Notons  $\mathcal{G}$  le sous-espace vectoriel de  $\mathcal{C}(E)$  engendré par la famille  $\{g_p : p \geq 0\}$  et considérons  $f \in \mathcal{G}$ . Un tel  $f$  s'écrit sous la forme  $f = \alpha_1 g_{p_1} + \dots + \alpha_m g_{p_m}$  et on déduit de la linéarité de l'intégrale que

$$\sum_{i=1}^m \alpha_i \Lambda(g_{p_i}) = \lim_{n \rightarrow \infty} \int_E f d\mu_n,$$

égalité dont le membre de droite montre que le membre de gauche ne dépend pas de la manière dont on a écrit  $f$  comme combinaison linéaire des  $g_p$ . On note  $\Lambda(f)$  la valeur commune des deux membres de cette égalité.

À partir de  $\Lambda$ , nous allons construire une forme linéaire sur  $\mathcal{C}(E)$  qui vérifie les hypothèses du théorème de Riesz.

La fonction  $\Lambda : \mathcal{G} \rightarrow \mathbb{R}$ , limite simple lorsque  $n$  tend vers l'infini des fonctions  $f \mapsto \int_E f d\mu_n$ , est linéaire. Elle vérifie par ailleurs  $|\Lambda(f)| \leq \|f\|$  pour tout  $f \in \mathcal{G}$ . Elle est donc uniformément continue sur  $\mathcal{G}$ , et s'étend de manière unique à l'adhérence de  $\mathcal{G}$  dans  $\mathcal{C}(E)$ , qui n'est autre que  $\mathcal{C}(E)$  tout entier. Ce prolongement est une forme linéaire sur  $\mathcal{C}(E)$  qui vérifie  $\Lambda(1) = 1$ , comme on le voit en appliquant sa définition à la fonction  $g_0$ . Soit maintenant  $f \in \mathcal{C}(E)$  une fonction positive. Soit  $(g'_k)_{k \geq 0}$  une suite d'éléments de  $\mathcal{G}$  qui converge vers  $f$ . Soit  $\varepsilon > 0$ . À partir d'un certain rang, la fonction  $g'_k + \varepsilon$  est positive, donc  $\Lambda(g'_k) \geq -\varepsilon$ . Ceci entraîne que  $\Lambda(f) \geq -\varepsilon$ . Puisque ceci a lieu pour tout  $\varepsilon > 0$ , on a  $\Lambda(f) \geq 0$ .

Finalement,  $\Lambda$  est une forme linéaire positive sur  $\mathcal{C}(E)$  et telle que  $\Lambda(1) = 1$ . Le théorème de Riesz (théorème D.1.1) nous assure de l'existence d'une (unique) mesure de probabilité  $\mu$  sur  $E$  telle que pour tout  $f \in \mathcal{C}(E)$ , on ait

$$\Lambda(f) = \int_E f d\mu.$$

Il reste à montrer que  $\mu$  est bien la limite faible de la suite  $(\mu_n)_{n \geq 0}$ . Pour cela, considérons

$f \in \mathcal{C}(E)$  et  $\varepsilon > 0$ . Soit  $k \geq 0$  tel que  $\|g_k - f\| < \varepsilon/2$ . On a

$$\begin{aligned} \left| \int_E f \, d\mu_n - \int_E f \, d\mu \right| &\leq \left| \int_E f \, d\mu_n - \int_E g_k \, d\mu_n \right| + \left| \int_E g_k \, d\mu_n - \int_E g_k \, d\mu \right| + \left| \int_E g_k \, d\mu - \int_E f \, d\mu \right| \\ &\leq \varepsilon + \left| \int_E g_k \, d\mu_n - \int_E g_k \, d\mu \right|. \end{aligned}$$

On en déduit que  $\overline{\lim}_{n \rightarrow \infty} \left| \int_E f \, d\mu_n - \int_E f \, d\mu \right| \leq \varepsilon$ , ce qui achève la démonstration.  $\square$

## 1.5 Le théorème de Prokhorov

Le théorème de Prokhorov caractérise les parties de  $\mathcal{M}(E)$  qui sont relativement compactes lorsque  $E$  est un espace Polonais.

**Définition 1.5.1.** Soit  $(E, d)$  un espace métrique. Soit  $\Gamma$  une partie de  $\mathcal{M}(E)$ . On dit que  $\Gamma$  est tendue si pour tout  $\varepsilon > 0$  il existe une partie  $K$  de  $E$  compacte et telle que

$$\forall \mu \in \Gamma, \mu(K) \geq 1 - \varepsilon.$$

Avant de démontrer le théorème de Prokhorov, commençons par une assertion beaucoup plus faible, mais utile, et instructive.

**Proposition 1.5.2.** Soit  $(E, d)$  un espace polonais. Soit  $\mu \in \mathcal{M}(E)$ . Alors  $\{\mu\}$  est tendue. Autrement dit, pour tout  $\varepsilon > 0$ , il existe un compact  $K$  de  $E$  tel que  $\mu(K) \geq 1 - \varepsilon$ .

*Démonstration.* Donnons-nous  $\varepsilon > 0$  et cherchons à construire le compact  $K$ . Pour cela, considérons une famille  $\{z_n : n \geq 0\}$  dénombrable et dense dans  $E$ . Puisque  $E = \bigcup_{n \geq 0} B(z_n, 1)$ , il existe un entier  $N_1$  tel que

$$\mu \left( \bigcup_{n=0}^{N_1} B(z_n, 1) \right) \geq 1 - \varepsilon/2.$$

De même, puisque  $E = \bigcup_{n \geq 0} B(z_n, 1/2)$ , il existe un entier  $N_2$  tel que

$$\mu \left( \bigcup_{n=0}^{N_2} B(z_n, \frac{1}{2}) \right) \geq 1 - \varepsilon/4.$$

Plus généralement, pour tout  $k \geq 1$ , il existe un entier  $N_k$  tel que

$$\mu \left( \bigcup_{n=0}^{N_k} B(z_n, \frac{1}{k}) \right) \geq 1 - 2^{-k} \varepsilon.$$

Posons, pour tout  $k \geq 1$ ,  $B_k = \bigcup_{n=0}^{N_k} B(z_n, \frac{1}{k})$ . Posons également  $A = \bigcap_{k=1}^{\infty} B_k$ .

Par construction, la partie  $A$  est précompacte : en effet, pour tout  $k \geq 1$ , elle peut être recouverte par  $N_k$  boules de rayon  $\frac{1}{k}$ . D'après le lemme 1.4.7, il s'ensuit que  $\overline{A}$  est précompacte.

Comme  $\bar{A}$  est une partie fermée d'un espace complet, elle est elle-même un espace complet, et précompact, donc compact.

Par ailleurs,

$$\mu(\bar{A}^c) \leq \mu(A^c) \leq \sum_{k=1}^{\infty} \mu(B_k^c) \leq \varepsilon,$$

c'est-à-dire que  $\mu(\bar{A}) \geq 1 - \varepsilon$ . □

La proposition que nous venons de démontrer est un cas particulier du théorème qui suit.

**Théorème 1.5.3** (Théorème de Prokhorov). *Soit  $(E, d)$  un espace polonais. Une partie de  $\mathcal{M}(E)$  est relativement compacte si et seulement si elle est tendue.*

L'implication la plus utile de ce théorème, pour nous en tout cas, est l'assertion qu'une partie tendue est relativement compacte. La stratégie de preuve est la suivante :

- se donner une partie  $\Gamma$  tendue de  $\mathcal{M}(E)$  et une suite d'éléments  $(\mu_n)_{n \geq 0}$  de cette partie ;
- remplacer la distance sur  $E$  par une distance qui le rend précompact, puis le compléter, et observer que ce complété  $\hat{E}$  est compact ;
- associer à la suite  $(\mu_n)_{n \geq 0}$  une suite  $(\hat{\mu}_n)_{n \geq 0}$  de mesures sur  $\hat{E}$  et en extraire, par compacité de  $\mathcal{M}(\hat{E})$ , une sous-suite  $(\hat{\mu}_{\psi(n)})_{n \geq 0}$  convergente, vers une mesure  $\nu$  ;
- utiliser la tension de  $\Gamma$  pour montrer que  $\nu$  est, en un certain sens, portée par  $E$  ;
- montrer que la suite  $(\mu_{\psi(n)})_{n \geq 0}$  converge vers  $\nu$ .

Le détail de la démonstration est rendu un peu plus délicat qu'on n'aurait pu s'y attendre par le fait que  $E$ , qui est une partie dense de son complété  $\hat{E}$ , n'en est pas forcément une partie borélienne.<sup>15</sup> Nous allons énoncer un lemme qui tirera ce point au clair. Dans tout ce qui suit, nous identifions un espace métrique à une partie (dense) de son complété.

**Lemme 1.5.4.** *Soit  $E$  un espace métrique et  $\hat{E}$  son complété.*

1. *Les tribus boréliennes de  $E$  et  $\hat{E}$  sont liées par la relation*

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

2. *Soit  $\mu$  un élément de  $\mathcal{M}(E)$ . On définit un élément  $\hat{\mu}$  de  $\mathcal{M}(\hat{E})$  en posant, pour tout  $A \in \mathcal{B}_{\hat{E}}$ ,*

$$\hat{\mu}(A) = \mu(A \cap E).$$

3. *Soit  $f$  une fonction borélienne bornée sur  $\hat{E}$ . Soit  $\check{f}$  sa restriction à  $E$ . Alors  $\check{f}$  est borélienne bornée sur  $E$  et on a l'égalité*

$$\int_E \check{f} \, d\mu = \int_{\hat{E}} f \, d\hat{\mu}.$$

15. Considérons par exemple dans  $[0, 1]$  muni de la distance usuelle une partie  $E$  dense et non borélienne. On peut par exemple prendre pour  $E$  le complémentaire d'une partie négligeable (c'est-à-dire incluse dans un borélien de mesure nulle) non borélienne. Alors le complété de  $E$  est l'intervalle  $[0, 1]$ , et il n'en est pas un borélien.

*Démonstration.* 1. Rappelons-nous d'abord que les ouverts de  $E$  sont exactement les intersections avec  $E$  des ouverts de  $\hat{E}$ .<sup>16</sup> Ainsi, la classe  $\mathcal{C}_E = \{A \cap E : A \in \mathcal{B}_{\hat{E}}\}$  de parties de  $E$ , qui contient toutes les intersections avec  $E$  des ouverts de  $\hat{E}$ , contient tous les ouverts de  $E$ . Par ailleurs,  $\mathcal{C}_E$  est une tribu sur  $E$ , la chose peut-être la moins évidente à vérifier étant sa stabilité par passage au complémentaire, qui résulte de l'égalité

$$E \setminus (A \cap E) = (\hat{E} \setminus A) \cap E.$$

La classe  $\mathcal{C}_E$  contient donc la tribu borélienne de  $E$  : nous avons montré l'inclusion

$$\mathcal{B}_E \subseteq \{A \cap E : A \in \mathcal{B}_{\hat{E}}\}.$$

Pour montrer l'autre inclusion, considérons la classe de parties de  $\hat{E}$

$$\mathcal{D}_{\hat{E}} = \{A \in \mathcal{B}_{\hat{E}} : A \cap E \in \mathcal{B}_E\}.$$

Elle contient tous les ouverts de  $\hat{E}$ . Par ailleurs, c'est une tribu, sa stabilité par passage au complémentaire se vérifiant grâce à la même égalité qui nous a servi à démontrer celle de  $\mathcal{C}_E$ . Ainsi,  $\mathcal{D}_{\hat{E}} = \mathcal{B}_{\hat{E}}$ , ce qui montre l'inclusion

$$\{A \cap E : A \in \mathcal{B}_{\hat{E}}\} \subseteq \mathcal{B}_E.$$

2. L'assertion 1 entraîne que  $\hat{\mu}$  est une mesure sur  $\mathcal{B}_{\hat{E}}$  et on vérifie immédiatement que c'est une mesure de probabilité.

3. L'égalité est vraie si  $f$  est l'indicatrice d'un borélien de  $\hat{E}$ . Le cas général s'en déduit par linéarité et par un argument d'approximation.  $\square$

Nous pouvons maintenant démontrer le théorème de Prokhorov.

*Démonstration du théorème de Prokhorov.* 1. Tendue  $\Rightarrow$  relativement compacte. Grâce à la proposition 1.4.10, commençons par munir  $E$  d'une distance  $\delta$  qui induit la même topologie que  $d$  et qui rend  $(E, \delta)$  précompact. Notons  $(\hat{E}, \hat{\delta})$  le complété de  $(E, \delta)$ , qui, étant précompact et complet, est compact.<sup>17</sup> Rappelons que  $E$  s'identifie à une partie de  $\hat{E}$ , et nous le considérerons comme tel dans toute la suite de cette démonstration. Rappelons aussi que  $E$  est dense dans  $\hat{E}$ .

Nous avons déjà dit, au lemme 1.5.4, que toute mesure  $\mu$  sur  $(E, \mathcal{B}_E)$  induisait une mesure  $\hat{\mu}$  sur  $(\hat{E}, \mathcal{B}_{\hat{E}})$ , en posant, pour tout  $A \in \mathcal{B}_{\hat{E}}$ ,

$$\hat{\mu}(A) = \mu(A \cap E).$$

Nous avons aussi déjà souligné le fait qu'il n'était pas vrai que tout borélien de  $E$  soit un borélien de  $\hat{E}$ . En revanche, notons que toute partie compacte  $K$  de  $E$  est encore une partie

16. Soit  $(X, \mathcal{T})$  un espace topologique et  $Y$  une partie quelconque de  $X$ . L'ensemble  $\{O \cap Y : O \in \mathcal{T}\}$  est une topologie sur  $Y$ , dite induite par la topologie de  $X$ . Supposons la topologie de  $X$  induite par une distance  $d$ . La restriction de la distance  $d$  à  $Y$  est une distance sur  $Y$ . La chose à vérifier est que la topologie sur  $Y$  induite par la restriction de la distance  $d$  est égale à la topologie induite par la topologie de  $X$ . Ce n'est pas très difficile.

17. En remplaçant  $d$  par  $\delta$ , on a gagné la précompacité, mais on a pu perdre la complétude, comme on le voit avec l'exemple de  $\mathbb{R}$  muni de la distance  $\delta(x, y) = |\arctan x - \arctan y|$ .

compacte de  $\hat{E}$ , car le fait pour  $K$  d'être compacte est une propriété intrinsèque de l'espace métrique  $(K, \delta)$ . En particulier, tout compact de  $E$  est un borélien de  $\hat{E}$ .

Soit  $\Gamma \subseteq \mathcal{M}(E)$  une partie tendue. Nous allons montrer que de toute suite de  $\Gamma$  on peut extraire une sous-suite qui converge dans  $\mathcal{M}(E)$ .

Considérons maintenant une suite  $(\mu_n)_{n \geq 0}$  d'éléments de  $\Gamma$ . La suite  $(\hat{\mu}_n)_{n \geq 0}$  est une suite de mesures sur l'espace compact  $\hat{E}$ , donc d'après le théorème 1.4.13, il en existe une sous-suite  $(\hat{\mu}_{\psi(n)})_{n \geq 0}$  qui converge vers une mesure  $\nu \in \mathcal{M}(\hat{E})$ .

Pour tout  $r \geq 1$ , on peut puisque  $\Gamma$  est tendue trouver un compact  $K_r$  inclus dans  $E$  tel que pour tout  $n \geq 1$ , on ait

$$\hat{\mu}_n(K_r) \geq 1 - \frac{1}{r}.$$

Posons  $S = \bigcup_{r=1}^{\infty} K_r$ . La partie  $S$  est un borélien de  $E$  et aussi, d'après ce que nous venons de dire, un borélien de  $\hat{E}$ . On a de plus pour tout  $n \geq 0$  les égalités

$$\hat{\mu}_n(S) = 1 = \mu_n(S \cap E) = \mu_n(S).$$

Par ailleurs, pour tout  $r \geq 1$ , puisque  $K_r$  est fermé, on déduit du théorème de Portmanteau que

$$\nu(K_r) \geq \overline{\lim}_{n \rightarrow \infty} \hat{\mu}_{\psi(n)}(K_r) \geq 1 - \frac{1}{r},$$

si bien que  $\nu(S) = 1$ .

Nous voudrions maintenant définir, à partir de la mesure  $\nu$ , une mesure  $\mu$  sur  $E$ . Pour cela, considérons un borélien de  $E$ . Il s'écrit  $A \cap E$ , avec  $A$  borélien de  $\hat{E}$ , et nous voudrions poser  $\mu(A \cap E) = \nu(A)$ , mais l'écriture  $A \cap E$  de notre borélien n'a aucune raison d'être unique. Supposons donc que  $A \cap E = B \cap E$ , avec  $A, B \in \mathcal{B}_{\hat{E}}$ . Alors, puisque  $\nu(S) = 1$ , nous avons

$$\nu(A) = \nu(A \cap S) = \nu(B \cap S) = \nu(B),$$

l'égalité centrale provenant du fait que  $A \cap S = B \cap S$ . Ainsi, la relation

$$\mu(A \cap E) = \nu(A)$$

définit sans ambiguïté une mesure de probabilité  $\mu$  sur  $(E, \mathcal{B}_E)$ .

Définissons maintenant une mesure  $\mu$  sur  $(E, \mathcal{B}_E)$  par restriction de  $\nu$ , en posant, pour tout  $A \in \mathcal{B}_{\hat{E}}$ ,

$$\mu(A \cap E) = \nu(A) = \nu(A \cap S).$$

On a bien sûr  $\mu(S) = 1$ . De plus, l'égalité  $\mu(A \cap E) = \nu(A)$  montre que  $\nu = \hat{\mu}$ .

Il nous reste à montrer que la suite  $(\mu_{\psi(n)})_{n \geq 0}$  converge vers  $\mu$ . Pour cela, considérons une fonction  $f \in \mathcal{U}_b^{\delta}(E)$  et notons  $\hat{f} \in C_b(\hat{E})$  son prolongement continu. On a, avec les notations du lemme,  $f = \hat{f}$ , si bien que

$$\int_E f \, d\mu_{\psi(n)} = \int_E \hat{f} \, d\hat{\mu}_{\psi(n)} \xrightarrow{n \rightarrow \infty} \int_E \hat{f} \, d\hat{\mu} = \int_E f \, d\mu.$$

Ceci achève la démonstration de cette implication.

2. Relativement compacte  $\Rightarrow$  tendue. Quitte à remplacer  $\Gamma$  par son adhérence, ce qui ne fera que renforcer la conclusion, on peut la supposer compacte. Soit donc  $\Gamma$  une partie compacte de  $\mathcal{M}(E)$ .

Nous allons faire une démonstration qui ressemble beaucoup à celle de la proposition 1.5.2. Commençons par démontrer un résultat intermédiaire. Soit  $(O_n)_{n \geq 0}$  une suite croissante d'ouverts de  $E$  dont la réunion est  $E$  tout entier. Montrons qu'il existe alors, pour tout  $\varepsilon > 0$ , un entier  $N$  tel que pour tout  $\mu \in \Gamma$ , on ait  $\mu(O_N) \geq 1 - \varepsilon$ .<sup>18</sup>

Dans le cas contraire, il existerait  $\varepsilon > 0$  et une suite  $(\mu_n)_{n \geq 0}$  d'éléments de  $\Gamma$  tels que pour tout  $n \geq 0$  on ait  $\mu_n(O_n) < 1 - \varepsilon$ . Quitte à extraire une sous-suite de la suite  $(\mu_n)_{n \geq 0}$ , on pourrait supposer qu'elle converge faiblement vers une mesure  $\mu$ . Alors le théorème de Portmanteau nous donnerait, pour tout  $n \geq 0$ ,

$$\mu(O_n) \leq \liminf_{m \rightarrow \infty} \mu_m(O_n) \leq \liminf_{m \rightarrow \infty} \mu_m(O_m) \leq 1 - \varepsilon.$$

Ce serait contradictoire avec le fait que  $\lim_{n \rightarrow \infty} \mu(O_n) = \mu(E) = 1$ .

Nous pouvons maintenant nous donner  $\varepsilon > 0$  et reprendre la méthode de démonstration de la proposition 1.5.2. En effet, étant donné une famille  $\{z_n : n \geq 0\}$  dénombrable et dense dans  $E$ , et d'après le résultat que nous venons de démontrer, il existe pour tout  $k \geq 1$  un entier  $N_k$  tel que pour tout  $\mu \in \Gamma$ , on ait

$$\mu \left( \bigcup_{n=0}^{N_k} B(z_n, \frac{1}{k}) \right) \geq 1 - 2^{-k} \varepsilon.$$

La partie

$$K = \bigcap_{k=1}^{\infty} \bigcup_{n=0}^{N_k} \bar{B}(z_n, \frac{1}{k})$$

est compacte et vérifie, pour tout  $\mu \in \Gamma$ , l'inégalité  $\mu(K) \geq 1 - \varepsilon$ . □

Il vaut la peine d'aller vérifier dans la démonstration de la première implication, selon laquelle une partie tendue est relativement compacte, que nous n'y avons pas utilisé la complétude de  $E$ , mais seulement l'hypothèse que  $(E, d)$  était séparable.

---

18. Autrement dit, le "théorème de convergence monotone pour les mesures" a lieu uniformément sur  $\Gamma$  pour les suites croissantes d'ouverts.



# 2

## Convergence en loi des processus continus

---

Nous allons appliquer l'étude de l'espace des mesures sur un espace polonais que nous avons menée au chapitre précédent au cas de l'espace des fonctions continues sur un intervalle compact de  $\mathbb{R}$  à valeurs dans un espace polonais. Notre but est d'établir des critères de convergence qui soient pratiques à mettre en œuvre, et de les appliquer à la démonstration d'un résultat de convergence en loi de marches aléatoires convenablement renormalisées vers le mouvement brownien.

Les fonctions que nous allons considérer seront définies sur l'intervalle compact  $[0, 1]$ . Le cas d'un autre intervalle compact serait évidemment absolument identique. Le cas d'un intervalle non compact, en revanche, tel que l'intervalle  $\mathbb{R}_+$  d'usage très courant en probabilités, pose quelques problèmes techniques supplémentaires, et nous en dirons un mot à la fin de ce chapitre.

### 2.1 L'espace des fonctions continues à valeurs dans un espace polonais

Soit  $(E, d)$  un espace métrique. Nous notons

$$\mathcal{W}(E) = \mathcal{C}([0, 1], E) = \{x : [0, 1] \rightarrow E \text{ continue}\}$$

l'espace des fonctions continues de  $[0, 1]$  dans  $E$ . Nous noterons parfois, avec l'usage probabiliste en tête,  $x = (x_t)_{t \geq 0}$  un élément de  $\mathcal{W}(E)$ . Nous munirons toujours l'ensemble  $\mathcal{W}(E)$  de la distance

$$D(x, y) = \sup\{d(x_t, y_t) : t \in [0, 1]\}.$$

Notre premier objectif est de démontrer que si  $(E, d)$  est polonais, alors  $(\mathcal{W}(E), D)$  est aussi polonais. Un premier morceau de la démonstration est un exercice classique.

**Exercice 2.1.** Montrer que si  $(E, d)$  est complet, alors  $(\mathcal{W}(E), D)$  est complet.

Le second morceau est le résultat suivant.

**Proposition 2.1.1.** *Si  $(E, d)$  est séparable, alors  $(\mathcal{W}(E), D)$  est séparable.*

Notons que dans le cas où  $E = \mathbb{R}$ , qui est un cas que nous considérerons fréquemment, cette proposition découle immédiatement de la proposition 1.4.11, car  $\mathcal{W}(\mathbb{R}) = \mathcal{C}([0, 1]) = \mathcal{U}([0, 1])$ .

En revanche, dans le cas général, la démonstration de cette proposition n'est pas tout à fait élémentaire. Nous avons tout d'abord besoin de savoir que la topologie de  $\mathcal{W}(E)$  ne dépend que de la topologie de  $E$ .

**Lemme 2.1.2.** *Deux distances sur un ensemble  $E$  qui induisent la même topologie déterminent sur  $\mathcal{W}(E)$  des distances qui induisent la même topologie.*

On peut sans doute démontrer cette proposition en travaillant directement avec les distances, mais une manière peut-être meilleure de la démontrer consiste à décrire la topologie de  $(\mathcal{W}(E), D)$  à partir de celle de  $(E, d)$  sans faire référence aux distances de part ni d'autre. La bonne notion pour cela est celle de topologie *compacte-ouverte* sur l'espace des applications continues entre deux espaces topologiques.

**Définition 2.1.3.** *Soient  $X$  et  $Y$  deux espaces topologiques. Soit  $\mathcal{C}(X, Y)$  l'ensemble des applications continues de  $X$  dans  $Y$ . Pour toute partie  $K \subseteq X$  compacte et toute partie  $U \subseteq Y$  ouverte, on note*

$$V_{K,U} = \{f \in \mathcal{C}(X, Y) : f(K) \subseteq U\}.$$

On appelle topologie *compacte-ouverte* sur  $\mathcal{C}(X, Y)$  la topologie la moins fine pour laquelle toutes les parties de la forme  $V_{K,U}$  sont ouvertes.

Dans la topologie compacte-ouverte, un voisinage d'une fonction  $f$  est une partie de  $\mathcal{C}(X, Y)$  qui contient une partie de la forme  $V_{K_1, U_1} \cap \dots \cap V_{K_r, U_r}$  qui contient  $f$ . Le fait qui nous importe est le suivant.

**Lemme 2.1.4.** *Soit  $(E, d)$  un espace métrique. La topologie sur  $\mathcal{W}(E) = \mathcal{C}([0, 1], E)$  induite par la distance  $D$  est la topologie compacte-ouverte.*

Soulignons que le lemme 2.1.2 découle immédiatement de ce lemme.

*Démonstration.* Soit  $x$  un élément de  $\mathcal{C}([0, 1], E)$ . Nous allons montrer que tout voisinage compact-ouvert de  $x$  contient un voisinage pour la topologie induite par  $D$ , et réciproquement.

Soit  $V$  un voisinage compact-ouvert de  $x$ . Il existe donc un entier  $r \geq 1$ , des compacts  $K_1, \dots, K_r$  de  $[0, 1]$ , et des ouverts  $U_1, \dots, U_r$  de  $E$  tels que  $x \in V_{K_1, U_1} \cap \dots \cap V_{K_r, U_r} \subseteq V$ . Notons

$$\varepsilon = \min\{d(x(K_i), U_i^c) : i = 1, \dots, r\}$$

qui par compacité de  $K_1, \dots, K_r$ , est un nombre strictement positif. Alors la  $D$ -boule ouverte de centre  $x$  et de rayon  $\varepsilon$  est incluse dans  $V$ .

Réciproquement, donnons-nous  $\varepsilon > 0$  et considérons la  $D$ -boule de rayon  $\varepsilon$  autour de  $x$ . Soit  $\delta$  un module de  $\varepsilon/2$ -uniforme continuité de  $x$  : si  $s$  et  $t$  sont dans  $[0, 1]$  tels que  $|s - t| < \delta$ , alors  $d(x_s, x_t) < \varepsilon/2$ . Soit  $r \geq 1$  un entier tel que  $1/r < \delta$ . Pour tout  $i \in \{1, \dots, r\}$ , posons  $K_i = [(i-1)/r, i/r]$  et  $U_i = B(x_{i/r}, \varepsilon/2)$ . Alors  $x \in V_{K_1, U_1} \cap \dots \cap V_{K_r, U_r}$ . De plus, considérons  $y \in V_{K_1, U_1} \cap \dots \cap V_{K_r, U_r}$ . Soit  $t \in [0, 1]$ . Soit  $i \in \{1, \dots, r\}$  tel que  $t \in K_i$ . Alors

$$d(x_t, y_t) \leq d(x_t, x_{i/n}) + d(x_{i/n}, y_t) < \varepsilon$$

donc  $y$  appartient à la  $D$ -boule de rayon  $\varepsilon$  centrée en  $x$ . Ceci conclut la preuve.  $\square$

Nous allons maintenant démontrer la proposition 2.1.1 lorsque  $E$  est un espace vectoriel normé.

**Lemme 2.1.5.** *Soit  $(V, \|\cdot\|)$  un espace vectoriel normé séparable. Alors  $\mathcal{W}(V)$  est séparable.*

*Démonstration.* Soit  $D \subset V$  une partie dénombrable dense. Notons  $\mathcal{A}$  l'ensemble des éléments  $y \in \mathcal{W}(V)$  tels qu'il existe  $n \geq 1$  et  $v_0, \dots, v_n \in D$  tels que pour tout  $i \in \{0, \dots, n-1\}$  et tout  $t \in [i/n, (i+1)/n]$ , on ait

$$y_t = ((i+1) - nt)v_{n+1} + (nt - i)v_n.$$

En français, ce sont les application affines par morceaux, avec des morceaux de largeur  $1/n$ , et dont les valeurs aux multiples de  $1/n$  appartiennent à  $D$ . L'ensemble  $\mathcal{A}$  est une partie dénombrable de  $\mathcal{W}(V)$ . Montrons qu'elle est également dense.

Soit  $x \in \mathcal{W}(V)$ . Soit  $\varepsilon > 0$ . Nous allons montrer qu'il existe  $y \in \mathcal{A}$  tel que  $D(x, y) < \varepsilon$ . Pour cela, considérons un entier  $n \geq 1$  suffisamment grand pour que pour tous  $s, t \in [0, 1]$  tels que  $|s - t| < 1/n$ , on ait  $d(x_s, x_t) < \varepsilon/2$ . Pour tout  $i \in \{0, \dots, n\}$ , choisissons un point  $v_i$  de  $D$  dans la boule  $B(x_{i/n}, \varepsilon/2)$ . Soit  $y$  l'interpolation affine par morceaux de  $v_0, \dots, v_n$ . Alors un calcul analogue à celui que nous avons fait dans la démonstration du lemme 2.1.4 montre que  $D(x, y) < \varepsilon$ .  $\square$

Nous pouvons maintenant rassembler les morceaux.

*Démonstration de la proposition 2.1.1.* Utilisons la proposition 1.4.10 pour munir  $E$  d'une distance  $\delta$  qui induit la même topologie que  $d$  sur  $E$ , et qui rend  $(E, \delta)$  précompact. Notons  $\Delta$  la distance sur  $\mathcal{W}(E)$  induite par  $\delta$ . Nous allons montrer que  $(\mathcal{W}(E), \Delta)$  est séparable. D'après le lemme 2.1.2, il s'ensuivra que  $(\mathcal{W}(E), D)$  est séparable.

Considérons le complété  $(\hat{E}, \hat{\delta})$  de  $(E, \delta)$ . Comme nous l'avons déjà observé en d'autres occasions, ce complété est compact. Nous allons construire une isométrie de  $(\hat{E}, \hat{\delta})$  avec une partie de l'espace de Banach  $\mathcal{C}(\hat{E}, \hat{\delta})$ . Pour cela, nous considérons l'application

$$\begin{aligned} \iota : \hat{E} &\longrightarrow \mathcal{C}(\hat{E}) \\ x &\longmapsto \hat{\delta}(x, \cdot). \end{aligned}$$

L'inégalité  $|\hat{\delta}(x, z) - \hat{\delta}(y, z)| \leq \hat{\delta}(x, y)$  valable pour tous  $x, y, z \in \hat{E}$  montre que  $\|\iota(x) - \iota(y)\| \leq \hat{\delta}(x, y)$ . Par ailleurs,  $|\iota(x)(y) - \iota(y)(y)| = \hat{\delta}(x, y)$ , si bien que  $\iota$  est une isométrie.

L'inclusion de  $E$  dans  $\hat{E}$  composée par l'application  $\iota$  envoie  $(E, \delta)$  isométriquement dans  $(\mathcal{C}(\hat{E}, \hat{\delta}), \|\cdot\|)$ . Ainsi,  $\mathcal{W}(E)$ , muni de la distance induite par  $\delta$  est isométrique à une partie de  $\mathcal{W}(\mathcal{C}(\hat{E}, \hat{\delta}))$  muni de la distance induite par la norme  $\|\cdot\|$ . Le lemme 2.1.5 montre que  $\mathcal{W}(\mathcal{C}(\hat{E}, \hat{\delta}))$  est séparable. Il en est donc de même de  $\mathcal{W}(E)$ .  $\square$

Notons que nous avons utilisé le fait qu'une partie d'un espace séparable était séparable, chose qu'il est plus facile de voir avec la définition topologique de la séparabilité qu'avec sa définition métrique.

Nous avons finalement démontré le fait suivant.

**Théorème 2.1.6.** *Si  $(E, d)$  est polonais, alors  $(\mathcal{W}(E), D)$  est polonais.*

## 2.2 Compacité dans $\mathcal{W}(E)$

Notre objectif est d'étudier la convergence de certaines suites de mesures sur  $\mathcal{W}(E)$ . Pour cela, nous utiliserons le théorème de Prokhorov, et pour utiliser ce théorème, nous avons besoin de comprendre quelles sont les parties relativement compactes de  $\mathcal{W}(E)$ . Le théorème d'Ascoli, ou d'Arzelà–Ascoli, donne une caractérisation de ces parties fondée sur la notion d'équicontinuité.

**Définition 2.2.1.** Soit  $(E, d)$  un espace métrique. Une partie  $A$  de  $\mathcal{C}([0, 1], E)$  est dite équicontinue si pour tout  $\varepsilon > 0$ , il existe  $\delta > 0$  tel que pour tout  $x \in A$  et tous  $s, t \in [0, 1]$  tels que  $|s - t| < \delta$ , on ait  $d(x_s, x_t) < \varepsilon$ .

Il sera commode d'adopter une notation pour le module d'uniforme continuité d'une fonction  $x \in \mathcal{C}([0, 1], \mathcal{W})$  : c'est la fonction  $\omega(x, \cdot) : \mathbb{R}_+^* \rightarrow \mathbb{R}_+$  définie en posant, pour tout  $\delta > 0$ ,

$$\omega(x, \delta) = \sup\{d(x_s, x_t) : s, t \in [0, 1], |s - t| < \delta\}.$$

Dire qu'une fonction  $x : [0, 1] \rightarrow E$  est uniformément continue signifie que  $\omega(x, \delta)$  tend vers 0 lorsque  $\delta$  tend vers 0, ce qui peut s'écrire, puisque la fonction  $\delta \mapsto \omega(x, \delta)$  est croissante,

$$\forall \varepsilon > 0, \exists \delta > 0, \omega(x, \delta) < \varepsilon.$$

L'équicontinuité d'une partie  $A$  de  $\mathcal{C}([0, 1], E)$  peut s'écrire

$$\forall \varepsilon > 0, \exists \delta > 0, \forall x \in X, \omega(x, \delta) < \varepsilon.$$

Autrement dit, une partie est équicontinue si elle est "uniformément uniformément continue".

**Théorème 2.2.2** (Théorème d'Arzelà–Ascoli). Soit  $(E, d)$  un espace métrique complet. Une partie  $\mathcal{K}$  de  $\mathcal{W}(E)$  est relativement compacte si et seulement si les deux propriétés suivantes sont satisfaites.

1. Pour tout  $t \in [0, 1]$ , la partie  $\{x_t : x \in \mathcal{K}\}$  de  $E$  est relativement compacte.
2. La partie  $\mathcal{K}$  est équicontinue.

*Démonstration.* Supposons  $\mathcal{K}$  relativement compacte. Pour tout  $t \in [0, 1]$ , l'application d'évaluation  $ev_t : \mathcal{W}(E) \rightarrow E$  qui à  $x$  associe  $x_t$  est continue. L'image par  $ev_t$  de l'adhérence de  $\mathcal{K}$  est donc compacte. Cette image contient l'image par  $ev_t$  de  $\mathcal{K}$ , qui est exactement la partie  $\{x_t : x \in \mathcal{K}\}$  de  $E$ , qui est donc relativement compacte. La première assertion est donc vérifiée.

Montrons maintenant que  $\mathcal{K}$  est équicontinue. Quitte à remplacer  $\mathcal{K}$  par son adhérence, on peut la supposer compacte. Donnons-nous  $\varepsilon > 0$ . Soit  $\{x_1, \dots, x_r\}$  une partie finie de  $\mathcal{K}$  telle que  $\mathcal{K}$  soit recouverte par les boules de rayon  $\varepsilon/3$  centrées en  $x_1, \dots, x_r$ . Soit  $\delta > 0$  tel que pour tout  $i \in \{1, \dots, r\}$ , on ait  $\omega(x_i, \delta) < \varepsilon/3$ .

Soit  $y$  un élément de  $\mathcal{K}$ . Nous allons montrer que  $\omega(y, \delta) < \varepsilon$ . En effet, soit  $i \in \{1, \dots, r\}$  tel que  $D(x_i, y) < \varepsilon/3$ . Alors pour tous  $s, t \in [0, 1]$  tels que  $|s - t| < \delta$ , on a

$$d(y_s, y_t) \leq d(y_s, x_s) + d(x_s, x_t) + d(x_t, y_t) < 3\varepsilon/3,$$

ce qui montre la deuxième assertion.

Considérons maintenant une partie  $\mathcal{K}$  qui satisfait les deux assertions et montrons qu'elle est relativement compacte. Soit  $(x_n)_{n \geq 0} = ((x_{n,t})_{t \in [0,1]})_{n \geq 0}$  une suite d'éléments de  $\mathcal{K}$ . De la première assertion, on déduit que pour tout  $t \in [0,1]$ , il est possible d'extraire de la suite  $(x_{n,t})_{n \geq 0}$  une suite convergente dans  $E$ . Par le procédé d'extraction diagonale, on peut trouver une extraction  $\psi : \mathbb{N} \rightarrow \mathbb{N}$  telle que la suite  $(x_{\psi(n),t})_{n \geq 0}$  converge pour tout  $t \in \mathbb{Q} \cap [0,1]$ , vers une limite qu'on note  $x_t$ .

Montrons que cette fonction  $x$ , qui n'est pour l'instant définie que sur  $\mathbb{Q} \cap [0,1]$ , y est uniformément continue. Donnons-nous  $\varepsilon > 0$ . L'équicontinuité de  $\mathcal{K}$  nous donne  $\delta > 0$  tel que  $\omega(x_{\psi(n)}, \delta) < \varepsilon/2$  pour tout  $n \geq 0$ . Ainsi, pour tous  $s, t \in \mathbb{Q} \cap [0,1]$  tels que  $|s - t| < \delta$ , on a pour tout  $n \geq 0$  l'inégalité  $d(x_{\psi(n),s}, x_{\psi(n),t}) < \varepsilon/2$ , qui en faisant tendre  $n$  vers l'infini donne  $d(x_s, x_t) \leq \varepsilon/2 < \varepsilon$ .

La fonction  $x : \mathbb{Q} \cap [0,1] \rightarrow E$ , uniformément continue sur une partie dense de  $[0,1]$ , se prolonge en une fonction continue  $x : [0,1] \rightarrow E$ . Il reste à montrer que la suite  $(x_{\psi(n)})_{n \geq 0}$  converge vers  $x$ . Donnons-nous  $\varepsilon > 0$ . Soit  $\delta$  tel que pour tout  $y \in \mathcal{K}$ , on ait  $\omega(y, \delta) < \varepsilon/3$ . Alors d'après le raisonnement fait au paragraphe précédent,  $\omega(x, \delta) \leq \varepsilon/3$ . Soit  $r \geq 1$  tel que  $1/r < \delta$ . Soit  $n_0 \geq 1$  tel que pour tout  $n \geq n_0$ , on ait, pour tout  $i \in \{0, \dots, r\}$ ,  $d(x_{\psi(n),i/r}, x_{i/r}) < \varepsilon/3$ . Soit maintenant  $t \in [0,1]$ . Soit  $i \in \{0, \dots, r-1\}$  tel que  $t \in [i/r, (i+1)/r]$ . Alors pour tout  $n \geq n_0$ ,

$$d(x_{\psi(n),t}, x_t) \leq d(x_{\psi(n),t}, x_{\psi(n),i/r}) + d(x_{\psi(n),i/r}, x_{i/r}) + d(x_{i/r}, x_t) < \varepsilon.$$

Ceci achève la démonstration. □

En examinant la preuve, on voit que quitte à en adapter légèrement le dernier argument, on peut affaiblir la première assertion, de la manière suivante.

**Corollaire 2.2.3.** *Soit  $(E, d)$  un espace métrique complet. Une partie  $\mathcal{K}$  de  $\mathcal{W}(E)$  est relativement compacte si et seulement si les deux propriétés suivantes sont satisfaites.*

1. *Il existe une partie dense  $Q \subset [0,1]$  telle que pour tout  $t \in Q$ , la partie  $\{x_t : x \in \mathcal{K}\}$  de  $E$  soit relativement compacte.*

2. *La partie  $\mathcal{K}$  est équicontinue.*

**Exercice 2.2.** Considérons  $E = [0,1[$  muni de la distance usuelle. Définissons

$$L = \{t \mapsto ct : [0,1] \rightarrow E, c \in [0,1[ \} \subset \mathcal{W}(E).$$

La partie  $L$  est-elle relativement compacte? Si non, quelles hypothèses du corollaire précédent ne sont-elles pas vérifiées?

On travaille souvent dans le cas où  $E$  est égal à  $\mathbb{R}$  ou à un espace vectoriel normé de dimension finie. Dans ces espaces, les parties relativement compactes sont les parties bornées, si bien qu'on peut encore affaiblir les hypothèses de la manière suivante.

Convenons d'appeler, dans un espace métrique, *bornée* une partie de diamètre fini. Avec des symboles, une partie  $B$  de  $(E, d)$  est bornée si  $\sup\{d(x, y) : x, y \in B\} < \infty$ . Il est équivalent de demander que pour tout  $x \in B$ , la fonction  $d(x, \cdot)$  soit bornée sur  $B$ , ou encore qu'il existe un  $x \in B$  tel que la fonction  $d(x, \cdot)$  soit bornée sur  $B$ . Notons que l'adhérence d'une partie bornée d'un espace métrique est toujours bornée.

**Corollaire 2.2.4.** Soit  $(E, d)$  un espace métrique dans lequel les parties fermées et bornées sont compactes. Une partie  $\mathcal{K}$  de  $\mathcal{W}(E)$  est relativement compacte si et seulement si les deux propriétés suivantes sont satisfaites.

1. La partie  $\{x_0 : x \in \mathcal{K}\}$  de  $E$  est bornée.
2. La partie  $\mathcal{K}$  est équicontinue.

*Démonstration.* Soit  $\mathcal{K}$  une partie de  $\mathcal{W}(E)$  qui satisfait les deux propriétés. Nous allons montrer que pour tout  $t \in [0, 1]$ , la partie  $\{x_t : x \in \mathcal{K}\}$  de  $E$  est bornée. Soit  $\delta > 0$  tel que pour tout  $x \in \mathcal{K}$ , on ait  $\omega(x, \delta) < 1$ . Soit  $r \geq 1$  tel que  $1/r < \delta$ . Alors de tout point de  $[0, 1]$  on peut aller à 0 par une suite d'au plus  $r$  sauts de largeur inférieure à  $\delta$ . Il s'ensuit que pour tout  $x \in \mathcal{K}$  et tout  $t \in [0, 1]$ , on a  $d(x_t, x_0) \leq r$ . Pour tout  $t \in [0, 1]$  et tous  $y, z \in \mathcal{K}$ , on a donc

$$d(y_t, z_t) \leq d(y_t, y_0) + d(y_0, z_0) + d(z_0, z_t) \leq r + \text{diam}\{x_0 : x \in \mathcal{K}\} + r < \infty,$$

ce qui montre la propriété annoncée. □

## 2.3 Convergence en loi dans $\mathcal{W}(E)$

### 2.3.1 Tribu borélienne et tribu cylindrique

Soit  $(E, d)$  un espace métrique. Nous allons maintenant étudier des convergences de mesures de probabilités sur  $\mathcal{W} = \mathcal{W}(E) = \mathcal{C}([0, 1], E)$ . La tribu dont nous munissons cet ensemble est la tribu borélienne  $\mathcal{B}_{\mathcal{W}}$  de la topologie induite par la distance  $D$ . Il y a cependant une autre tribu naturelle sur cet ensemble. Rappelons que nous avons défini, pour tout  $t \in [0, 1]$ , l'application d'évaluation

$$\begin{aligned} \text{ev}_t : \mathcal{C}([0, 1], E) &\longrightarrow E \\ x &\longmapsto x_t. \end{aligned}$$

On peut considérer la tribu

$$\mathcal{C}_{\mathcal{W}} = \sigma(\text{ev}_t : t \in [0, 1])$$

qui est la plus petite tribu sur  $\mathcal{W}$  qui rende mesurables toutes les applications d'évaluation. Avec les applications probabilistes en tête, c'est la tribu  $\mathcal{C}_{\mathcal{W}}$  qui nous intéresse. En revanche, pour pouvoir utiliser les résultats que nous avons établis jusqu'ici, il nous faut considérer  $\mathcal{B}_{\mathcal{W}}$ .

**Proposition 2.3.1.** Les tribus  $\mathcal{B}_{\mathcal{W}}$  et  $\mathcal{C}_{\mathcal{W}}$  sur  $\mathcal{C}([0, 1], E)$  sont égales.

*Démonstration.* Les applications d'évaluation sont continues, donc boréliennes. On a donc l'inclusion  $\mathcal{C}_{\mathcal{W}} \subseteq \mathcal{B}_{\mathcal{W}}$ .

Pour montrer l'autre inclusion, nous allons montrer que toute boule ouverte de  $\mathcal{W}(E)$  appartient à  $\mathcal{C}_{\mathcal{W}}$ . Soit  $x \in \mathcal{W}(E)$  et  $r > 0$ . Tout d'abord, on a, en mettant en indice des boules l'espace dans lequel elles sont définies,

$$B_{\mathcal{W}}(x, r) = \bigcup_{s \in \mathbb{Q}, 0 \leq s < r} \bar{B}_{\mathcal{W}}(x, s).$$

Il suffit donc de montrer que toute boule fermée appartient à  $\mathcal{C}_{\mathcal{W}}$ . Or

$$\begin{aligned}\bar{B}_{\mathcal{W}}(x, r) &= \{y \in \mathcal{W}(E) : \forall t \in [0, 1], d(x_t, y_t) \leq r\} \\ &= \{y \in \mathcal{W}(E) : \forall t \in \mathbb{Q} \cap [0, 1], d(x_t, y_t) \leq r\} \\ &= \bigcap_{s \in \mathbb{Q} \cap [0, 1]} \text{ev}_s^{-1}(\bar{B}_E(x_s, r))\end{aligned}$$

qui appartient à  $\mathcal{C}_{\mathcal{W}}$ . □

Définissons, pour tout  $r \geq 1$  et tout  $r$ -uplet  $(t_1, \dots, t_r)$  de réels tels que  $0 \leq t_1 < \dots < t_r \leq 1$ , une application

$$\begin{aligned}\text{ev}_{t_1, \dots, t_r} : \mathcal{C}([0, 1], E) &\longrightarrow E^r \\ x &\longmapsto (x_{t_1}, \dots, x_{t_r}).\end{aligned}$$

Les parties de  $\mathcal{W}(E)$  de la forme  $\text{ev}_{t_1, \dots, t_r}^{-1}(B)$ , où  $B$  est un borélien de  $E^r$ , s'appellent des parties cylindriques. La classe des parties cylindriques de  $\mathcal{W}(E)$  est un  $\pi$ -système (il faut le vérifier) qui engendre la tribu  $\mathcal{C}_{\mathcal{W}}$ , donc la tribu borélienne.

**Définition 2.3.2.** Soit  $\mu$  une mesure de probabilité borélienne sur  $\mathcal{C}([0, 1], E)$ . Les mesures images de  $\mu$  par les applications  $\text{ev}_{t_1, \dots, t_r}$ , avec  $r \geq 1$  et  $0 \leq t_1 < \dots < t_r \leq 1$ , s'appellent les lois marginales fini-dimensionnelles, ou plus simplement les lois marginales, ou marginales, ou encore marges de la mesure  $\mu$ .

Pour toute mesure de probabilité  $\mu \in \mathcal{M}(\mathcal{W}(E))$ , pour tout entier  $r \geq 1$  et tous réels  $0 \leq t_1 < \dots < t_r \leq 1$ , nous noterons

$$\mu^{t_1, \dots, t_r} = \mu \circ \text{ev}_{t_1, \dots, t_r}^{-1} \in \mathcal{M}(E^r). \quad (2.1)$$

**Proposition 2.3.3.** Deux mesures de probabilité sur  $\mathcal{C}([0, 1], E)$  sont égales si et seulement si elles ont les mêmes lois marginales.

*Démonstration.* En effet, dire que deux mesures de probabilités sur  $\mathcal{C}([0, 1], E)$  ont les mêmes lois marginales équivaut à dire qu'elles co $\tilde{\mathcal{A}}$ -ncident sur la classe des cylindres. Or cette classe est un  $\pi$ -système, qui engendre la tribu borélienne. Par ailleurs, la classe des parties mesurables sur lesquelles deux mesures de probabilité co $\tilde{\mathcal{A}}$ -ncident est toujours un  $\lambda$ -système. Le  $\lambda$ -système sur lequel nos deux mesures co $\tilde{\mathcal{A}}$ -ncident contient donc le  $\pi$ -système des parties cylindriques, donc, d'après le lemme de la classe monotone, la tribu engendrée par ce  $\pi$ -système : les deux mesures sont donc égales. □

### 2.3.2 Convergence faible dans $\mathcal{W}(E)$

On dit qu'une suite de mesures de probabilités  $(\mu_n)_{n \geq 0}$  sur  $\mathcal{W}(E)$  converge vers une mesure  $\mu$  au sens des lois marginales fini-dimensionnelles si pour tous  $r \geq 1$  et tous  $0 \leq t_1 < \dots < t_r \leq 1$ , on a dans  $\mathcal{M}(E^r)$  la convergence faible

$$\mu_n^{t_1, \dots, t_r} \xrightarrow[n \rightarrow \infty]{} \mu^{t_1, \dots, t_r}.$$

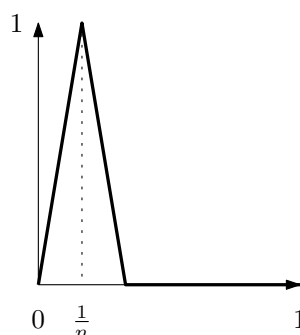
**Proposition 2.3.4.** *Si une suite de mesures sur  $\mathcal{W}(E)$  converge faiblement, elle converge au sens des lois marginales fini-dimensionnelles vers la même limite.*

*Démonstration.* Il suffit en effet d'observer que les applications  $ev_{t_1, \dots, t_r}$  sont continues, et de se rappeler que l'image par une application continue d'une suite faiblement convergente de mesures est une suite faiblement convergente.  $\square$

La réciproque est fautive. Considérons par exemple le cas où  $E = \mathbb{R}$  et, pour tout  $n \geq 1$ , définissons la fonction  $z_n : [0, 1] \rightarrow \mathbb{R}$  en posant, pour tout  $t \in [0, 1]$ ,

$$z_{n,t} = \max(1 - n|t - \frac{1}{n}|, 0).$$

Le graphe de cette fonction est le suivant :



On vérifie que la suite  $(\delta_{z_n})_{n \geq 0}$  des masses de Dirac en les fonctions  $z_n$  converge au sens des lois marginales fini-dimensionnelles vers la masse de Dirac  $\delta_0$  en la fonction nulle. Toutefois, la fonction  $f : \mathcal{C}([0, 1], \mathbb{R}) \rightarrow \mathbb{R}$  définie par  $f(x) = \min(\|x\|, 1)$  est continue et bornée et elle vérifie

$$1 = \int_{\mathcal{C}([0,1], \mathbb{R})} f d\delta_{z_n} \not\xrightarrow[n \rightarrow \infty]{} \int_{\mathcal{C}([0,1], \mathbb{R})} f d\delta_0 = 0.$$

Ceci montre qu'on n'a pas convergence faible de la suite  $(\delta_{z_n})_{n \geq 0}$  vers  $\delta_0$ .

**Exercice 2.3.** Vérifier précisément les assertions qui précèdent.

Avec le résultat suivant, nous commençons à voir les fils de notre étude se rassembler.

**Théorème 2.3.5.** *Soit  $(E, d)$  un espace polonais. Une suite  $(\mu_n)_{n \geq 0}$  de mesures de probabilités sur  $\mathcal{C}([0, 1], E)$  converge faiblement vers une mesure  $\mu$  si et seulement si les deux propriétés suivantes sont vérifiées.*

1. La suite  $(\mu_n)_{n \geq 0}$  est tendue.
2. La suite  $(\mu_n)_{n \geq 0}$  converge au sens des lois marginales fini-dimensionnelles vers  $\mu$ .

*Démonstration.* Supposons que la suite  $(\mu_n)_{n \geq 0}$  converge faiblement vers  $\mu$ . Nous avons déjà dit à la proposition 2.3.4 que la convergence avait lieu au sens des lois marginales fini-dimensionnelles. Par ailleurs, il est vrai dans tout espace topologique que l'ensemble formé par une suite convergente et sa limite est compact. Dans notre cas, grâce au théorème de Prokhorov, il s'ensuit que la suite  $(\mu_n)_{n \geq 0}$  est tendue.



Supposons maintenant les deux propriétés vérifiées. Pour montrer que la suite  $(\mu_n)_{n \geq 0}$  converge faiblement vers  $\mu$ , nous allons montrer que toute sous-suite convergente de cette suite converge faiblement vers  $\mu$ . Mais si une sous-suite de la suite  $(\mu_n)_{n \geq 0}$  converge faiblement vers une limite  $\nu$ , elle converge au sens des lois marginales fini-dimensionnelles vers  $\nu$ , si bien que  $\nu$  a les mêmes lois marginales que  $\mu$ , et est donc égale à  $\mu$ , par la proposition 2.3.3.

Pour conclure la démonstration, on invoque la proposition suivante, que nous énonçons indépendamment car il est souvent utile.  $\square$

**Proposition 2.3.6.** *Soit  $\mu$  une mesure sur un espace polonais. Si une suite  $(\mu_n)_{n \geq 0}$  de mesures sur cet espace polonais est tendue, et si toute sous-suite convergente de cette suite converge vers  $\mu$ , alors la suite  $(\mu_n)_{n \geq 0}$  elle-même converge vers  $\mu$ .*

*Démonstration.* Raisonnons par l'absurde et supposons que la suite  $(\mu_n)_{n \geq 0}$  ne converge pas vers  $\mu$ . Alors il existe un voisinage  $V$  de  $\mu$  et une sous-suite  $(\mu_{\psi(n)})_{n \geq 0}$  de notre suite qui reste à l'extérieur de  $V$ . Mais la sous-suite est encore tendue, donc relativement compacte d'après le théorème de Prokhorov. On peut donc en extraire une sous-sous-suite convergente. Or d'une part cette sous-sous-suite reste hors de  $V$  et d'autre part, elle converge vers  $\mu$  par hypothèse : c'est contradictoire.  $\square$

La convergence au sens des lois marginales fini-dimensionnelles n'est en général pas, des deux hypothèses du théorème précédent, la plus difficile à établir. La partie délicate de l'étude des convergences faibles de mesures de probabilités sur des espaces de fonctions est celle qui consiste à montrer que les suites de mesures qu'on considère sont tendues.

## 2.4 Critères de tension dans $\mathcal{M}(\mathcal{C}([0, 1], E))$

### 2.4.1 Des traductions successives

Pour appliquer le théorème 2.3.5, nous avons besoin d'avoir des critères aussi pratiques et concrets que possibles de tension d'un ensemble de mesures sur  $\mathcal{W}(E)$ . Pour cela, nous allons bien entendu utiliser le théorème d'Arzelà–Ascoli (théorème 2.2.2).

Commençons par une transcription assez littérale du théorème d'Arzelà–Ascoli. Rappelons la notation (2.1) pour les lois marginales d'une mesure sur  $\mathcal{W}(E)$ .

**Proposition 2.4.1.** *Soit  $(E, d)$  un espace polonais. Soit  $\Gamma$  une partie de  $\mathcal{M}(\mathcal{W}(E))$ . La partie  $\Gamma$  est tendue si et seulement si les deux propriétés suivantes sont vérifiées.*

1. Pour tout  $t \in [0, 1]$ , la famille  $\{\mu^t : \mu \in \Gamma\} \subseteq \mathcal{M}(E)$  est tendue.
2. Pour tous  $\varepsilon > 0$  et  $\eta > 0$ , il existe  $\delta > 0$  tel que

$$\forall \mu \in \Gamma, \mu(\{x \in \mathcal{W}(E) : \omega(x, \delta) \leq \eta\}) \geq 1 - \varepsilon.$$

*Si les fermés bornés de  $E$  sont compacts, on peut remplacer la propriété 1 par la suivante.*

- 1'. La famille  $\{\mu^0 : \mu \in \Gamma\} \subseteq \mathcal{M}(E)$  est tendue.

*Démonstration.* Supposons  $\Gamma$  tendue. Soit  $t \in [0, 1]$ . Soit  $\varepsilon > 0$ . Soit  $\mathcal{K} \subset \mathcal{W}(E)$  un compact tel que  $\mu(\mathcal{K}) \geq 1 - \varepsilon$  pour tout  $\mu \in \Gamma$ . Alors la partie  $\text{ev}_t(\mathcal{K})$  de  $E$  est compacte et pour tout  $\mu \in \Gamma$ ,

$$\mu^t(\text{ev}_t(\mathcal{K})) = \mu(\{x : x_t \in \text{ev}_t(\mathcal{K})\}) \geq \mu(\mathcal{K}) \geq 1 - \varepsilon.$$

Ceci prouve la première propriété. Pour la seconde, donnons-nous  $\varepsilon$  et  $\eta$  strictement positifs. Soit comme précédemment  $\mathcal{K}$  un compact tel que  $\mu(\mathcal{K}) \geq 1 - \varepsilon$  pour tout  $\mu \in \Gamma$ . D'après le théorème d'Arzelà–Ascoli,  $\mathcal{K}$  est équicontinue. Il existe donc  $\delta > 0$  tel que  $\omega(x, \delta) < \eta$  pour tout  $x \in \mathcal{K}$ . Alors pour tout  $\mu \in \Gamma$ , on a

$$\mu(\{x \in \mathcal{W}(E) : \omega(x, \delta) \leq \eta\}) \geq \mu(\mathcal{K}) \geq 1 - \varepsilon,$$

ce qui prouve la seconde propriété.

Démontrons maintenant la réciproque. Soit  $\Gamma$  satisfaisant les deux assertions. Soit  $\varepsilon > 0$ . Pour tout entier  $n \geq 1$ , soit  $\delta_n > 0$  tel que pour tout  $\mu \in \Gamma$  on ait

$$\mu(\{x \in \mathcal{W}(E) : \omega(x, \delta_n) \leq \frac{1}{n}\}) \geq 1 - 2^{-n} \frac{\varepsilon}{2}.$$

Par ailleurs, soit  $(q_n)_{n \geq 1}$  une énumération des rationnels entre 0 et 1. Pour tout  $n \geq 1$ , soit  $K_n$  une partie compacte de  $E$  telle que pour tout  $\mu \in \Gamma$ , on ait

$$\mu^{q_n}(K_n) \geq 1 - 2^{-n} \frac{\varepsilon}{2}.$$

Posons

$$\mathcal{K} = \bigcap_{n \geq 1} \{x \in \mathcal{W}(E) : \omega(x, \delta_n) \leq \frac{1}{n}\} \cap \bigcap_{n \geq 1} \text{ev}_{q_n}^{-1}(K_n).$$

Alors  $\mathcal{K}$  est une partie fermée de  $\mathcal{W}(E)$  et d'après le corollaire 2.2.3, elle est relativement compacte. Elle est donc compacte. De plus, pour tout  $\mu \in \Gamma$ , on a

$$\mu(\mathcal{K}) \geq 1 - \varepsilon,$$

ce qui montre la tension de  $\Gamma$ .

Le cas où les fermés bornés de  $E$  sont compacts est tout à fait similaire.  $\square$

On peut traduire le résultat que nous venons d'obtenir dans le langage des probabilités, plutôt que celui de la théorie de la mesure. Nous appelons *processus (stochastique) continu à valeurs dans  $E$*  une variable aléatoire définie sur un espace de probabilité quelconque et à valeurs dans  $(\mathcal{W}(E), \mathcal{B}_{\mathcal{W}})$ . On dit qu'une famille de telles variables aléatoires, ou de variables aléatoires à valeurs dans un espace métrique quelconque, est tendue si la famille des lois de ces variables aléatoires est tendue.

Nous profitons de cette nouvelle traduction pour modifier légèrement l'expression de la deuxième condition.

**Proposition 2.4.2.** *Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité sur lequel est définie une suite  $(X_n)_{n \geq 0}$  de processus continus à valeurs dans  $E$ . La suite  $(X_n)_{n \geq 0}$  est tendue si et seulement si les deux propriétés suivantes sont vérifiées.*

1. Pour tout  $t \in [0, 1]$ , la suite  $(X_{n,t})_{n \geq 0}$  est tendue.
2. Pour tous  $\varepsilon > 0$  et  $\eta > 0$ , il existe  $\delta > 0$  tel que

$$\underline{\lim}_{n \rightarrow \infty} \mathbb{P}(\omega(X_n, \delta) \leq \eta) \geq 1 - \varepsilon.$$

Si les fermés bornés de  $E$  sont compacts, on peut remplacer la propriété 1 par la suivante.

- 1'. La famille  $(X_{n,0})_{n \geq 0}$  est tendue.

*Démonstration.* Une transcription littérale de l'énoncé précédent nous aurait donné un infimum sur  $n$  à la place d'une limite inférieure dans la deuxième condition, ce qui semble plus faible. Supposons la propriété 2 vérifiée et donnons-nous  $\varepsilon > 0$  et  $\eta > 0$ . Soit  $\delta > 0$  tel qu'on ait  $\underline{\lim}_{n \rightarrow \infty} \mathbb{P}(\omega(X_n, \delta) \leq \eta) \geq 1 - \varepsilon/2$ . Il existe  $n_0$  tel que

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

Quitte à replacer  $\delta$  par un nombre strictement plus petit, on peut supposer que pour tout entier  $n \in \{0, \dots, n_0 - 1\}$ , on a aussi  $\mathbb{P}(\omega(X_n, \delta) \leq \eta) \geq 1 - \varepsilon$ . Ainsi, la propriété 2 de la proposition 2.4.1 est vérifiée.  $\square$

## 2.4.2 Un premier critère utile

La proposition 2.4.2 est difficile à utiliser, car elle demande de vérifier une propriété des modules de continuité des processus considérés, ce qui en général n'est pas très commode. Nous allons voir qu'il est possible de donner une condition suffisante de tension qui ne fait intervenir que les lois marginales fini-dimensionnelles des processus.

**Proposition 2.4.3.** Soit  $(X_n)_{n \geq 0}$  une suite de processus continus à valeurs dans  $E$ . On suppose que les deux propriétés suivantes sont vérifiées.

1. Pour tout  $t \in [0, 1]$ , la suite  $(X_{n,t})_{n \geq 0}$  est tendue.
2. Pour tous  $\varepsilon > 0$  et  $\eta > 0$ , il existe un entier  $k \geq 1$  tel que

$$\overline{\lim}_{n \rightarrow \infty} k \max_{0 \leq i \leq k-1} \mathbb{P}\left(\sup \{d(X_{n, \frac{i}{k}}, X_{n,t}) : t \in [\frac{i}{k}, \frac{i+1}{k}]\} \geq \eta\right) \leq \varepsilon.$$

Alors la suite  $(X_n)_{n \geq 0}$  est tendue.

Si les fermés bornés de  $E$  sont compacts, on peut remplacer la propriété 1 par la suivante.

- 1'. La famille  $(X_{n,0})_{n \geq 0}$  est tendue.

*Démonstration.* Soit  $x \in \mathcal{W}(E)$  et  $k \geq 1$  un entier. Notons

$$\alpha = \max_{0 \leq i \leq k-1} \sup \{d(x_{\frac{i}{k}}, x_t) : t \in [\frac{i}{k}, \frac{i+1}{k}]\}.$$

Soient  $s, t \in [0, 1]$  deux réels tels que  $s \leq t$  et  $|s - t| \leq \frac{1}{k}$ . De deux choses l'une : soit il existe  $i \in \{0, \dots, k-1\}$  tel que  $\frac{i}{k} \leq s \leq t < \frac{i+1}{k}$ , auquel cas

$$d(x_s, x_t) \leq d(x_s, x_{\frac{i}{k}}) + d(x_{\frac{i}{k}}, x_t) \leq 2\alpha,$$

soit il existe  $i \in \{0, \dots, k-2\}$  tel que  $\frac{i}{k} \leq s < \frac{i+1}{k} \leq t < \frac{i+2}{k}$ , auquel cas

$$d(x_s, x_t) \leq d(x_s, x_{\frac{i}{k}}) + d(x_{\frac{i}{k}}, x_{\frac{i+1}{k}}) + d(x_{\frac{i+1}{k}}, x_t) \leq 3\alpha.$$

Dans tous les cas, on a  $d(x_s, x_t) \leq 3\alpha$ . Nous avons donc montré que

$$\omega(x, \frac{1}{k}) \leq 3 \max_{0 \leq i \leq k-1} \sup \{d(x_{\frac{i}{k}}, x_t) : t \in [\frac{i}{k}, \frac{i+1}{k}]\}.$$

Pour tout  $n \geq 1$  et tout  $\eta > 0$ , on a donc

$$\begin{aligned} \mathbb{P}(\omega(X_n, \frac{1}{k}) > \eta) &\leq \mathbb{P}(3 \max_{0 \leq i \leq k-1} \sup \{d(X_{n, \frac{i}{k}}, X_{n, t}) : t \in [\frac{i}{k}, \frac{i+1}{k}]\} > \eta) \\ &\leq k \max_{0 \leq i \leq k-1} \mathbb{P}(\sup \{d(X_{n, \frac{i}{k}}, X_{n, t}) : t \in [\frac{i}{k}, \frac{i+1}{k}]\} > \frac{\eta}{3}). \end{aligned}$$

Les hypothèses de la proposition entraînent donc, pour tout  $\varepsilon > 0$ , l'existence de  $k$  tel que la limite supérieure de cette dernière quantité est inférieure à  $\varepsilon$ . On peut donc appliquer la proposition 2.4.2 et conclure.  $\square$

### 2.4.3 Le critère de Kolmogorov

Dans les situations où il s'applique, le critère de Kolmogorov est de loin le plus simple de ceux que nous allons étudier. Il donne, en plus de la tension, une mesure de la régularité des trajectoires des processus limites. Pour  $x \in \mathcal{W}(E)$  et  $\alpha \geq 0$ , notons

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

On appelle ce nombre norme de Hölder d'exposant  $\alpha$  de  $x$ . On dit que  $x$  est höldérien d'exposant  $\alpha$  si  $N_\alpha(x) < \infty$ .

**Exercice 2.4.** Qu'est-ce qu'un chemin 0-höldérien? Un chemin 1-höldérien? Un chemin  $\alpha$ -höldérien pour  $\alpha > 1$ ?

**Théorème 2.4.4** (Critère de Kolmogorov). Soit  $(X_n)_{n \geq 0}$  une suite de processus continus à valeurs dans  $E$ . On suppose que les deux propriétés suivantes sont vérifiées.

1. Pour tout  $t \in [0, 1]$ , la suite  $(X_{n, t})_{n \geq 0}$  est tendue.
2. Il existe des réels strictement positifs  $\beta, p, C$  tels que pour tout  $n \geq 0$  et tous  $s, t \in [0, 1]$ , on ait

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

Alors la suite  $(X_n)_{n \geq 0}$  est tendue. De plus, pour tout  $\alpha$  tel que  $0 < \alpha < \frac{\beta}{p}$ , et pour tout  $\varepsilon > 0$ , il existe  $M \geq 0$  tel que

$$\sup_{n \geq 0} \mathbb{P}(N_\alpha(X_n) \geq M) \leq \varepsilon. \quad (2.2)$$

Si les fermés bornés de  $E$  sont compacts, on peut remplacer la propriété 1 par la suivante.

- 1'. La famille  $(X_{n, 0})_{n \geq 0}$  est tendue.

Commençons par démontrer un lemme combinatoire. Pour tout  $k \geq 0$ , notons

$$D_k = \{0, \frac{1}{2^k}, \dots, \frac{2^k-1}{2^k}, 1\}$$

l'ensemble des multiples de  $2^{-k}$  compris entre 0 et 1. Notons

$$D = \bigcup_{k \geq 0} D_k$$

l'ensemble des nombres dyadiques compris entre 0 et 1.

**Lemme 2.4.5** (Lemme de chaînage). *Soit  $x$  un élément de  $\mathcal{W}(E)$ . Supposons qu'il existe  $\alpha > 0$  et  $K \geq 0$  tels que pour tout  $k \geq 0$  et tous  $s$  et  $t$  éléments consécutifs de  $D_k$  on ait*

$$d(x_s, x_t) \leq K|t - s|^\alpha.$$

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

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

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

*Démonstration.* Soient  $s, t$  deux éléments de  $D$  tels que  $s \leq t$ . Notons  $\ell \geq 1$  l'unique entier tel que

$$2^{-\ell} < t - s \leq 2^{-(\ell-1)}.$$

Les réels  $s$  et  $t$  ne sont pas dans le même intervalle fermé délimité par  $D_\ell$  mais il peuvent, ou non, être dans le même intervalle fermé délimité par  $D_{\ell-1}$ . Nous allons discuter suivant qu'ils le sont ou non.

Supposons d'abord qu'il existe  $q \in D_{\ell-1}$  tel que  $q \leq s < t < q + 2^{-(\ell-1)}$ . On peut alors écrire

$$s = q + \sum_{k \geq \ell} \frac{a_k}{2^k} \quad \text{et} \quad t = q + \sum_{k \geq \ell} \frac{b_k}{2^k},$$

où  $(a_k)_{k \geq \ell}$  et  $(b_k)_{k \geq \ell}$  sont deux suites presque nulles de 0 et de 1. Notons, pour tout  $n \geq \ell - 1$ ,

$$s_n = q + \sum_{k=\ell}^n \frac{a_k}{2^k} \quad \text{et} \quad t_n = q + \sum_{k=\ell}^n \frac{b_k}{2^k},$$

étant entendu que  $s_{\ell-1} = t_{\ell-1} = q$ . Notons enfin  $L \geq 1$  un entier tel que  $s_L = s$  et  $t_L = t$ . Alors

$$\begin{aligned} d(x_s, x_t) &= d(x_{s_L}, x_{t_L}) \\ &\leq d(x_{s_L}, x_{s_{L-1}}) + \dots + d(x_{s_\ell}, x_q) + d(x_q, x_{t_\ell}) + \dots + d(x_{t_{L-1}}, x_{t_L}) \\ &\leq \sum_{k \geq \ell-1} d(x_{s_k}, x_{s_{k+1}}) + \sum_{k \geq \ell-1} d(x_{t_k}, x_{t_{k+1}}). \end{aligned}$$

Or pour tout  $k \geq \ell - 1$ ,  $s_k$  et  $s_{k+1}$  (de même que  $t_k$  et  $t_{k+1}$ ) sont égaux ou consécutifs dans  $k + 1$ . Ainsi,

$$d(x_s, x_t) \leq 2K \sum_{k \geq \ell} 2^{-k\alpha} = \frac{2K}{1 - 2^{-\alpha}} 2^{-\ell\alpha} \leq \frac{2K}{1 - 2^{-\alpha}} |t - s|^\alpha.$$

Supposons maintenant qu'il existe  $q \in \ell - 1$  tel que  $s < q \leq t$ . Quitte à écarter le cas où  $s = q - 2^{-(\ell-1)}$  et  $t = q$  sont consécutifs dans  $\ell - 1$ , cas où nous savons déjà que l'inégalité cherchée est vraie, on a donc les inégalités  $q - 2^{-(\ell-1)} < s < q \leq t < q + 2^{-(\ell-1)}$ . On peut alors écrire

$$s = q - \sum_{k \geq \ell} \frac{a_k}{2^k} \quad \text{et} \quad t = q + \sum_{k \geq \ell} \frac{b_k}{2^k},$$

où  $(a_k)_{k \geq \ell}$  et  $(b_k)_{k \geq \ell}$  sont deux suites presque nulles de 0 et de 1. Un raisonnement strictement analogue au précédent amène dans ce cas encore à la conclusion.  $\square$

Notons que nous n'avons utilisé nulle part dans la démonstration qui précède l'hypothèse que  $x$  était continue. On ne l'utilise que pour déduire de la majoration de  $d(x_s, x_t)$  lorsque  $s, t \in D$  la majoration de  $N_\alpha(x)$ .

*Démonstration du théorème 2.4.4.* Pour tout  $k \geq 0$  et tout  $n \geq 0$ , notons

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

L'inégalité de Markov nous donne

$$\begin{aligned} \mathbb{P}(Z_{n,k} \geq K2^{-k\alpha}) &\leq 2^k \max_{0 \leq i \leq 2^k - 1} \mathbb{P}(d(X_{n, \frac{i}{2^k}}, X_{n, \frac{i+1}{2^k}}) \geq K2^{-k\alpha}) \\ &\leq 2^k C 2^{-k(1+\beta)} K^{-p} 2^{pk\alpha} \\ &= CK^{-p} 2^{-k(\beta-p\alpha)}. \end{aligned}$$

Si  $0 < \alpha < \frac{\beta}{p}$ , alors la puissance de 2 est négative. Dans ce cas, le lemme 2.4.5 nous donne

$$\mathbb{P}\left(N_\alpha(X_n) > \frac{2K}{1 - 2^{-\alpha}}\right) \leq \mathbb{P}(\exists k \geq 0, Z_{n,k} > K2^{-k\alpha}) \leq CK^{-p} \sum_{k \geq 0} 2^{-k(\beta-p\alpha)} = \frac{C}{K^p(1 - 2^{-(\beta-p\alpha)}}.$$

Cette majoration qui ne dépend pas de  $n$  suffit à démontrer (2.2).

Enfin, l'inégalité  $\omega(X_n, \delta) \leq N_\alpha(X_n)\delta^\alpha$  entraîne

$$\mathbb{P}(\omega(X_n, \delta) > \eta) \leq \mathbb{P}(N_\alpha(X_n) > \eta\delta^{-\alpha})$$

et les hypothèses de la proposition 2.4.2 sont vérifiées.  $\square$

## 2.5 Le théorème de Donsker

### 2.5.1 La mesure de Wiener

Dans cette courte section, il s'agit moins pour nous de faire un cours complet sur le mouvement brownien que de rappeler ce dont il s'agit, et ce qui nous sera utile pour la démonstration du théorème de Donsker.

Rappelons qu'on appelle mesure de Wiener une mesure borélienne de probabilités sur  $\mathcal{C}([0, 1], \mathbb{R})$  sous laquelle le processus canonique est nul en 0, centré, à accroissement indépendants, et d'accroissement sur l'intervalle  $[s, t]$  gaussien centré de variance  $t - s$ .

Qu'il existe une telle mesure ne va pas du tout de soi, et peut se démontrer de plusieurs manières.

- On peut exhiber les marges fini-dimensionnelles de la mesure de Wiener, vérifier qu'elles sont cohérentes, appliquer le théorème d'extension de Kolmogorov, puis appliquer le critère de régularité de Kolmogorov pour montrer que le processus construit admet une version continue.
- On peut aussi partir d'un isomorphisme abstrait  $W$  entre l'espace de Hilbert  $L^2([0, 1])$  et un espace de Hilbert de variables aléatoires gaussiennes, en définissant pour tout  $t \geq 0$   $B_t = W(\mathbb{1}_{[0, t]})$ , puis en appliquant le même critère de régularité de Kolmogorov pour obtenir une version continue de ce processus.
- On peut donner une variante plus explicite de la construction précédente, et plus fidèle à l'esprit de la construction originale de Wiener, en réalisant le mouvement brownien comme somme d'une série presque sûrement normalement convergente de fonctions continues aléatoires sur  $[0, 1]$ . Cette construction est détaillée à l'appendice E.
- On peut aussi déduire l'existence de la mesure de Wiener du théorème de Donsker, que nous allons démontrer.

La seule chose que nous ayons besoin de savoir sur la mesure de Wiener est la suivante : si un processus gaussien centré  $(B_t)_{t \in [0, 1]}$  est à trajectoires continues et de covariance  $\mathbb{E}[B_s B_t] = \min(s, t)$ , alors sa loi est la mesure de Wiener. Autrement dit, c'est un mouvement brownien.

### 2.5.2 Convergence des marches aléatoires vers le mouvement brownien

Nous pouvons maintenant appliquer notre étude au cas de la convergence des marches aléatoires vers le mouvement brownien.

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité. Sur cet espace, soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles indépendantes et identiquement distribuées, admettant un moment d'ordre 2. Quitte à leur appliquer une transformation affine, on peut supposer, et nous supposons, qu'elles sont centrées et de variance 1 :

$$\mathbb{E}[X_1] = 0 \quad \text{et} \quad \text{Var}(X_1) = 1.$$

Posons  $S_0 = 0$  et, pour tout réel  $u \geq 0$ ,

$$S_u = X_1 + \dots + X_{[u]} + (u - [u])X_{[u]+1}.$$

Enfin, pour tout  $n \geq 1$ , définissons une variable aléatoire  $M_n$  à valeurs dans  $\mathcal{C}([0, 1], \mathbb{R})$  en posant, pour tout  $t \in [0, 1]$ ,

$$M_{n,t} = \frac{S_{nt}}{\sqrt{n}}.$$

**Théorème 2.5.1** (Théorème de Donsker). *La suite  $(M_n)_{n \geq 1}$  converge, dans  $\mathcal{M}(\mathcal{C}([0, 1], \mathbb{R}))$ , vers la mesure de Wiener.*

*Démonstration.* Commençons par étudier la convergence des marges fini-dimensionnelles de la suite  $(M_n)_{n \geq 1}$ . Choisissons  $k$  réels  $0 \leq t_1 \leq \dots \leq t_k \leq 1$ . Les accroissements de  $(M_{n,t_1}, M_{n,t_2} - M_{n,t_1}, \dots, M_{n,t_k} - M_{n,t_{k-1}})$  ne sont pas exactement indépendants, et nous devons commencer par en étudier une approximation.

Pour tout  $n \geq 1$ , les variables aléatoires  $(S_{\lfloor nt_1 \rfloor}, S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor}, \dots, S_{\lfloor nt_k \rfloor} - S_{\lfloor nt_{k-1} \rfloor})$  sont indépendantes, et on a pour tout entier  $i \in \{0, \dots, k-1\}$  l'égalité en loi de  $S_{\lfloor nt_{i+1} \rfloor} - S_{\lfloor nt_i \rfloor}$  et de  $S_{\lfloor nt_{i+1} \rfloor - \lfloor nt_i \rfloor}$ . Le théorème central limite<sup>1</sup> permet donc d'affirmer que

$$\left( \frac{S_{\lfloor nt_1 \rfloor}}{\sqrt{n}}, \dots, \frac{S_{\lfloor nt_k \rfloor}}{\sqrt{n}} \right) \xrightarrow[n \rightarrow \infty]{} (N_1, N_1 + N_2, \dots, N_1 + \dots + N_k),$$

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

Pour tout  $t \in [0, 1]$  et tout  $n \geq 1$ , on a

$$\left| M_{n,t} - \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \right| \leq \frac{|X_{\lfloor nt \rfloor + 1}|}{\sqrt{n}}$$

et cette différence converge donc, lorsque  $n$  tend vers l'infini à  $t$  fixé, vers 0 en probabilité. Le lemme de Slutsky permet d'ajouter à une convergence en loi une autre convergence en loi, pourvu que la limite de la seconde soit constante. On a donc

$$(M_{n,t_1}, \dots, M_{n,t_k}) \xrightarrow[n \rightarrow \infty]{} (N_1, N_1 + N_2, \dots, N_1 + \dots + N_k),$$

ce qui conclut l'étude de la convergence des marges fini-dimensionnelles de  $M_n$ .

Il nous faut maintenant démontrer la tension de la suite  $(M_n)_{n \geq 1}$ . Pour ce faire, nous allons utiliser la proposition 2.4.3.<sup>2</sup>

1. Nous appliquons le théorème central limite à chaque composante de cette suite de vecteurs aléatoires, puis nous combinons les convergences en loi en utilisant l'indépendance, et, par exemple, le théorème de Lévy qui assure l'équivalence de la convergence en loi et de la convergence simple des fonctions caractéristiques.

2. Utiliser le théorème de Kolmogorov nécessiterait plus d'information que nous n'en avons sur la suite  $(X_n)_{n \geq 1}$ , par exemple de savoir qu'elles ont des moments finis d'ordre strictement plus grand que 2. En fait, pour utiliser confortablement le critère de Kolmogorov, il faudrait supposer que nos variables aléatoires admettent un moment d'ordre 4. La situation est ici très analogue à celle que l'on rencontre lorsque l'on démontre la loi forte des grands nombres : si l'on ne suppose rien de plus sur la loi des variables aléatoires que le fait qu'elles sont intégrables, la démonstration est beaucoup plus difficile que si l'on suppose qu'elles admettent un moment d'ordre 2, ou mieux, d'ordre 4.



Nous allons commencer par démontrer une inégalité maximale. Nous allons montrer que pour tout réel  $\lambda > 0$  et tout  $n \geq 0$ , on a

$$\mathbb{P}\left(\max_{0 \leq i \leq n} |S_i| > \lambda\sqrt{n}\right) \leq 2\mathbb{P}\left(|S_n| > (\lambda - \sqrt{2})\sqrt{n}\right). \quad (2.3)$$

Fixons  $\lambda > 0$  et  $n \geq 0$ . Notons  $T = \inf\{i \geq 0 : |S_i| > \lambda\sqrt{n}\}$ . Le membre de gauche de l'inégalité à démontrer n'est autre que  $\mathbb{P}(T \leq n)$ . Calculons le :

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

Sur l'événement  $\{T = j, |S_n| \leq (\lambda - \sqrt{2})\sqrt{n}\}$ , on a  $|S_j| \geq \lambda\sqrt{n}$ , si bien que  $|S_j - S_n| \geq \sqrt{2n}$ . Ainsi, en observant que l'événement  $\{|S_j - S_n| \geq \sqrt{2n}\}$  est indépendant de  $\{T = j\}$ , et de probabilité bornée, par l'inégalité de Bienaymé–Chebyshev, par  $\frac{1}{2}$ , on trouve

$$\begin{aligned} \mathbb{P}(T \leq n) &\leq \mathbb{P}(|S_n| > (\lambda - \sqrt{2})\sqrt{n}) + \sum_{j=0}^n \mathbb{P}(T = j) \mathbb{P}(|S_n| \leq (\lambda - \sqrt{2})\sqrt{n}) \\ &\leq \mathbb{P}(|S_n| > (\lambda - \sqrt{2})\sqrt{n}) + \frac{1}{2} \mathbb{P}(T \leq n), \end{aligned}$$

ce qui montre l'inégalité voulue.

Nous allons maintenant montrer que les hypothèses de la proposition 2.4.3 sont satisfaites. Puisque  $M_{n,0} = 0$ , la première hypothèse est bien satisfaite. Pour la seconde, commençons par observer que pour tous  $0 \leq t \leq s$  et tout  $n \geq 1$ , on a

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

De plus, le fait que pour tous réels  $a$  et  $b$  on ait  $\lfloor a + b \rfloor \in \{\lfloor a \rfloor + \lfloor b \rfloor, \lfloor a \rfloor + \lfloor b \rfloor + 1\}$  entraîne

$$\lfloor ns \rfloor \leq \lfloor nt \rfloor + \lfloor n(s - t) \rfloor + 1,$$

si bien que  $\lfloor ns \rfloor + 1 - \lfloor nt \rfloor \leq \lfloor n(s - t) \rfloor + 2$ , et finalement, que pour tous  $t \geq 0$ ,  $\delta > 0$  et  $n \geq 1$ , on a

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

Donnons-nous maintenant  $\varepsilon > 0$  et  $\eta > 0$ . Soit  $k \geq 1$  un entier et posons  $\delta = \frac{1}{k}$ . Pour tout  $t \in [0, 1 - \delta]$ , et pour tout  $n \geq 1$ , on a, d'après ce qui précède,

$$\begin{aligned} \mathbb{P}\left(\sup_{r \in [0, \delta]} |M_{n, t+r} - M_{n, t}| > \eta\right) &\leq \mathbb{P}\left(\max_{0 \leq i \leq \lfloor n\delta \rfloor + 2} |S_{\lfloor nt \rfloor + i} - S_{\lfloor nt \rfloor}| > \frac{\eta\sqrt{n}}{2}\right) \\ &= \mathbb{P}\left(\max_{0 \leq i \leq \lfloor n\delta \rfloor + 2} |S_i| > \frac{\eta\sqrt{n}}{2}\right). \end{aligned}$$

Pour utiliser l'inégalité maximale (2.3), nous écrivons cette dernière probabilité sous la forme

$$\mathbb{P}\left(\max_{0 \leq i \leq [n\delta] + 2} |S_i| > \frac{\eta\sqrt{n}}{2\sqrt{[n\delta] + 2}} \sqrt{[n\delta] + 2}\right)$$

et trouvons qu'elle est inférieure ou égale à

$$2\mathbb{P}\left(\frac{|S_{[n\delta] + 2}|}{\sqrt{[n\delta] + 2}} > \frac{\eta\sqrt{n}}{2\sqrt{[n\delta] + 2}} - \sqrt{2}\right).$$

Lorsque  $n$  tend vers l'infini, cette dernière probabilité converge, d'après le théorème central limite, la dernière assertion du théorème de Portmanteau et le fait que la loi gaussienne standard n'a pas d'atomes, vers

$$2\mathbb{P}\left(|N| \geq \frac{\eta}{2\sqrt{\delta}} - \sqrt{2}\right),$$

où  $N$  suit la loi  $\mathcal{N}(0, 1)$ . Une inégalité classique (voir par exemple (E.2)) nous donne alors

$$2\mathbb{P}\left(|N| \geq \frac{\eta}{2\sqrt{\delta}} - \sqrt{2}\right) \leq 2\left(\frac{\eta}{2\sqrt{\delta}} - \sqrt{2}\right)^{-1} e^{-\frac{1}{2}\left(\frac{\eta}{2\sqrt{\delta}} - \sqrt{2}\right)^2}.$$

Puisque la majoration que nous avons faite ne dépendait pas de  $t$ , nous trouvons que pour tout entier  $k \geq 1$ ,

$$\overline{\lim}_{n \rightarrow \infty} k \max_{0 \leq i \leq k-1} \mathbb{P}\left(\sup\{|M_{n, \frac{i}{k}} - M_{n, t}| : t \in [\frac{i}{k}, \frac{i+1}{k}]\} \geq \eta\right) \leq \frac{2k}{\frac{\eta}{2}\sqrt{k} - \sqrt{2}} e^{-\frac{1}{2}\left(\frac{\eta}{2}\sqrt{k} - \sqrt{2}\right)^2}.$$

Cette quantité tend vers 0 lorsque  $k$  tend vers l'infini, ce qui achève de montrer que la deuxième condition de la proposition 2.4.3 est satisfaite, et, grâce au théorème 2.3.5, conclut la démonstration.  $\square$

## **Deuxième partie**

# **Introduction aux grandes déviations**

Cette deuxième partie du cours est consacrée à l'étude d'une classe de résultats appelés principes de grandes déviations, qui expriment d'une manière mêlant subtilement topologie et mesure la concentration exponentiellement rapide d'une suite de mesures boréliennes de probabilités sur un espace topologique. L'aspect topologique sera cependant moins central que dans la première partie du cours, et l'accent sera plus mis sur des phénomènes de nature probabiliste.

Notre objectif sera de comprendre ce qu'est un principe de grandes déviations (voir la définition 3.3.2), d'en étudier quelques propriétés générales, et deux exemples génériques fondamentaux que sont le théorème de Sanov (théorème 4.3.1) pour les lois empiriques d'échantillons i.i.d. et le théorème de Cramér (théorème 5.4) pour leurs moyennes empiriques.

Presque toute cette partie est directement issue du livre de Dembo et Zeitouni [**DemboZeitouni**], qu'on pourra consulter avec grand profit, que ce soit pour plus de détails ou pour des développements plus avancés. Notons que le théorème de Gärtner–Ellis, dont nous ne parlerons pas du tout, fait partie des résultats de base de la théorie des grandes déviations et aurait sans doute constitué, avec le « lemme de Varadhan » que nous évoquons sans le nommer dans l'introduction (voir (3.6)), la prochaine étape d'un traitement plus approfondi du sujet.

D'autres textes classiques de théorie des grandes déviations sont les livres de Varadhan [**Varadhan**] (un des pères de la théorie), de Stroock [**Stroock**], et de Ellis [**Ellis**].

# 3

## Principes de grandes déviations

---

### 3.1 Des événements rares

Sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ , soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes et identiquement distribuées, admettant un moment d'ordre 1. Notons  $m = \mathbb{E}[X_1]$  leur moyenne. La loi faible des grands nombres assure que pour tout  $x > m$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_1 + \dots + X_n > nx) = 0. \quad (3.1)$$

Si l'on suppose que nos variables aléatoires admettent un moment d'ordre 2, si nous notons  $\sigma^2 = \text{Var}(X_1)$ , et si nous supposons  $\sigma > 0$ , alors le théorème central limite permet de préciser le résultat précédent : pour tout réel  $a$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_1 + \dots + X_n > nm + \sqrt{n}\sigma a) = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{t^2}{2}} dt. \quad (3.2)$$

Nous voudrions préciser (3.1) d'une autre manière, en déterminant la vitesse de la convergence que cette équation énonce, c'est-à-dire en déterminant l'ordre de grandeur, lorsque  $n$  tend vers l'infini, de la probabilité qui figure au membre de gauche. Le réel  $a$  n'étant pas autorisé à dépendre de  $n$  dans l'énoncé du théorème central limite, nous ne pouvons pas poser  $a = \frac{x}{\sigma}\sqrt{n}$  dans (3.2). Faisons-le tout de même : nous trouvons alors

$$\mathbb{P}(X_1 + \dots + X_n > nx) \stackrel{??}{\underset{n \rightarrow \infty}{\approx}} \frac{1}{\sqrt{2\pi}} \int_{\frac{x}{\sigma}\sqrt{n}}^\infty e^{-\frac{t^2}{2}} dt \sim \frac{\sigma}{\sqrt{2\pi}nx} e^{-n\frac{x^2}{2\sigma^2}}.$$

Le dernier équivalent est juste, mais inutilement précis, puisque le premier est, en général, faux. Toutefois, tout n'est pas à jeter dans l'estimation heuristique que nous venons de faire : l'idée que la probabilité qui nous intéresse décroît avec  $n$  comme  $e^{-nc}$ , où  $c$  est une constante, est, sous certaines hypothèses, correcte. Lorsqu'elle l'est, c'est sur la constante  $c$  que nous allons concentrer nos efforts, et la quantité dont nous étudierons le comportement asymptotique est

$$\frac{1}{n} \log \mathbb{P}(X_1 + \dots + X_n > nx).$$

Nous verrons qu'il n'est pas toujours possible de déterminer la limite de cette quantité, mais c'est néanmoins dans cette échelle que nous allons travailler.

Supposons maintenant que nos variables aléatoires admettent des moments exponentiels de tous ordres, c'est-à-dire que pour tout réel  $\alpha$ , on a  $\mathbb{E}[e^{\alpha X_1}] < \infty$ . C'est par exemple le cas pour des variables aléatoires bornées. Une inégalité classique, de Chernov, nous donne une indication un peu plus solide du fait que l'heuristique que nous avons déduite d'une application illégitime du théorème central limite est correcte. Rappelons cette inégalité et la manière dont elle se démontre.

**Proposition 3.1.1.** *Sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ , soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes et identiquement distribuées, admettant des moments exponentiels de tous ordres.*

*Soit  $x > \mathbb{E}[X_1]$  un réel. Il existe un réel  $c > 0$  tel que pour tout  $n \geq 1$ , on ait*

$$\mathbb{P}(X_1 + \dots + X_n \geq nx) \leq e^{-nc}.$$

*Démonstration.* Soit  $n \geq 1$  un entier. Soit  $t > 0$  un réel. L'inégalité de Markov<sup>1</sup> nous donne

$$\mathbb{P}(X_1 + \dots + X_n \geq nx) = \mathbb{P}(e^{t(X_1 + \dots + X_n)} > e^{ntx}) \leq e^{-ntx} \mathbb{E}[e^{t(X_1 + \dots + X_n)}] = e^{-ntx} \mathbb{E}[e^{tX_1}]^n.$$

Notons, pour tout réel  $t$ ,

$$\Lambda(t) = \log \mathbb{E}[e^{tX_1}].$$

Nous avons donc démontré que pour tout réel  $t > 0$ ,

$$\mathbb{P}(X_1 + \dots + X_n \geq nx) \leq e^{-n(tx - \Lambda(t))}.$$

Il nous suffit de démontrer qu'il existe  $t > 0$  tel que  $tx - \Lambda(t) > 0$ . Notons  $c(t) = tx - \Lambda(t)$ . C'est une fonction  $C^1$  sur  $\mathbb{R}$ , qui vérifie  $c(0) = 0$  et  $c'(0) = x - \Lambda'(0) = x - \mathbb{E}[X_1]$ , qui est strictement positif par hypothèse. Il s'ensuit que  $c$  est strictement positive pour  $t > 0$  assez petit, et le résultat est prouvé.  $\square$

Pour tirer le maximum du raisonnement que nous venons de faire, cherchons la valeur de  $t$  qui maximise  $c(t)$ . Définissons donc, pour tout réel  $x$ ,

$$\Lambda^*(x) = \sup\{xt - \Lambda(t) : t > 0\}.$$

Notre raisonnement montre que pour tout  $n \geq 1$ ,

$$\mathbb{P}(X_1 + \dots + X_n \geq nx) \leq e^{-n\Lambda^*(x)}.$$

Cette assertion est une instance d'une partie de ce que nous appellerons bientôt un principe de grandes déviations. Nous l'écrirons plutôt sous la forme suivante, un peu moins précise :

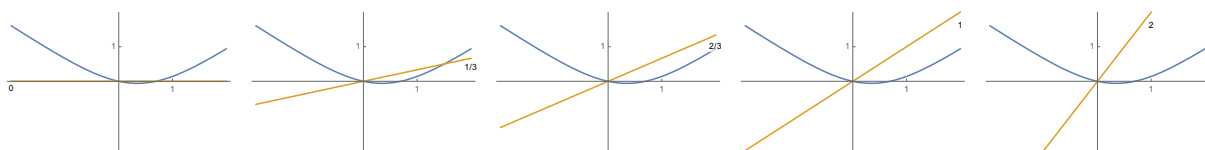
$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(X_1 + \dots + X_n \geq nx) \leq -\Lambda^*(x).$$

1. Il peut paraître très grossier d'utiliser ici l'inégalité de Markov, d'une manière qui revient à dire que pour tout réel  $u$  et tout réel  $t > 0$ , on a  $\mathbb{1}_{\mathbb{R}_+}(u) \leq e^{tu}$ . Toutefois, on observera que nous avons la liberté de choisir  $t$ , et que pour tout réel  $u$ , on a  $\inf\{e^{tu} : t > 0\} = \mathbb{1}_{\mathbb{R}_+}(u)$ .

Faisons le calcul sur un exemple simple. Considérons  $p \in ]0, 1[$  et supposons que la loi commune de nos variables aléatoires soit  $p\delta_1 + (1 - p)\delta_{-1}$ , de moyenne  $2p - 1$ . Alors, pour tout réel  $t$ , on a

$$\Lambda(t) = \log(pe^t + (1 - p)e^{-t}).$$

La figure ci-dessous montre le graphe de  $\Lambda$  pour  $p = \frac{1}{3}$ , avec celui de droites de différentes pentes, indiquées sur chaque graphe :  $x = 0, \frac{1}{3}, \frac{2}{3}, 1, 2$ .



On voit que pour  $x = 0, x = \frac{1}{3}$  ou  $x = \frac{2}{3}$ , le supremum de  $xt - \Lambda(t)$  est atteint pour un unique  $t$ , et on peut se convaincre sur le dessin que ce  $t$  vérifie l'équation  $\Lambda'(t) = x$ , ce qui permet de le calculer. Pour  $x = 1$ , le supremum de  $xt - \Lambda(t)$  en est la limite lorsque  $t$  tend vers l'infini, qui vaut  $-\log p = -\log \frac{1}{3}$ . Et en effet,

$$\mathbb{P}(X_1 + \dots + X_n \geq n) = \mathbb{P}(X_1 + \dots + X_n = n) = \frac{1}{3^n}.$$

Pour  $x = 2$ , on voit que le supremum est infini, ce qui est cohérent avec le fait que la probabilité que nous cherchons à estimer est nulle.

Revenons au cas général. Un de nos objectifs est de démontrer le théorème de Cramér, qui affirme que sous de bonnes hypothèses, on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(X_1 + \dots + X_n \geq nx) = -\Lambda^*(x).$$

Nous voyons sur l'exemple que nous venons d'étudier qu'il ne serait pas vrai, par exemple, que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(X_1 + \dots + X_n > nx) = -\Lambda^*(x).$$

En effet, dans notre exemple, cet énoncé est faux pour  $x = 1$ , puisque le membre de gauche vaut  $-\infty$  et le membre de droite  $\log \frac{1}{3}$ .

Cet exemple montre qu'un cadre général dans lequel nous pourrions écrire des énoncés comme ceux auxquels nous venons de nous intéresser devra prendre en compte la topologie d'une manière sérieuse. C'est un point commun avec l'étude de la convergence faible que nous avons faite dans la première partie de ce cours.

### 3.2 Une concentration exponentielle

En termes vagues, un principe de grandes déviations affirme qu'une suite de mesures sur un certain espace se concentre exponentiellement vite sur un sous-ensemble de cet espace, et décrit, à l'échelle exponentielle, la vitesse à laquelle la masse de tout ensemble disjoint du lieu

de concentration de la suite de mesures tend vers 0. Dans cette section, nous allons examiner ce que peut signifier une telle assertion, et la forme qu'elle pourrait prendre, sans encore en donner la forme exacte.

Dans l'exemple précédent, l'espace était  $\mathbb{R}$ , et la suite de mesures la suite des lois des moyennes empiriques  $\frac{1}{n}(X_1 + \dots + X_n)$ . Cette suite de mesure se concentre sur le singleton  $\{m\}$ , et nous avons discuté du fait qu'il était plausible que pour tout  $x > m$ , on ait

$$\mathbb{P}\left(\frac{1}{n}(X_1 + \dots + X_n) \in [x, +\infty[ \right) \approx e^{-n\Lambda^*(x)}$$

pour une certaine fonction  $\Lambda^*$ .

**Sommes et maximums : la fonction de taux** — Pour fixer les idées dans un cadre un peu plus général, considérons un espace métrique  $(E, d)$  et une suite  $(\mu_n)_{n \geq 0}$  de mesures boréliennes de probabilité sur  $(E, d)$ . Nous nous intéressons à des situations où pour une classe aussi grande que possible de parties  $A$  de  $E$ , la mesure  $\mu_n(A)$  décroît exponentiellement vite avec  $n$ , c'est-à-dire qu'il existe un réel  $F(A) \geq 0$  tel que

$$\mu_n(A) \approx e^{-nF(A)}.$$

Pour comprendre ce que peut être la forme de la fonctionnelle  $F$  qui donne le taux de décroissance de la mesure des parties de  $E$ , examinons ce qui se passe pour des réunions disjointes. Pour préciser un peu les choses, supposons donnés deux boréliens  $A$  et  $B$  disjoints tels que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) = -F(A) \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B) = -F(B).$$

Alors on vérifie<sup>2</sup>, et c'est un fait important de cette histoire, que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A \cup B) = -\min(F(A), F(B)).$$

En subdivisant une partie  $A$  de plus en plus finement, on voit qu'une forme plausible de la fonctionnelle  $F$  serait

$$F(A) = \inf_{x \in A} I(x),$$

où la fonction  $I$  serait en quelque sorte définie par  $I(x) = F(\{x\})$ . Une approche aussi naïve ne fonctionne bien entendu pas en général, mais la forme que nous venons d'induire pour la fonctionnelle  $F$  est tout de même correcte, et un principe de grandes déviations ne sera pas très loin d'affirmer que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) = -\inf_{x \in A} I(x) \tag{3.3}$$

pour une certaine fonction  $I$  qui s'appellera la *fonction de taux*.

**Sommes et maximums : l'absence de mesure de référence** — Pour aller un peu plus loin dans l'exploration heuristique, prenons encore un instant au sérieux l'égalité  $I(x) = F(\{x\})$  et

2. Soient  $(a_n)_{n \geq 1}$  et  $(b_n)_{n \geq 1}$  des suites de réels convergeant respectivement vers  $a$  et  $b$ . Supposons  $a \geq b$ . Alors pour tout  $n \geq 1$ , on a  $e^{na_n} \leq e^{na_n} + e^{nb_n} \leq 2e^{n \max(a_n, b_n)}$ , si bien que  $\lim_{n \rightarrow \infty} \frac{1}{n} \log(e^{na_n} + e^{nb_n}) = a = \max(a, b)$ .



supposons qu'il y ait une mesure de référence  $\nu$  sur notre espace  $(E, d)$ . On aurait donc, pour une bonne partie  $A$ ,

$$\mu_n(A) \simeq \int_A e^{-nI(x)} d\nu(x). \quad (3.4)$$

Il n'y a en général ni une mesure de référence  $\nu$ , ni une expression comme (3.4), mais il est tout de même vrai, par exemple, que si  $\nu$  est une mesure finie sur  $(E, d)$ , et  $I$  une fonction borélienne positive sur  $E$ , alors pour tout borélien  $A$  de  $E$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \int_A e^{-nI(x)} d\nu(x) = -\inf_A \text{ess}(I), \quad (3.5)$$

où l'infimum essentiel de  $I$  sur  $A$  (relativement à  $\nu$ ) est défini par

$$\inf_A \text{ess}(I) = \sup\{a : \nu(A \cap \{I \geq a\}) = 0\}.$$

Une version sans infimum essentiel de cette assertion, dans un cadre plus restreint mais qui tourne autour de la même idée, est la suivante : si  $f : [0, 1] \rightarrow \mathbb{R}_+$  est une fonction continue, alors

$$\lim_{n \rightarrow \infty} \left( \int_0^1 f^n(t) dt \right)^{\frac{1}{n}} = \|f\|_\infty.$$

Un aspect fondamental de cette dernière égalité est que la limite du membre de gauche ne dépend pas du nombre de points où  $f$  atteint son maximum, ou plus généralement de la taille de l'ensemble où  $f$  atteint son maximum. De même, dans (3.5), la limite ne dépend pas du tout de la mesure  $\nu$ , et en particulier pas de la mesure de l'ensemble où  $I$  est égale à, ou voisine de, son minimum.

Le fond commun de ces remarques est le fait suivant, que nous avons déjà évoqué.

**Lemme 3.2.1.** Soient  $(a_n)_{n \geq 1}$  et  $(b_n)_{n \geq 1}$  des suites de réels convergeant respectivement vers des réels  $a$  et  $b$ . Soient  $(\alpha_n)_{n \geq 1}$  et  $(\beta_n)_{n \geq 1}$  des suites de réels telles qu'il existe des réels  $c, C > 0$  tels que pour tout  $n \geq 1$ , on ait  $c \leq \alpha_n, \beta_n \leq C$ . Alors

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(\alpha_n e^{na_n} + \beta_n e^{nb_n}) = \max(a, b).$$

Une version plus générale de ce lemme est la suivante (voir [DemboZeitouni], lemme 1.2.15).

**Lemme 3.2.2.** Soit  $N \geq 1$  un entier. Soient  $a_1, \dots, a_N = \mathbb{R}_+^* \rightarrow \mathbb{R}_+$  des fonctions. On a

$$\overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log \sum_{i=1}^N a_i(\varepsilon) = \max_{i=1}^N \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon \log a_i(\varepsilon).$$

Un principe de grandes déviations va nous permettre de donner corps à une idée comme celle qu'essaye d'exprimer (3.4), sans avoir besoin de disposer sur notre espace d'une mesure de référence : le fait de travailler à l'échelle exponentielle, qui est très grossière, nous en dispense.

**Une version fonctionnelle** — Donnons une expression fonctionnelle (toujours heuristique) de la concentration exponentielle que nous essayons de décrire. Reprenons (3.4), en nous donnant maintenant une fonction test  $\phi$  sur notre espace  $(E, d)$ . Alors nous pouvons nous attendre à ce que

$$\int_A e^{n\phi(x)} d\mu_n \simeq \int_A e^{-n(I(x)-\phi(x))} d\nu.$$

Pour nous débarrasser de la mesure  $\nu$ , que nous avons introduite assez artificiellement, nous pouvons prendre la limite lorsque  $n$  tend vers l'infini de  $\frac{1}{n} \log$  de ces expressions : nous trouvons

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \int_A e^{n\phi(x)} d\mu_n = - \inf_{x \in A} (I(x) - \phi(x)), \quad (3.6)$$

qui est une version fonctionnelle de (3.3).

### 3.3 Principes de grandes déviations

Nous allons maintenant donner des définitions précises. Les ingrédients d'un principe de grandes déviations (PGD) sont les suivants.

1. Un espace topologique  $(E, \mathcal{T})$ .
2. Une suite  $(\mu_n)_{n \geq 1}$  de mesures boréliennes de probabilité sur  $E$ .
3. Une fonction  $I : E \rightarrow [0, +\infty]$  sur  $E$ , qu'on appelle *fonction de taux* du PGD.

La fonction de taux n'est pas une fonction arbitraire. Avant ou après avoir lu la définition qui suit, il pourra être utile de consulter l'annexe C.

**Définition 3.3.1.** Soit  $(E, \mathcal{T})$  un espace topologique. On appelle fonction de taux sur  $E$  une fonction semi-continue inférieurement  $I : E \rightarrow [0, +\infty]$ .

On dit qu'une fonction de taux  $I$  est une bonne fonction de taux si pour tout réel  $a$ , l'ensemble  $\{I \leq a\}$  est non seulement fermé, mais compact.

Nous pouvons maintenant donner la définition principale de ce chapitre.

**Définition 3.3.2** (Principe de grandes déviations). Soit  $(E, \mathcal{T})$  un espace topologique. Soit  $(\mu_n)_{n \geq 1}$  une suite de mesures boréliennes de probabilité sur  $E$ . Soit  $I$  une fonction de taux sur  $E$ . On dit que la suite  $(\mu_n)_{n \geq 1}$  satisfait un principe de grandes déviations de fonction de taux  $I$  si pour tout borélien  $A$  de  $E$ , on a

$$- \inf_{x \in \overset{\circ}{A}} I(x) \leq \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \leq \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \leq - \inf_{x \in \bar{A}} I(x). \quad (3.7)$$

Cette définition est très générale, mais dans l'immense majorité des cas, l'espace topologique sera un espace métrique, le plus souvent polonais. Ajoutons qu'il est parfois nécessaire de remplacer le facteur  $\frac{1}{n}$  par une autre suite tendant vers 0, par exemple  $\frac{1}{n^2}$ , auquel cas on préciserait que le principe de grande déviations est de vitesse  $n^2$ . Nous ne rencontrerons pas d'exemples de cette situation dans ce cours, et tous nos principes de grandes déviations seront de vitesse  $n$ .

La définition 3.3.2 n'est pas tout à fait simple, ce qui justifie peut-être a posteriori les efforts que nous avons faits à la section précédente pour nous y préparer. Elle appelle cependant encore quelques remarques.

1. Il est convenu que  $\log 0 = -\infty$  et  $\inf \emptyset = +\infty$ .
2. Il est utile de prendre quelques secondes pour réfléchir une bonne fois pour toutes<sup>3</sup> au fait que la fonction d'ensemble  $A \mapsto -\inf\{I(x) : x \in A\}$  est croissante.
3. Un PGD est la conjonction de deux inégalités, qu'on appelle naturellement borne inférieure et borne supérieure, et dont les démonstrations sont en général assez différentes, nous y reviendrons. La topologie joue dans chacune de ces bornes un rôle essentiel. On peut par exemple voir qu'il ne serait pas très raisonnable d'espérer remplacer, dans la borne inférieure,  $\mathring{A}$  par  $A$ , en pensant au cas où les mesures  $\mu_n$  sont diffuses<sup>4</sup> et où  $A$  est un singleton.
4. En appliquant la borne supérieure à  $A = E$ , on trouve que  $\inf_{x \in E} I(x) = 0$ . Si la fonction de taux est bonne, elle atteint, comme fonction s.c.i., son minimum sur le compact  $\{I \leq 1\}$ , et l'ensemble  $\{I = 0\}$  est donc un compact non vide de  $E$ .
5. Considérons une suite  $(x_n)_{n \geq 1}$  de réels strictement positifs qui tend vers 0, aussi vite que l'on veut. Posons, pour tout  $n \geq 1$ ,  $\mu_n = \delta_{x_n}$ . Si les principes de grandes déviations sont censés permettre d'exprimer des phénomènes de concentration, il serait souhaitable qu'une telle suite en satisfasse un. La masse donnée par la mesure  $\mu_n$  à un intervalle ouvert d'extrémités strictement positives est nulle pour  $n$  assez grand. La borne inférieure du PGD impose donc que la fonction de taux soit identiquement infinie sur  $\mathbb{R}_+^*$ . Il en est de même sur  $\mathbb{R}^*$ . En utilisant la remarque précédente, on voit que la fonction de taux doit être nulle en 0. En prenant  $A = \mathbb{R}_+^*$  qui satisfait  $\mu_n(A) = 1$  pour tout  $n \geq 1$ , et qui est une partie sur laquelle la fonction de taux est identiquement infinie, on voit l'importance de prendre l'adhérence dans la borne supérieure.

**Proposition 3.3.3.** *Une suite  $(\mu_n)_{n \geq 1}$  de mesures boréliennes de probabilité sur un espace topologique  $E$  satisfait un PGD de fonction de taux  $I$  si et seulement si les deux assertions suivantes sont vraies.*

1. (Borne supérieure) *Pour tout réel  $\alpha < \infty$  et tout borélien  $A$  tel que  $\overline{A} \subset \{I > \alpha\}$ , on a*

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \leq -\alpha. \quad (3.8)$$

2. (Borne inférieure) *Pour tout  $x \in E$  tel que  $I(x) < \infty$  et tout borélien  $A$  tel que  $x \in \mathring{A}$ , on a*

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \geq -I(x). \quad (3.9)$$

*Démonstration.* Le fait qu'une suite  $(\mu_n)_{n \geq 1}$  qui satisfait un PGD de fonction de taux  $I$  et de vitesse  $(a_n)_{n \geq 0}$  vérifie les deux propriétés ci-dessus est immédiat.

Considérons maintenant une suite  $(\mu_n)_{n \geq 1}$  qui vérifie les deux propriétés ci-dessus, et donnons-nous un borélien  $A$ .

D'une part, pour tout réel  $\alpha$  tel que  $\inf_{x \in \overline{A}} I(x) > \alpha$ , on a l'inclusion  $\overline{A} \subset \{I > \alpha\}$ , donc  $\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \leq -\alpha$ . Ceci entraîne que  $\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \leq -\inf_{x \in \overline{A}} I(x)$ .

3. Est-ce vraiment possible en mathématiques? Serait-ce vraiment souhaitable?

4. Une mesure diffuse est une mesure qui donne une masse nulle aux singletons.

D'autre part, la borne (3.9) est aussi vraie pour tout  $x$  intérieur à  $A$  tel que  $I(x) = \infty$ . Ainsi,  $\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \geq -\inf_{x \in \overset{\circ}{A}} I(x)$ .  $\square$

Il serait erroné, malgré ce que pouvait nous suggérer la remarque 5 ci-dessus, de penser que le fait de satisfaire un PGD est une version très forte de la convergence faible.

**Exemple 3.3.4.** Sur  $\mathbb{R}$ , la suite de mesures  $(\mu_n)_{n \geq 1}$  définie par  $\mu_n = \delta_{(-1)^n}$  satisfait un PGD de fonction de taux  $I$  donnée par  $I(-1) = I(1) = 0$  et  $I(x) = +\infty$  pour tout  $x \in \mathbb{R} \setminus \{-1, 1\}$ .

Une question que pose la définition d'un PGD et que nous n'avons pas encore soulevée est de savoir pourquoi on y demande que  $I$  soit s.c.i. Un élément de réponse est donné par la proposition suivante.

**Proposition 3.3.5.** Sur un espace métrique  $(E, d)$ , soit  $(\mu_n)_{n \geq 1}$  une suite de mesures boréliennes de probabilité qui satisfait un PGD et de fonction de taux  $I$ . Alors pour tout  $x \in E$ , on a

$$I(x) = -\lim_{r \rightarrow 0} \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B(x, r)).$$

En particulier, une suite de mesures de probabilités ne peut satisfaire deux PGD de fonctions de taux différentes.

*Démonstration.* Soit  $x$  un élément de  $E$ . Pour tout  $r > 0$ , on a  $I(x) \geq \inf\{I(y) : y \in B(x, r)\}$ . Lorsque  $r$  tend vers 0, cet infimum croît, et  $I(x) \geq \lim_{r \rightarrow 0} \inf\{I(y) : y \in B(x, r)\}$ . Supposons que cette inégalité soit stricte. Alors il existerait un réel  $\alpha$  tel que

$$\forall r > 0, \quad I(x) > \alpha > \inf\{I(y) : y \in B(x, r)\}.$$

Or la fonction  $I$  est s.c.i., donc l'ensemble  $\{I > \alpha\}$  est ouvert. Puisqu'il contient  $x$ , il contient une boule autour de  $x$ , ce qui contredit l'assertion que nous venons de faire. On a donc

$$I(x) = \lim_{r \rightarrow 0} \inf\{I(y) : y \in B(x, r)\}.$$

Il est aussi vrai que pour tout  $r > 0$ , on a  $I(x) \geq \inf\{I(y) : y \in \overline{B}(x, r)\}$ . S'il existait un réel  $\alpha$  tel que pour tout  $r > 0$  on ait  $I(x) > \alpha > \inf\{I(y) : y \in \overline{B}(x, r)\}$ , il existerait un réel  $s > 0$  tel que  $\overline{B}(x, s/2) \subseteq B(x, s) \subseteq \{I > \alpha\}$  et on aboutirait à une contradiction. Finalement,

$$I(x) = \lim_{r \rightarrow 0} \inf\{I(y) : y \in B(x, r)\} = \lim_{r \rightarrow 0} \inf\{I(y) : y \in \overline{B}(x, r)\}.$$

On a donc, pour tout  $x \in E$ ,

$$\begin{aligned} -\lim_{r \rightarrow 0} \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B(x, r)) &\leq I(x) \leq -\lim_{r \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(\overline{B}(x, r)) \\ &\leq -\lim_{r \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B(x, r)), \end{aligned}$$

donc

$$I(x) = -\lim_{r \rightarrow 0} \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B(x, r)) = -\lim_{r \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B(x, r)).$$

Cette expression montre que la fonction de taux  $I$  est complètement déterminée par la suite de mesures  $(\mu_n)_{n \geq 1}$ .  $\square$

Nous n'étudierons pas en détail les propriétés générales des principes de grandes déviations : nous nous intéresserons plutôt à deux situations particulières, et assez typiques. Toutefois, mentionnons tout de même, parce que c'est à la fois important et assez simple à énoncer et à démontrer, la manière dont un PGD se transporte par une application continue.

**Proposition 3.3.6** (Principe de contraction). *Soient  $(E, d)$  et  $(F, \delta)$  deux espaces métriques. Soit  $(\mu_n)_{n \geq 1}$  une suite de mesures boréliennes de probabilité sur  $E$  qui satisfait un principe de grandes déviations de bonne fonction de taux  $I$ .*

*Soit  $f : E \rightarrow F$  une application continue. Pour tout  $y \in F$ , posons*

$$J(y) = \inf\{I(x) : f(x) = y\}.$$

*Alors  $J$  est une bonne fonction de taux sur  $F$  et la suite de mesures  $(\mu_n \circ f^{-1})_{n \geq 1}$  sur  $F$  satisfait un PGD de fonction de taux  $J$ .*

*Démonstration.* Commençons par vérifier que  $J$  est une fonction de taux sur  $F$ . C'est par construction une fonction à valeurs dans  $[0, +\infty]$ . Donnons-nous un réel positif  $\alpha$ . Par définition de  $J$ , on a, pour tout  $y \in F$ ,

$$J(y) \leq \alpha \iff \forall \varepsilon > 0, \exists x \in E, f(x) = y \text{ et } I(x) \leq \alpha + \varepsilon.$$

Ainsi,

$$\{J \leq \alpha\} = \bigcap_{n \geq 1} f(\{I \leq \alpha + \frac{1}{n}\}).$$

Puisque  $I$  est une bonne fonction de taux, l'ensemble  $\{I \leq \alpha + \frac{1}{n}\}$  est compact pour tout  $n \geq 1$ , et  $\{J \leq \alpha\}$  est fermé, et même compact. Ainsi,  $J$  est une bonne fonction de taux sur  $F$ .<sup>5</sup>

Utilisons maintenant la proposition 3.3.3 pour montrer que la suite  $(\mu_n \circ f^{-1})_{n \geq 1}$  satisfait un PGD de fonction de taux  $J$ .

1. Donnons-nous un réel  $\alpha$  et un borélien  $A$  de  $F$  tel que  $\overline{A} \subset \{J > \alpha\}$ . Alors, par définition de  $J$ , on a  $f^{-1}(\overline{A}) \subset \{I > \alpha\}$ . Or puisque  $f$  est continue,  $f^{-1}(\overline{A})$  est un fermé qui contient  $f^{-1}(A)$ , donc son adhérence, donc on a  $\overline{f^{-1}(A)} \subset \{I > \alpha\}$ . On a donc, grâce à (3.8) appliquée à la suite  $(\mu_n)_{n \geq 1}$  et au borélien  $f^{-1}(A)$ ,

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log(\mu_n \circ f^{-1})(A) = \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(f^{-1}(A)) \leq -\alpha.$$

2. Donnons-nous maintenant  $y \in F$  tel que  $J(y) < \infty$ , et un borélien  $A$  dont l'intérieur contient  $y$ . Alors il existe  $x \in E$  tel que  $f(x) = y$  et  $I(x) < \infty$ . De plus, par continuité de  $f$ , la partie  $f^{-1}(A)$  de  $E$  est un voisinage de  $x$ . Ainsi,

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log(\mu_n \circ f^{-1})(A) = \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(f^{-1}(A)) \geq -I(x).$$

Puisque ceci est vrai pour tout  $x$  tel que  $f(x) = y$ , on a

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log(\mu_n \circ f^{-1})(A) \geq \sup\{-I(x) : x \in E, f(x) = y\} = -J(y).$$

Ceci démontre la borne inférieure, et achève la démonstration.  $\square$

5. Notons que l'hypothèse que  $I$  est une bonne fonction de taux nous a été nécessaire, ou tout du moins, cruciallement utile, pour montrer que  $J$  est une fonction de taux.

# 4

## Grandes déviations pour des lois empiriques

La but de ce chapitre est d'étudier un théorème de Sanov qui quantifie la probabilité qu'un grand échantillon i.i.d. d'une loi donnée ait une distribution empirique significativement différente de sa loi. Ceci sera notre premier exemple d'un principe de grandes déviations, et l'occasion de voir apparaître une quantité très importante : l'entropie relative de deux mesures de probabilités.

### 4.1 Entropie relative de deux mesures

Soit  $\Sigma = \{a_1, \dots, a_N\}$  un ensemble fini fixé une fois pour toutes. Une mesure de probabilité sur  $\Sigma$  est un vecteur de  $\mathbb{R}^N$  à composantes positives et de somme 1. Ainsi,  $\mathcal{M}(\Sigma)$  s'identifie à une partie de  $\mathbb{R}^N$ , convexe et compacte, et la topologie correspondante est la topologie faible correspondant à la topologie discrète sur  $\Sigma$ . Pour tout  $\mu \in \Sigma$ , nous noterons au besoin  $\Sigma_\mu$  le support de  $\mu$ , c'est-à-dire la réunion des singletons de  $\Sigma$  de masse strictement positive.

Choisissons une mesure de probabilité  $\mu$  sur  $\Sigma$ . Quitte à restreindre  $\Sigma$ , nous supposons que  $\Sigma_\mu = \Sigma$ . Donnons-nous, sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ , une suite  $(Y_n)_{n \geq 1}$  de variables aléatoires indépendantes et identiquement distribuées, de loi  $\mu$ . Pour tout  $n \geq 1$ , notons

$$L_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}.$$

Il s'agit d'une mesure de probabilités aléatoires sur  $\Sigma$ . Autrement dit,  $L_n$  est une variable aléatoire à valeurs dans  $\mathcal{M}(\Sigma)$ . Pour tout  $a \in \Sigma$ , la loi forte des grands nombres appliquée à la suite  $(\mathbb{1}_{\{Y_n=a\}})_{n \geq 1}$  nous donne

$$L_n(a) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Y_i=a\}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{P}(Y_1 = a) = \mu(a).$$

Ainsi, sur l'événement  $\mathbb{P}$ -presque sûr où cette convergence a lieu pour tout  $a \in \Sigma$ , on a, pour toute fonction  $f : \Sigma \rightarrow \mathbb{R}$ ,

$$\int_{\Sigma} f \, dL_n = \sum_{j=1}^N f(a_j) L_n(a_j) \xrightarrow[n \rightarrow \infty]{} \sum_{j=1}^N f(a_j) \mu(a_j) = \int_{\Sigma} f \, d\mu.$$

Autrement dit, la suite  $(L_n)_{n \geq 1}$  converge presque sûrement faiblement vers  $\mu$ .

Nous voulons estimer la probabilité que pour un certain  $n$ , la mesure  $L_n$  soit éloignée de  $\mu$ . Pour cela, donnons-nous une mesure  $\nu$  et essayons de calculer la probabilité que  $L_n$  soit égale à  $\nu$ . Il s'agit d'un problème combinatoire. Tout d'abord, les seules valeurs que peut prendre la mesure  $L_n$  sont les mesures qui donnent à chaque point de  $\Sigma$  une masse multiple de  $\frac{1}{n}$ . Notons  $\mathcal{L}_n$  l'ensemble de ces mesures :

$$\mathcal{L}_n = \left\{ \pi \in \mathcal{M}(\Sigma) : \forall a \in \Sigma, \pi(a) \in \frac{1}{n}\mathbb{N} \right\}.$$

Choisissons  $\nu \in \mathcal{L}_n$  et soit  $(k_1, \dots, k_N)$  le vecteur d'entiers tel que pour tout  $j \in \{1, \dots, N\}$ , on ait  $\nu(a_j) = \frac{k_j}{n}$ . Notons  $T_n(\nu)$  l'ensemble des suites de  $n$  éléments de  $\Sigma$  dont la loi empirique est  $\nu$  :

$$T_n(\nu) = \left\{ (y_1, \dots, y_n) \in \Sigma^n : \frac{1}{n} \sum_{i=1}^n \delta_{y_i} = \nu \right\}.$$

On a donc

$$\mathbb{P}(L_n = \nu) = \sum_{(y_1, \dots, y_n) \in T_n(\nu)} \mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n)$$

et on observe que la probabilité ne dépend pas de la suite  $(y_1, \dots, y_n)$  dans  $T_n(\nu)$ , si bien que

$$\mathbb{P}(L_n = \nu) = |T_n(\nu)| \prod_{j=1}^N \mu(a_j)^{k_j}.$$

Le cardinal de l'ensemble  $T_n(\nu)$  peut se calculer comme suit : cet ensemble contient la suite

$$\underbrace{(a_1, \dots, a_1)}_{k_1 \text{ fois}}, \dots, \underbrace{(a_N, \dots, a_N)}_{k_N \text{ fois}}$$

et tous les autres éléments de cet ensemble sont les suites qu'on obtient en permutant les termes de celle-ci. Il y a donc une action transitive du groupe symétrique  $S_n$  sur  $T_n(\nu)$ , et le stabilisateur de l'élément que nous avons écrit est de cardinal  $k_1! \dots k_N!$ . Ainsi,

$$|T_n(\nu)| = \frac{n!}{k_1! \dots k_N!} \quad (4.1)$$

et

$$\mathbb{P}(L_n = \nu) = \frac{n!}{k_1! \dots k_N!} \prod_{j=1}^N \mu(a_j)^{k_j}.$$

Réécrivons cette expression sous la forme

$$\mathbb{P}(L_n = \nu) = \frac{n!}{k_1! \dots k_N!} \exp \left( n \sum_{j=1}^N \nu(a_j) \log \mu(a_j) \right). \quad (4.2)$$

Cette expression est exacte et nous nous en servons à nouveau. Pour l'instant, essayons de déterminer, à l'échelle exponentielle, la taille du quotient de factorielles. Pour cela, nous utilisons

la formule de Stirling, en ne gardant que les termes d'ordre au moins  $e^n$  :

$$\begin{aligned} \frac{n!}{k_1! \dots k_N!} &\approx \exp \left( n \log n - n - \sum_{j=1}^N (k_j \log k_j - k_j) \right) \\ &= \exp \left( - \sum_{j=1}^N k_j \log \frac{k_j}{n} \right) \\ &= \exp \left( - n \sum_{j=1}^N v(a_j) \log v(a_j) \right). \end{aligned} \quad (4.3)$$

En mettant tout ensemble, nous obtenons l'expression approchée

$$\mathbb{P}(L_n = \nu) \approx \exp \left( - n \sum_{j=1}^N v(a_j) \log \frac{v(a_j)}{\mu(a_j)} \right).$$

Ce calcul préliminaire, bien qu'approché, est fondamental, et nous motive pour la définition suivante. Il est entendu dans ce qui suit que  $0 \log 0 = 0$ ,  $0 \log(0/0) = 0$  et pour tout  $p > 0$ ,  $p \log(p/0) = +\infty$ .

**Définition 4.1.1** (Entropie relative). Soient  $\nu$  et  $\mu$  deux mesures de probabilité sur  $\Sigma$ .

1. On appelle entropie de la mesure  $\nu$  la quantité

$$H(\nu) = - \sum_{j=1}^N \nu(a_j) \log \nu(a_j).$$

2. On appelle entropie relative de  $\nu$  par rapport à  $\mu$  la quantité

$$H(\nu|\mu) = \sum_{j=1}^N \nu(a_j) \log \frac{\nu(a_j)}{\mu(a_j)}.$$

L'entropie relative de  $\nu$  par rapport à  $\mu$  est infinie si  $\nu$  n'est pas absolument continue par rapport à  $\mu$ , et vaut sinon

$$H(\nu|\mu) = \int_{\Sigma} \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu.$$

Il est utile de se rappeler que la fonction  $x \mapsto x \log x$  est nulle en 0 et en 1, strictement négative sur l'intérieur de l'intervalle  $[0, 1]$ , et strictement convexe. En particulier, grâce à l'inégalité de Jensen, on vérifie que  $H(\nu|\mu) \geq 0$ . L'égalité a lieu seulement si la densité  $\frac{d\nu}{d\mu}$  est constante, c'est-à-dire si  $\nu = \mu$ .

Quant à l'entropie de  $\nu$ , elle est manifestement positive, et nulle seulement si  $\nu$  est une masse de Dirac. Le commentaire que nous venons de faire sur l'entropie relative de deux mesures montre qu'on peut la définir dans une grande généralité. Par contre, l'entropie telle que nous l'avons définie n'a de sens que sur un espace fini. Toutefois, en notant  $\sigma$  la mesure uniforme sur  $\Sigma$ , on a

$$H(\nu|\sigma) = \log N - H(\nu).$$

Cette observation nous montre, en prime, que  $H(\nu) \leq \log N$ , avec égalité uniquement lorsque  $\nu$  est la mesure uniforme.



## 4.2 Résultats précis

Réécrivons l'égalité (4.2) en utilisant (4.1) et la définition que nous venons de donner. Nous allons aussi désormais utiliser la notation  $\mathbb{P}_\mu$  à la place de  $\mathbb{P}$ , pour nous souvenir que sous  $\mathbb{P}_\mu$ , les  $(Y_n)_{n \geq 1}$  sont de loi  $\mu$ . Nous allons aussi utiliser la mesure  $\mathbb{P}_\nu$  pour d'autres mesures  $\nu$  que la mesure  $\mu$ .<sup>1</sup>

**Proposition 4.2.1.** *Soit  $\nu \in \mathcal{L}_n$  et soit  $(y_1, \dots, y_n)$  un élément de  $T_n(\nu)$ . Alors*

$$\mathbb{P}_\mu((Y_1, \dots, Y_n) = (y_1, \dots, y_n)) = e^{-n(H(\nu) + H(\nu|\mu))}.$$

En particulier,

$$\mathbb{P}_\nu((Y_1, \dots, Y_n) = (y_1, \dots, y_n)) = e^{-nH(\nu)}. \quad (4.4)$$

Nous avons obtenu en (4.3) l'ordre de grandeur  $|T_n(\nu)| \approx e^{-nH(\nu)}$ . Nous allons maintenant préciser rigoureusement ce point.

**Lemme 4.2.2.** *Soit  $\nu \in \mathcal{L}_n$ . On a*

$$(n+1)^{-N} e^{nH(\nu)} \leq |T_n(\nu)| \leq e^{nH(\nu)}.$$

*Démonstration.* Commençons par montrer la borne supérieure. En utilisant (4.4), nous trouvons

$$\mathbb{P}_\nu(L_n = \nu) = \mathbb{P}_\nu((Y_1, \dots, Y_n) \in T_n(\nu)) = |T_n(\nu)| e^{-nH(\nu)}.$$

Le fait que cette probabilité soit inférieure à 1 entraîne l'inégalité

$$|T_n(\nu)| \leq e^{nH(\nu)}.$$

Montrons maintenant la borne inférieure. Pour cela, nous allons utiliser le fait suivant, que nous démontrerons dans un instant : la fonction  $\pi \mapsto \mathbb{P}_\nu(L_n = \pi)$ , définie sur  $\mathcal{L}_n$ , atteint son maximum en  $\nu$ . Autrement dit, sous  $\mathbb{P}_\nu$ , la distribution empirique la plus probable de  $(Y_1, \dots, Y_n)$  est  $\nu$ . Écrivons alors

$$1 = \sum_{\pi \in \mathcal{L}_n} \mathbb{P}_\nu(L_n = \pi) \leq |\mathcal{L}_n| \mathbb{P}_\nu(L_n = \nu) = |\mathcal{L}_n| |T_n(\nu)| e^{-nH(\nu)}.$$

Puisqu'un élément de  $\mathcal{L}_n$  attribue à un élément de  $\Sigma$  une masse qui appartient à l'ensemble  $\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$ , on a  $|\mathcal{L}_n| \leq (n+1)^{|\Sigma|}$ . En reportant cette inégalité dans la ligne de calcul précédente, on aboutit à

$$(n+1)^{-N} e^{nH(\nu)} \leq |T_n(\nu)|,$$

ce qui conclut la démonstration.

---

1. Le plus simple, conceptuellement, mais peut-être pas pédagogiquement, aurait été de se donner l'espace canonique  $\Sigma^{\mathbb{N}^*}$  muni de la tribu engendrée par le processus canonique, noté  $(Y_n)_{n \geq 1}$ , et de définir sur cet espace, pour toute mesure  $\nu \in \mathcal{M}(\Sigma)$ , la mesure  $\mathbb{P}_\nu = \nu^{\otimes \mathbb{N}^*}$ .

Reste à vérifier que pour tout  $\pi \in \mathcal{L}_n$ , on a  $\mathbb{P}_v(L_n = \pi) \leq \mathbb{P}_v(L_n = v)$ . Utilisons l'équation qui précédait (4.2) pour calculer le quotient de ces probabilités<sup>2</sup> : on a

$$\frac{\mathbb{P}_v(L_n = \pi)}{\mathbb{P}_v(L_n = v)} = \prod_{j=1}^N \frac{(nv(a_j))! v(a_j)^{n\pi(a_j)}}{(n\pi(a_j))! v(a_j)^{nv(a_j)}} = \prod_{j=1}^N \frac{(nv(a_j))! (nv(a_j))^{n\pi(a_j)}}{(n\pi(a_j))! (nv(a_j))^{nv(a_j)'}}$$

où nous avons glissé un  $n$  au numérateur et au dénominateur en tirant parti du fait que les sommes des puissances en haut et en bas sont égales, à  $n$ . Chaque terme du produit est maintenant de la forme

$$\frac{m!}{p!} m^{p-m} = \frac{m^p}{p!} \bigg/ \frac{m^m}{m!}$$

avec  $m, p \geq 0$ . Or  $m$  étant fixé, il n'est pas difficile de vérifier que la suite  $(\frac{m^p}{p!})_{p \geq 0}$  croît jusqu'à  $p = m$  et décroît ensuite. Le quotient que nous étudions est donc plus petit que 1, et notre affirmation est démontrée.  $\square$

En mettant ensemble les deux résultats précédents, on aboutit à l'estimation suivante.

**Proposition 4.2.3.** *Pour tout  $v \in \mathcal{L}_n$ , on a*

$$(n+1)^{-N} e^{-nH(v|\mu)} \leq \mathbb{P}_\mu(L_n = v) \leq e^{-nH(v|\mu)}.$$

*Démonstration.* En effet, après avoir choisi  $(y_1, \dots, y_n) \in T_n(v)$ , on a

$$\mathbb{P}_\mu(L_n = v) = |T_n(v)| \mathbb{P}_\mu((Y_1, \dots, Y_n) = (y_1, \dots, y_n))$$

que nous estimons et calculons grâce à la proposition 4.2.1 et au lemme 4.2.2.  $\square$

### 4.3 Le théorème de Sanov

Nous allons maintenant démontrer un principe de grandes déviations.

**Théorème 4.3.1** (Théorème de Sanov). *Sous  $\mathbb{P}_\mu$ , la suite des lois des mesures aléatoires  $(L_n)_{n \geq 1}$  satisfait sur  $\mathcal{M}(\Sigma)$  un principe de grandes déviations de bonne fonction de taux  $H(\cdot|\mu)$ .*

*Autrement dit, pour tout borélien  $\Gamma \in \mathcal{M}(\Sigma)$ , on a*

$$-\inf_{v \in \hat{\Gamma}} H(v|\mu) \leq \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\mu(L_n \in \Gamma) \leq \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\mu(L_n \in \Gamma) \leq -\inf_{v \in \Gamma} H(v|\mu).$$

On remarquera que dans la borne supérieure, l'infimum est pris sur  $\Gamma$  et non sur son adhérence : c'est délibéré, et l'énoncé est donc plus précis qu'un principe de grandes déviations tel que nous l'avons défini au chapitre précédent.

Nous allons avoir besoin d'un lemme qui nous dise que l'ensemble  $\mathcal{L}_n$  remplit bien l'espace  $\mathcal{M}(\Sigma)$ . Rappelons que la distance en variation entre deux mesures de probabilité  $\nu$  et  $\pi$  sur  $\Sigma$  est définie par

$$d_V(\nu, \pi) = \sup\{|\nu(A) - \pi(A)| : A \subseteq \Sigma\}$$

2. Nous raisonnons comme si le support de  $\nu$  était  $\Sigma$  tout entier. S'il ne l'est pas, il suffit de restreindre les produits qui suivent aux éléments du support de  $\nu$ .

et qu'elle est égale à

$$d_V(\nu, \pi) = \frac{1}{2} \sum_{a \in \Sigma} |\nu(a) - \pi(a)|,$$

comme on peut s'en assurer au prix d'un moment de réflexion.

Rappelons que  $N$  est le cardinal de l'ensemble  $\Sigma$ .

**Lemme 4.3.2.** *Pour tout  $n \geq 1$  et tout  $\pi \in \mathcal{M}(\Sigma)$ , il existe  $\nu \in \mathcal{L}_n$  tel que*

$$d_V(\pi, \nu) \leq \frac{N}{2n}.$$

*Démonstration.* Pour chaque  $j \in \{1, \dots, N\}$ , notons  $k_j = \lfloor n\pi(a_j) \rfloor$ , si bien qu'on a les inégalités  $\frac{k_j}{n} \leq \pi(a_j) < \frac{k_j+1}{n}$ . On vérifie qu'il existe un unique entier  $J \in \{0, \dots, N\}$  tel que la mesure

$$\nu = \left( \frac{k_1}{n} + \frac{1}{n}, \dots, \frac{k_J}{n} + \frac{1}{n}, \frac{k_{J+1}}{n}, \dots, \frac{k_N}{n} \right)$$

soit une mesure de probabilité. C'est un élément de  $\mathcal{L}_n$ , et

$$d_V(\pi, \nu) = \frac{1}{2} \sum_{j=1}^N |\pi(a_j) - \nu(a_j)| \leq \frac{N}{2n},$$

si bien que  $\nu$  remplit la condition souhaitée. □

Nous pouvons maintenant démontrer le théorème de Sanov.

*Démonstration du théorème 4.3.1.* Soit  $\Gamma$  un borélien de  $\mathcal{M}(\Sigma)$ . D'après la proposition 4.2.3, nous avons tout d'abord

$$\begin{aligned} \mathbb{P}_\mu(L_n \in \Gamma) &= \sum_{\nu \in \Gamma \cap \mathcal{L}_n} \mathbb{P}_\mu(L_n = \nu) \\ &\leq \sum_{\nu \in \Gamma \cap \mathcal{L}_n} e^{-nH(\nu|\mu)} \\ &\leq |\Gamma \cap \mathcal{L}_n| e^{-n \inf\{H(\nu|\mu) : \nu \in \Gamma \cap \mathcal{L}_n\}} \end{aligned}$$

et puisque nous avons déjà observé que  $|\mathcal{L}_n| \leq (n+1)^N$ , nous en déduisons que

$$\mathbb{P}_\mu(L_n \in \Gamma) \leq (n+1)^N e^{-n \inf\{H(\nu|\mu) : \nu \in \Gamma \cap \mathcal{L}_n\}}.$$

Ensuite, nous avons

$$\begin{aligned} \mathbb{P}_\mu(L_n \in \Gamma) &= \sum_{\nu \in \Gamma \cap \mathcal{L}_n} \mathbb{P}_\mu(L_n = \nu) \\ &\geq \sum_{\nu \in \Gamma \cap \mathcal{L}_n} (n+1)^{-N} e^{-nH(\nu|\mu)} \\ &\geq (n+1)^{-N} e^{-n \inf\{H(\nu|\mu) : \nu \in \Gamma \cap \mathcal{L}_n\}}. \end{aligned}$$

Les facteurs  $(n+1)^N$  et  $(n+1)^{-N}$  vous disparaître, puisque

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(n+1)^N = \lim_{n \rightarrow \infty} \frac{1}{n} \log(n+1)^{-N} = 0.$$

On déduit donc des deux bornes que nous avons données que

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\mu(L_n \in \Gamma) = - \underline{\lim}_{n \rightarrow \infty} \inf\{H(v|\mu) : v \in \Gamma \cap \mathcal{L}_n\}$$

et

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\mu(L_n \in \Gamma) = - \overline{\lim}_{n \rightarrow \infty} \inf\{H(v|\mu) : v \in \Gamma \cap \mathcal{L}_n\}.$$

De la première égalité, et de l'inclusion  $\Gamma \cap \mathcal{L}_n \subseteq \Gamma$ , on déduit

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\mu(L_n \in \Gamma) \leq - \inf\{H(v|\mu) : v \in \Gamma\},$$

c'est-à-dire la borne supérieure du théorème de Sanov.

Pour démontrer la borne inférieure, considérons un point  $v$  dans l'intérieur de  $\Gamma$ .<sup>3</sup> Il existe  $\delta > 0$  tel que la boule, en distance en variation, de centre  $v$  et de rayon  $\delta$ , soit incluse dans  $\Gamma$ . Il existe donc, d'après le lemme 4.3.2, et pour tout  $n > \frac{1}{\delta}$ , un élément  $v_n \in \Gamma \cap \mathcal{L}_n$  tel que  $d_V(v_n, v) \leq \frac{N}{2n}$ . En particulier, la suite  $(v_n)_n$  assez grand converge vers  $v$ , et  $\lim_{n \rightarrow \infty} H(v_n|\mu) = H(v|\mu)$ .<sup>4</sup> Ainsi,

$$- \overline{\lim}_{n \rightarrow \infty} \inf\{H(v|\mu) : v \in \Gamma \cap \mathcal{L}_n\} \geq - \overline{\lim}_{n \rightarrow \infty} H(v_n|\mu) = -H(v|\mu).$$

Ainsi,

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\mu(L_n \in \Gamma) \geq -H(v|\mu).$$

Puisque ceci est vrai pour tout  $v \in \overset{\circ}{\Gamma}$ , on a

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\mu(L_n \in \Gamma) \geq \sup\{-H(v|\mu) : v \in \overset{\circ}{\Gamma}\} = - \inf\{H(v|\mu) : v \in \overset{\circ}{\Gamma}\}$$

et le théorème est démontré. □

3. Il s'agit de l'intérieur de  $\Gamma$  comme partie de l'espace topologique  $\mathcal{M}(\Sigma)$ .

4. Après l'étude très raffinée que nous avons faite dans la première partie de ce cours de la convergence faible des mesures dans des espaces polonais généraux, ne perdons pas de vue que dans le cas présent, sur un ensemble fini, la convergence des mesures n'est rien d'autre que la convergence de la masse de chaque point.

# 5

## Grandes déviations pour des moyennes empiriques

Dans ce court chapitre (qui est court en partie parce qu'il s'appuie sur l'annexe F), nous allons démontrer un résultat prototypique de grandes déviations, qui concerne les moyennes empiriques d'échantillons indépendants et identiquement distribués. Le théorème principal, dû à Harald Cramér, peut être considéré comme un complément à la loi forte des grands nombres.

### 5.1 Transformée de Legendre et fonction de taux

Sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ , soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes et identiquement distribuées. Nous définissons la transformée de Laplace de la loi de nos variables aléatoires comme la fonction  $L$  définie sur  $\mathbb{R}$  (ou plutôt sur le dual  $\mathbb{R}^*$  de  $\mathbb{R}$ , que nous identifions à  $\mathbb{R}$ ) par

$$L(\lambda) = \mathbb{E}[e^{\lambda X_1}] \in ]0, +\infty],$$

et nous en considérons le logarithme

$$\Lambda(\lambda) = \log L(\lambda) \in ]-\infty, +\infty].$$

**Lemme 5.1.** *La fonction  $\Lambda$  est convexe.*

*Démonstration.* Soient  $\lambda, \theta$  des réels et soit  $\alpha \in [0, 1]$ . Alors l'inégalité de Hölder donne

$$\begin{aligned} \Lambda((1-\alpha)\lambda + \alpha\theta) &= \log \mathbb{E}[e^{(1-\alpha)\lambda X} e^{\alpha\theta X}] \\ &\leq \log (\mathbb{E}[e^{\lambda X}]^{(1-\alpha)} \mathbb{E}[e^{\theta X}]^\alpha) \\ &= (1-\alpha)\Lambda(\lambda) + \alpha\Lambda(\theta), \end{aligned}$$

ce qui est l'inégalité cherchée. □

La transformée de Legendre de  $\Lambda$  (voir la définition F.1.2) est la fonction  $\Lambda^*$  définie sur  $\mathbb{R}$  par

$$\Lambda^*(x) = \sup\{\lambda x - \Lambda(\lambda) : \lambda \in \mathbb{R}\}.$$

Il s'agit, d'après l'étude menée dans l'appendice F, d'une fonction semi-continue inférieurement et convexe. Puisque  $\Lambda(0) = 0$ , la fonction  $\Lambda^*$  est positive. Notons que si  $\Lambda$  est infinie en tout point autre que 0, alors  $\Lambda^*$  est identiquement nulle.

Le résultat suivant rassemble certaines propriétés de  $\Lambda^*$  qui nous seront utiles. Nous notons  $m$  l'espérance de  $X_1$  lorsque celle-ci existe, au sens où l'une au moins des deux variables aléatoires  $X_1^+$  et  $X_1^-$  (les parties positive et négative de  $X_1$ ) admet une intégrale finie.

**Proposition 5.2.** *Supposons qu'il existe  $\lambda > 0$  tel que  $\Lambda(\lambda) < \infty$ . Alors  $m < +\infty$  et pour tout  $x \geq m$ , on a*

$$\Lambda^*(x) = \sup\{\lambda x - \Lambda(\lambda) : \lambda \geq 0\}.$$

*De même, s'il existe  $\lambda < 0$  tel que  $\Lambda(\lambda) < \infty$ , alors  $m > -\infty$ , et pour tout  $x \leq m$ , on a*

$$\Lambda^*(x) = \sup\{\lambda x - \Lambda(\lambda) : \lambda \leq 0\}.$$

*Dans tous les cas où  $m$  existe, on a  $\Lambda^*(m) = 0$  et la fonction  $\Lambda^*$  est décroissante sur  $] -\infty, m]$  puis croissante sur  $[m, +\infty[$ .*

*Démonstration.* Supposons qu'il existe  $\lambda > 0$  tel que  $\Lambda(\lambda) < \infty$ . Alors

$$\mathbb{E}[X_1^+] = \mathbb{E}[X_1 \mathbb{1}_{\{X_1 \geq 0\}}] \leq \frac{1}{\lambda} \mathbb{E}[e^{\lambda X_1}] < \infty$$

et  $m$  existe et  $m < \infty$ . L'inégalité de Jensen donne, pour tout  $\lambda \in \mathbb{R}$ ,

$$\Lambda(\lambda) \geq \mathbb{E}[\log e^{\lambda X}] = \lambda m.$$

On en déduit d'une part que  $\Lambda^*(m) \leq 0$ , donc que  $\Lambda^*(m) = 0$ . D'autre part, pour tout  $x \geq m$  et tout  $\lambda < 0$ , on a

$$\lambda x - \Lambda(\lambda) \leq \lambda m - \Lambda(\lambda) \leq 0 = 0x - \Lambda(0),$$

si bien que  $\Lambda^*(x) = \sup\{\lambda x - \Lambda(\lambda) : \lambda \geq 0\}$ .

Le cas où  $\Lambda$  prend une valeur finie sur  $\mathbb{R}_-$  se traite de manière analogue.

Enfin, sur  $] -\infty, m]$ , la fonction  $\Lambda^*$ , sup de fonctions décroissantes, est décroissante; et sur  $[m, +\infty[$ , elle est un sup de fonctions croissantes, donc croissante.  $\square$

Nous aurons aussi besoin du lemme suivant.

**Lemme 5.3.** *La fonction  $\Lambda$  est différentiable sur l'intérieur du domaine où elle est finie. Pour tout  $\theta$  dans ce domaine, on a*

$$\Lambda'(\theta) = \frac{\mathbb{E}[X_1 e^{\theta X_1}]}{\mathbb{E}[e^{\theta X_1}]}$$

*et si  $\Lambda'(\theta) = y$ , alors  $\Lambda^*(y) = \theta y - \Lambda(\theta)$ .*

*Démonstration.* Soit  $\theta$  un point intérieur au domaine (qui est un intervalle) sur lequel  $\Lambda$  est finie. Le fait que  $\Lambda$  soit finie sur un voisinage de  $\theta$  et le théorème de convergence dominée permettent de dériver sous l'espérance et d'obtenir la dérivabilité de  $\Lambda$  et la formule annoncée.

Notons maintenant  $y = \Lambda'(\theta)$ . La fonction  $\lambda \mapsto \lambda y - \Lambda(\lambda)$  est concave et sa dérivée s'annule en  $\theta$ . Elle atteint donc son maximum en  $\theta$ , si bien que  $\Lambda^*(y) = \theta y - \Lambda(\theta)$ .  $\square$

## 5.2 Le théorème de Cramér

Le théorème que nous allons démontrer (sous des hypothèses qui nous faciliteront un peu le travail) est le suivant.

**Théorème 5.4** (Théorème de Cramér). *La suite des lois des moyennes empiriques  $(\frac{X_1 + \dots + X_n}{n})_{n \geq 1}$  satisfait sur  $\mathbb{R}$  un PGD de fonction de taux  $\Lambda^*$ . Autrement dit,*

1. *pour tout fermé  $F$  de  $\mathbb{R}$ , on a*

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{X_1 + \dots + X_n}{n} \in F \right) \leq - \inf \{ \Lambda^*(x) : x \in F \},$$

2. *pour tout ouvert  $G$  de  $\mathbb{R}$ , on a*

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{X_1 + \dots + X_n}{n} \in G \right) \geq - \inf \{ \Lambda^*(x) : x \in G \}.$$

Ce théorème est vrai sans aucune hypothèse d'intégrabilité sur nos variables aléatoires. Dans le cas où elles n'admettent aucun moment exponentiel, c'est-à-dire si  $\Lambda(\lambda)$  est infinie pour tout  $\lambda \neq 0$ , alors  $\Lambda^*$  est identiquement nulle. Dans ce cas, la borne supérieure du théorème ne dit rien, mais la borne inférieure donne encore une information non triviale.

Nous allons démontrer la borne supérieure en toute généralité (nous n'en étions déjà pas très loin dans la discussion de l'inégalité de Chernov au début de cette partie), et pour la borne inférieure, nous allons faire l'hypothèse simplificatrice que nos variables aléatoires sont bornées. Cela nous permettra de nous concentrer sur l'idée principale de cette démonstration, qui est le changement exponentiel de mesure.

*Démonstration.* 1. Si la fonction  $\Lambda^*$  est identiquement nulle, la borne supérieure est vraie. Si elle n'est pas identiquement nulle, alors il existe un  $\lambda$  non nul tel que  $\Lambda(\lambda) < \infty$ , et l'espérance  $m = \mathbb{E}[X_1]$  existe, éventuellement infinie. Nous allons raisonner en supposant qu'elle est finie, mais nous indiquerons comment la preuve doit être adaptée dans le cas où cette espérance est infinie.

Dans la démonstration de la proposition 3.1.1, nous avons montré que pour tout  $x$  réel et tout  $\lambda > 0$ , nous avons  $\mathbb{P}(X_1 + \dots + X_n \geq nx) \leq e^{-n(\lambda x - \Lambda(\lambda))}$ . En optimisant sur  $\lambda$ , pour  $x \geq m$ , et en utilisant<sup>1</sup> la proposition 5.2, on trouve

$$\mathbb{P} \left( \frac{X_1 + \dots + X_n}{n} \geq x \right) \leq e^{-n\Lambda^*(x)}.$$

Par un raisonnement similaire, on démontre que pour tout  $x \leq m$ , on a

$$\mathbb{P} \left( \frac{X_1 + \dots + X_n}{n} \leq x \right) \leq e^{-n\Lambda^*(x)}.$$

1. Il y a ici une incohérence dans ma présentation. Pour utiliser la proposition 5.2, il faut supposer que la fonction  $\Lambda$  est finie en un point strictement positif. L'argument que j'ai écrit est correct sous l'hypothèse qu'il existe un point strictement positif et un point strictement négatif où  $\Lambda$  est finie, ce qui équivaut à supposer qu'il existe  $\alpha > 0$  tel que  $\mathbb{E}[e^{\alpha|X_1|}] < \infty$ .

Il est entendu que si  $m$  est infini, une seule de ces inégalités est pertinente.

Soit maintenant  $F$  un fermé non vide de  $\mathbb{R}$ . S'il contient  $m$ , alors l'inf de  $\Lambda^*$  sur  $F$  vaut 0 et l'inégalité est vraie. Supposons donc que  $m$  n'appartient pas à  $F$ . Posons alors

$$a = \sup(F \cap ]-\infty, m[) \quad \text{et} \quad b = \inf(F \cap ]m, +\infty[).$$

On a donc  $a < m < b$  et puisque  $F$  est non vide, l'un au moins de  $a$  et  $b$  est fini. Supposons qu'ils le soient tous les deux, les autres cas se traitant de manière similaire. Alors puisque  $\Lambda^*$  décroît avant  $m$  et croît après, on a  $\inf\{\Lambda^*(x) : x \in F\} = \min(\Lambda^*(a), \Lambda^*(b))$ . Ainsi,

$$\begin{aligned} \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \in F\right) &\leq \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \leq a\right) + \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \geq b\right) \\ &\leq e^{-n\Lambda^*(a)} + e^{-n\Lambda^*(b)} \\ &\leq 2e^{-n \inf\{\Lambda^*(x) : x \in F\}}. \end{aligned}$$

Il s'agit d'une inégalité plus précise que la borne supérieure annoncée.

2. Pour la démonstration de la borne inférieure, nous allons supposer que nos variables aléatoires sont bornées, et nous allons noter  $[a, b]$  l'enveloppe convexe du support de leur loi. Il s'ensuit que  $\Lambda(\lambda)$  est fini pour tout  $\lambda$ , équivalent à  $a\lambda$  en  $-\infty$  et à  $b\lambda$  en  $+\infty$ . En particulier,  $\Lambda^*$  est infinie hors de l'intervalle  $[a, b]$ , et finie sur  $]a, b[$ . Notons aussi que  $\Lambda$  est lisse, et que  $\Lambda'$  prend toutes les valeurs strictement comprises entre  $a$  et  $b$ .

Nous allons montrer que pour tout  $x \in \mathbb{R}$  et tout  $\delta > 0$ , on a

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \in ]x - \delta, x + \delta[ \right) \geq -\Lambda^*(x).$$

Notons que,  $x$  étant fixé, il suffit de démontrer cette inégalité pour  $\delta$  assez petit pour qu'elle soit établie pour tout  $\delta$ .

Si  $x \notin [a, b]$ , l'inégalité est vraie car le membre de droite vaut  $-\infty$ . Il n'est pas nécessaire de traiter les cas  $x = a$  et  $x = b$ , car un ouvert qui contient  $a$  ou  $b$  contient un point intérieur à l'intervalle  $[a, b]$ . Il suffit donc de considérer le cas où  $a < x < b$ .

Pour étudier la probabilité que la moyenne empirique prenne une valeur proche de  $x$ , nous allons modifier la loi de nos variables aléatoires par un changement de mesure exponentiel, de manière à amener leur moyenne en  $x$ , et voir ce qu'il nous en aura coûté.

Notons  $\mu$  la loi de  $X_1$ . Nous cherchons à définir une mesure  $\tilde{\mu}$  telle que

$$d\tilde{\mu}(x) \propto e^{\theta x} d\mu(x)$$

pour un certain  $\theta \in \mathbb{R}$ , de telle sorte que la moyenne de  $\tilde{\mu}$  soit  $x$ . Tout d'abord, l'expression correcte de  $\tilde{\mu}$ , pour en faire une mesure de probabilité, est

$$d\tilde{\mu}(x) = \frac{1}{\mathbb{E}[e^{\theta X_1}]} e^{\theta x} d\mu(x) = e^{\theta x - \Lambda(\theta)} d\mu(x). \quad (5.1)$$

Ensuite, la moyenne de  $\tilde{\mu}$  vaut, d'après le lemme 5.3

$$\int_{\mathbb{R}} x d\tilde{\mu}(x) = \frac{\mathbb{E}[X_1 e^{\theta X_1}]}{\mathbb{E}[e^{\theta X_1}]} = \Lambda'(\theta).$$



Nous avons déjà observé que  $\Lambda'$  prenait toutes les valeurs dans l'intervalle  $]a, b[$ . Nous définissons donc  $\theta$  par la relation  $\Lambda'(\theta) = x$ . Ceci définit uniquement  $\theta$  (à moins que nos variables aléatoires ne soient toutes identiquement nulles), mais il nous suffit de savoir qu'il est possible de choisir un  $\theta$  qui satisfasse cette égalité. Nous définissons ensuite  $\tilde{\mu}$  par (5.1).

Nous pouvons maintenant estimer la probabilité qui nous intéresse. Fixons  $n \geq 1$  et  $\varepsilon > 0$  et définissons

$$D = \{(x_1, \dots, x_n) \in \mathbb{R}^n : \frac{x_1 + \dots + x_n}{n} \in ]x - \varepsilon, x + \varepsilon[ \}$$

Nous avons

$$\begin{aligned} \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \in ]x - \varepsilon, x + \varepsilon[ \right) &= \int_D d\mu(x_1) \dots d\mu(x_n) \\ &\geq e^{-n\theta(x+\varepsilon)} \int_D e^{\theta(x_1 + \dots + x_n)} d\mu(x_1) \dots d\mu(x_n) \\ &= e^{-n\theta(x+\varepsilon) + n\Lambda(\theta)} \int_D d\tilde{\mu}(x_1) \dots d\tilde{\mu}(x_n) \\ &= e^{-n\theta(x+\varepsilon) + n\Lambda(\theta)} \mathbb{P}\left(\frac{Y_1 + \dots + Y_n}{n} \in ]x - \varepsilon, x + \varepsilon[ \right), \end{aligned}$$

où  $(Y_n)_{n \geq 1}$  est une suite i.i.d. de loi  $\tilde{\mu}$ . En utilisant encore le lemme 5.3, on peut réécrire le préfacteur exponentiel sous la forme

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \in ]x - \varepsilon, x + \varepsilon[ \right) \geq e^{-n\Lambda^*(x) - n\theta\varepsilon} \mathbb{P}\left(\frac{Y_1 + \dots + Y_n}{n} \in ]x - \varepsilon, x + \varepsilon[ \right).$$

La loi forte des grands nombres nous assure que le probabilité du membre de droite tend, lorsque  $n$  tend vers l'infini, vers 1. On en déduit que pour tout  $\varepsilon$  tel que  $0 < \varepsilon \leq \delta$ , on a

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \in ]x - \delta, x + \delta[ \right) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \in ]x - \varepsilon, x + \varepsilon[ \right) \\ &\geq -\Lambda^*(x) - \theta\varepsilon. \end{aligned}$$

On obtient la borne inférieure en faisant tendre  $\varepsilon$  vers 0. □

**Troisième partie**

**Introduction à la percolation**

---

Cette troisième et dernière partie du cours est un peu différente des deux précédentes : la topologie n'y jouera à peu près aucun rôle. Elle est consacrée à une introduction au modèle de la percolation par arêtes, et à son étude sur le réseau carré (ou cubique, ou hyper-cubique). Ce sera l'occasion de rencontrer quelques méthodes de la mécanique statistique du point de vue des probabilités, et aussi de voir les probabilités interagir avec de la géométrie.

Le modèle de percolation est un modèle de graphes aléatoires dans lequel, à partir d'un graphe donné, et pour un paramètre  $p$  fixé, on ouvre (c'est-à-dire qu'on garde) avec probabilité  $p$  et on ferme (c'est-à-dire qu'on enlève) avec probabilité  $1 - p$  chaque arête, indépendamment de toutes les autres. On étudie alors, en fonction de  $p$ , la géométrie du sous-graphe aléatoire obtenu. Ce modèle illustre de façon frappante qu'un modèle très simple à définir n'est pas nécessairement simple à étudier, et peut donner lieu à une étude très raffinée.

Notre objectif est de démontrer que le modèle de percolation sur le réseau cubique de dimension  $d \geq 2$  présente une transition de phase (théorème 6.2.6), et que pour  $d = 2$ , cette transition de phase a lieu lorsque  $p = \frac{1}{2}$  (proposition 8.2.2). Pour ce faire, il nous faudra en particulier étudier le nombre d'agrégats infinis d'une configuration sur-critique (Proposition 7.2.4) et contrôler la taille des agrégats, finis, d'une configuration sous-critique (proposition 7.5.1). Nous étudierons aussi, dans le contexte de la percolation, certaines inégalités importantes comme l'inégalité FKG (proposition 7.4.1). L'ensemble des résultats que nous démontrons est résumé dans le théorème 6.1.5.

Ma source principale pour ce chapitre est un texte écrit par Hugo Duminil-Copin [**Duminil-Copin**] et disponible à l'adresse <http://www.ihes.fr/~duminil/publi/2017percolation.pdf>. J'ai aussi utilisé le livre de Geoffrey Grimmett [**Grimmett**] qui est un des ouvrages de référence sur le sujet.

# 6

## Une transition de phase

---

Le modèle de percolation (par arêtes) consiste à former un sous-graphe aléatoire d'un graphe donné en ne conservant qu'un sous-ensemble aléatoire de l'ensemble des arêtes, et à étudier la géométrie — aléatoire — de ce sous-graphe.

### 6.1 Le modèle de la percolation par arêtes

#### 6.1.1 Graphes

Il y a beaucoup de définitions d'un graphe, suivant qu'on veuille ou non en orienter les arêtes, autoriser des arêtes multiples entre une paire de sommets, autoriser ou non les boucles. Nous adopterons la suivante, qui a l'avantage de la simplicité, pour laquelle les arêtes ne sont pas orientées, et qui n'autorise ni boucles ni arêtes multiples.

**Définition 6.1.1.** On appelle graphe une paire  $G = (V, E)$ , où  $V$  est un ensemble non vide et  $E$  est une partie de  $\mathcal{P}_2(V)$ , l'ensemble des parties à deux éléments de  $V$ .

Les éléments de  $V$  s'appellent les sommets (*vertices* en anglais) de  $G$ , et les éléments de  $E$  ses arêtes (*edges*). Pour tous sommets  $x, y \in V$ , on écrira  $x \sim y$ , et on dira que  $x$  et  $y$  sont voisins, pour indiquer que la paire  $\{x, y\}$  appartient à  $E$ .

On dit qu'un graphe est *localement fini* si tous ses sommets ont un nombre fini de voisins. On dit qu'il est *régulier* si tous ses sommets ont le même nombre (fini) de voisins, et on dira qu'il est  *$m$ -régulier* si ce nombre de voisins est  $m$ . On dit qu'il est *transitif* si le groupe de ses automorphismes agit transitivement sur  $V$ , c'est-à-dire si pour tous  $x, y \in V$  il existe une bijection  $f : V \rightarrow V$  telle que  $f(x) = y$  et pour toute arête  $\{z, w\}$ , la paire  $\{f(z), f(w)\}$  soit encore une arête. Nous concentrerons notre étude de la percolation sur la famille suivante de graphes transitifs.

**Exemple 6.1.2.** Soit  $d \geq 1$  un entier. On note comme toujours  $\mathbb{Z}$  l'ensemble des entiers relatifs, et on définit

$$\mathbb{E}^d = \{(x, y) \in \mathbb{Z}^d : \|x - y\| = 1\},$$

la distance étant la distance euclidienne usuelle. Le graphe

$$\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$$

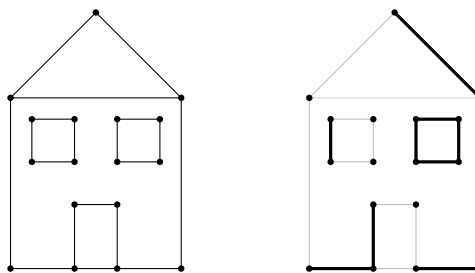
s'appelle le réseau (*lattice* en anglais) linéaire ( $d = 1$ ), ou carré ( $d = 2$ ), ou cubique ( $d = 3$ ), ou hypercubique ( $d \geq 4$ ), de dimension  $d$ . C'est un graphe  $2d$ -régulier et transitif.

On appelle *chemin* dans un graphe  $G = (V, E)$  une suite finie  $(x_0, \dots, x_n)$  de sommets dont chacun est voisin du suivant : pour tout  $i \in \{0, \dots, n-1\}$ , on a  $x_i \sim x_{i+1}$ . On dit que le chemin  $(x_0, \dots, x_n)$  est *issu* de  $x_0$ , et qu'il relie, ou joint,  $x_0$  à  $x_n$ . L'entier  $n$  s'appelle la longueur du chemin. C'est le nombre d'arêtes traversées par le chemin. Notons qu'avec notre définition, un chemin ne peut pas stationner en un sommet.

On remarquera par exemple que dans un graphe  $m$ -régulier, il y a  $n^m$  chemins de longueur  $n$  issus d'un sommet donné.

### 6.1.2 Espace de configuration et mesure de percolation

On appelle *configuration de percolation* sur un graphe  $G = (V, E)$  une partie de l'ensemble  $E$ . Les arêtes qui appartiennent à cette partie sont dites *ouvertes*, ou *présentes*, les autres sont dites *fermées*, ou *absentes*. La figure ci-dessous représente un graphe et une configuration de percolation sur ce graphe. Les arêtes ouvertes sont représentées en gras.



Le modèle de percolation consiste en une mesure de probabilité sur l'ensemble des configurations de percolation. Pour formaliser ceci, convenons de coder par 1 les arêtes ouvertes et 0 les arêtes fermées. L'espace de configuration est alors l'ensemble

$$\{0, 1\}^E \simeq \mathcal{P}(E)$$

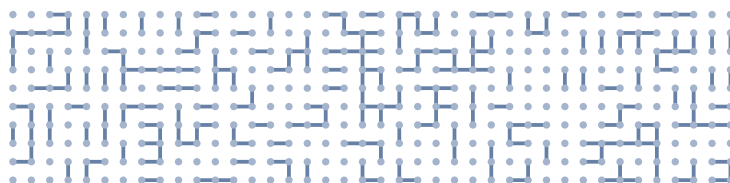
que nous munirons de la tribu  $\mathcal{F}$  engendrée par toutes les projections  $\{0, 1\}^E \rightarrow \{0, 1\}^F$  où  $F$  décrit l'ensemble des parties finies de  $E$ . En d'autres termes,  $\mathcal{F}$  est la tribu produit  $\mathcal{P}(\{0, 1\})^{\otimes E}$ .

**Définition 6.1.3.** Soit  $p \in [0, 1]$  un réel. La mesure de percolation (par arêtes) de paramètre  $p$  sur le graphe  $G = (V, E)$  est la mesure

$$\mathbb{P}_p = ((1-p)\delta_0 + p\delta_1)^{\otimes E}$$

sur l'espace mesurable  $(\{0, 1\}^E, \mathcal{F})$ .

Sous la mesure  $\mathbb{P}_p$ , chaque arête est ouverte avec probabilité  $p$  et fermée avec probabilité  $1-p$ , indépendamment de toutes les autres. Ci-dessous, une simulation de percolation sur une grille carrée  $10 \times 40$  avec  $p = 0.3$ .

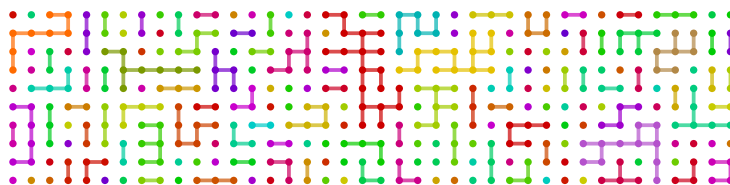


### 6.1.3 Événements de connexion et agrégat d'un sommet

Considérons une configuration de percolation sur notre graphe  $G = (V, E)$ , c'est-à-dire un élément  $\omega = (\omega_e)_{e \in E}$  de  $\{0, 1\}^E$ .

- On dit qu'une arête  $e$  est ouverte dans la configuration  $\omega$  si  $\omega_e = 1$ .
- On dit qu'un chemin  $c = (x_0, \dots, x_n)$  est ouvert dans la configuration  $\omega$  si toutes les arêtes qui le composent sont ouvertes dans la configuration  $\omega$ .
- On dit que deux sommets distincts  $x$  et  $y$  sont reliés dans la configuration  $\omega$  s'il existe un chemin de  $x$  à  $y$  qui soit ouvert dans la configuration  $\omega$ . On écrit dans ce cas  $x \xleftrightarrow{\omega} y$ .
- On dit que deux parties  $A$  et  $B$  de  $V$  sont reliées dans la configuration  $\omega$  s'il existe un chemin ouvert dans la configuration  $\omega$  qui relie un élément de  $A$  à un élément de  $B$ . On écrit dans ce cas  $A \xleftrightarrow{\omega} B$ .
- Pour toute configuration  $\omega$ , la relation  $\xleftrightarrow{\omega}$  est une relation d'équivalence sur  $V$ , dont les classes d'équivalence s'appellent les *agrégats* de la configuration  $\omega$ . Géométriquement, ces agrégats sont les composantes connexes du sous-graphe  $(V, \{e \in E : \omega_e = 1\})$  de  $G$ .

On voit ci-dessous les agrégats de la configuration sur une grille carrée  $10 \times 40$  représentée plus haut.



Une question typique que l'on pourrait poser sur cette configuration est de savoir s'il existe un agrégat qui contient à la fois un sommet du bord supérieur de la grille et un sommet de son bord inférieur. Dans ce cas, la réponse est non.

De manière duale, mais très naturelle, on notera par exemple

$$\{x \longleftrightarrow y\} = \{\omega \in \{0, 1\}^E : x \xleftrightarrow{\omega} y\}$$

l'ensemble des configurations pour lesquelles  $x$  et  $y$  sont dans le même agrégat. On pourra s'interroger, par exemple, sur la valeur de la probabilité

$$\mathbb{P}_p(x \longleftrightarrow y).$$

La définition la plus importante est toutefois la suivante : pour tout sommet  $x$  et toute configuration  $\omega$ , on appelle *agrégat* (ou cluster) de  $x$  (dans la configuration  $\omega$ ) la partie

$$C(x, \omega) = \{y \in V : x \xleftrightarrow{\omega} y\}.$$

Il s'agit d'une partie de  $V$  qui dépend de  $\omega$  et on y pense plus volontiers comme à une partie aléatoire de  $V$ . On la note alors  $C(x)$ .

**Exercice 6.1.** On considère la percolation sur le réseau hypercubique  $\mathbb{L}^d$ . Montrer que la loi de la taille  $|C(x)|$  de l'agrégat de  $x$  ne dépend pas de  $x \in \mathbb{Z}^d$ .

### 6.1.4 Probabilité de percolation

Une partie importante de la théorie de la percolation est consacrée à l'étude de l'événement

$$\{|C(x)| = \infty\},$$

sur lequel on dit qu'il y a percolation en  $x$ . Cet événement est aussi parfois noté  $\{x \longleftrightarrow \infty\}$ .

**Définition 6.1.4.** Soit  $d \geq 1$  un entier. On se place sur le graphe  $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$  et on définit la probabilité de percolation comme la fonction  $\theta_d : [0, 1] \rightarrow [0, 1]$  donnée par

$$\theta_d(p) = \mathbb{P}_p(|C(0)| = \infty).$$

Sous  $\mathbb{P}_0$ , l'agrégat de 0 est presque sûrement réduit à  $\{0\}$  et sous  $\mathbb{P}_1$ , il vaut presque sûrement  $\mathbb{Z}^d$ , si bien que

$$\theta_d(0) = 0 \text{ et } \theta_d(1) = 1.$$

Lorsque  $d = 1$ , et pour tout  $p < 1$ , on calcule aisément la loi de la taille de l'agrégat de 0, qui est celle de la somme de deux lois géométriques indépendantes.

**Exercice 6.2.** Montrer que pour  $d = 1$  et pour tout  $n \geq 1$ , on a

$$\mathbb{P}_p(|C(0)| = n) = np^{n-1}(1-p)^2.$$

En particulier, si  $p < 1$ , l'agrégat de 0 est fini  $\mathbb{P}_p$ -presque sûrement, si bien que pour tout  $p \in [0, 1[$ , on a  $\theta_d(p) = 0$ .

En revanche, dès que  $d \geq 2$ , la situation est beaucoup plus compliquée et il est exclu d'avoir une expression simple de la fonction  $\theta_d$ . Le mieux qu'on puisse espérer est de démontrer des propriétés qualitatives de  $\theta_d$ . On peut aller très loin dans cette direction, mais contentons-nous pour l'instant de quelques observations.

Tout d'abord, il est intuitivement plausible que la fonction  $\theta_d$  soit croissante : une tendance à avoir plus d'arêtes ouvertes devrait favoriser l'événement de percolation. C'est vrai, mais on ne peut pas le démontrer en une ligne, et nous en donnerons bientôt une preuve complète, à la section 6.2.1.

Ensuite, on peut étudier la continuité de  $\theta_d$ . Le cas  $d = 1$  montre qu'elle n'est pas en général continue sur  $[0, 1]$ , mais on peut penser que c'est un cas un peu particulier. Discuter du cas  $d \geq 2$  est l'occasion d'introduire quelques notations utiles.

Étant donné une partie  $A$  de  $V$ , on appelle *frontière* ou *bord* de  $A$  l'ensemble

$$\partial A = \{x \in A : \exists y \notin A, \{x, y\} \in E\}.$$

Dans le réseau  $\mathbb{L}^d$ , nous définissons pour tout  $n \geq 0$  la partie

$$B_n = B_n^d = \{x \in \mathbb{Z}^d : \|x\|_\infty \leq n\} = \{-n, \dots, n\}^d.$$

Une remarque importante est l'égalité d'événements

$$\{|C(0)| = \infty\} = \bigcap_{n \geq 0} \{0 \longleftrightarrow \partial B_n\}.$$

De plus, comme la suite d'événements  $(\{0 \longleftrightarrow \partial B_n\})_{n \geq 0}$  est décroissante, on a

$$\theta_d(p) = \lim_{n \rightarrow \infty} \downarrow \mathbb{P}_p(0 \longleftrightarrow \partial B_n). \quad (6.1)$$

Or la probabilité  $\mathbb{P}_p(0 \longleftrightarrow \partial B_n)$  est la somme sur toutes les configurations de percolation dans la boîte  $B_n$  pour lesquelles 0 est relié au bord de la boîte de  $p$  à la puissance le nombre d'arêtes ouvertes fois  $(1-p)$  puissance le nombre d'arêtes fermées. En particulier, c'est une fonction polynomiale de  $p$ .

À titre d'exemple (très simple, mais instructif sur les liens entre la percolation dans  $\mathbb{Z}^d$  et la percolation dans  $B_n$ ), on peut vérifier que  $\mathbb{P}_p(0 \longleftrightarrow \partial B_1) = 1 - (1-p)^{2d}$ .

**Exercice 6.3.** Donner une version rigoureuse et détaillée de l'argument précédent. On pourra introduire le sous-graphe fini de  $\mathbb{L}^d$  associé à la boîte  $B_n$ , vérifier que la restriction à ce sous-graphe de la percolation sur  $\mathbb{L}^d$  est la percolation sur ce sous-graphe, et que 0 est relié à  $\partial B_n$  dans une configuration de percolation sur  $\mathbb{L}^d$  si et seulement si il l'est dans la restriction à  $B_n$  de cette configuration.

**Exercice 6.4.** Si  $x$  et  $y$  sont deux sommets distincts de  $\mathbb{Z}^d$ , la probabilité  $\mathbb{P}_p(x \longleftrightarrow y)$  est-elle polynomiale en  $p$  ?

L'équation (6.1) nous montre donc que la fonction  $\theta_d$  est une limite décroissante de fonctions continues : elle est donc semi-continue supérieurement.

**Exercice 6.5.** Montrer qu'une fonction sur  $[0, 1]$  croissante et semi-continue supérieurement est continue à droite.

Une caractéristique remarquable du modèle de percolation est que la fonction  $\theta_d$  n'est pas du tout une fonction analytique de  $p$ , même sur l'intervalle  $]0, 1[$  : nous allons par exemple démontrer que pour tout  $d \geq 1$ , elle est identiquement nulle sur un intervalle non réduit à  $\{0\}$ .

Pour exprimer ce fait, on définit, pour tout  $d \geq 1$ , la *probabilité critique*

$$p_c(d) = \sup\{p \in [0, 1] : \theta_d(p) = 0\}.$$

Nous avons maintenant défini les objets et les quantités qui nous permettent d'énoncer le théorème suivant, qui résume l'ensemble des propriétés de la percolation par arêtes sur le réseau  $\mathbb{L}^d$  que nous espérons démontrer. Nous ne démontrerons peut-être pas tout, et certainement pas dans l'ordre de l'énoncé. Le cas  $d = 1$  étant élémentaire et un peu particulier, nous le laissons de côté.

**Théorème 6.1.5** (Percolation sur  $\mathbb{L}^d$ ). Soit  $d \geq 2$  un entier. Notons  $p_c = p_c(d)$ .

1. La probabilité critique vérifie  $0 < p_c < 1$ .
2. La fonction  $\theta_d$  est nulle sur l'intervalle  $[0, p_c[$ , continue à droite en  $p_c$ , strictement positive, croissante et continue sur l'intervalle  $]p_c, 1]$ , et elle vaut 1 en 1.
3. Supposons  $p < p_c$ . Alors presque sûrement sous  $\mathbb{P}_p$ , tous les agrégats sont finis. De plus, il existe une constante  $\alpha > 0$  telle que  $\mathbb{P}_p(0 \longleftrightarrow \partial B_n) \leq e^{-\alpha n}$ .
4. Supposons  $\theta_d(p) > 0$ . Alors il y a presque sûrement sous  $\mathbb{P}_p$  un unique agrégat infini.
5. On a  $p_c(2) = \frac{1}{2}$ . De plus,  $\theta_2(\frac{1}{2}) = 0$ .



Notons que cet énoncé ne précise pas, pour  $d \geq 3$ , si la fonction  $\theta_d$  est continue en 0, c'est-à-dire si  $\theta_d(p_c) = 0$ . Cette égalité a été démontrée lorsque  $d \geq 11$ , et les gens qui les connaissent estiment que les méthodes utilisées dans ce cas devraient permettre de traiter le cas  $d \geq 7$ . Pour  $d \in \{3, 4, 5, 6\}$ , la question semble encore ouverte (je ne suis pas sûr du statut de la question pour  $d \in \{4, 5, 6\}$ ), et ce d'autant plus que  $d$  est petit, et le cas  $d = 3$  constitue l'une des conjectures les plus célèbres de la théorie des probabilités aujourd'hui — en partie par le contraste extrême entre la simplicité de son énoncé et le fait qu'elle ait déjà résisté à plus de 60 ans d'efforts soutenus.

Notons aussi que la réponse à cette conjecture ne présuppose pas, comme on pourrait le croire, la détermination de  $p_c(d)$  pour  $d \geq 3$ . Il y a en fait très peu de graphes (même des réseaux bi-dimensionnels autres que le réseau carré) pour lesquels la valeur exacte de la probabilité critique soit connue, et il ne semble pas y avoir de raison pour laquelle ces valeurs critiques devraient avoir une expression simple.<sup>1</sup>

Ajoutons à ce bref panorama que lorsque  $d = 2$ , le comportement de  $\theta_2(p)$  pour  $p$  voisin de  $p_c$  est connu pour de nombreux réseaux (carré, triangulaire, ou plus exotiques), et ne dépend, dans une certaine mesure, pas du détail du réseau. Ainsi, sur  $\mathbb{L}^2$ , on sait que pour  $p$  voisin de (et supérieur à)  $p_c$ , on a

$$\theta_2(p) \sim (p - p_c)^{\frac{5}{36}}$$

et on s'attend à ce que ce comportement soit commun à de très nombreux modèles de percolation, et de physique statistique, en dimension 2. L'envie de comprendre pourquoi un exposant rationnel (et lequel) apparaît dans ce problème et dans de nombreuses questions connexes a suscité de très nombreux travaux pendant ces quinze ou vingt dernières années, et a conduit à l'appropriation en probabilités d'idées issues de la physique théorique, plus précisément de la théorie conforme des champs. Il s'agit d'un domaine de recherche actuellement très actif.

## 6.2 Existence d'une transition de phase

Notre premier objectif est de montrer que la probabilité critique n'est ni 0 ni 1 lorsque  $d \geq 2$ . Avant cela, nous allons construire un couplage des mesures de percolation associées à tous les paramètres dans l'intervalle  $[0, 1]$ . La notion de domination stochastique, que nous introduisons ici, est d'usage très fréquent en physique statistique.

### 6.2.1 Domination stochastique

Au début de cette section, nous pouvons travailler sur un graphe quelconque  $G = (V, E)$ . Une remarque fondamentale est que l'espace de configuration du modèle de percolation, qui est l'ensemble des parties de  $E$ , est ordonné par l'inclusion. Sur l'espace mesurable  $(\{0, 1\}^E, \mathcal{F})$ , nous allons définir une classe particulière d'événements.

**Définition 6.2.1.** *On dit qu'un événement est croissant si sa fonction indicatrice est croissante. En d'autres termes, un événement  $A \in \mathcal{F}$  est croissant si pour toutes configurations  $\omega$  et  $\omega'$ , on a*

$$(\omega \in A \text{ et } \omega \leq \omega') \Rightarrow \omega' \in A.$$

1. Cette question est un terrain idéal d'investigation numérique, comme on peut le voir en consultant par exemple la page [https://en.wikipedia.org/wiki/Percolation\\_threshold](https://en.wikipedia.org/wiki/Percolation_threshold).

Un événement croissant est un événement, c'est-à-dire une classe de parties de  $E$  qui, dès qu'elle contient une partie  $F$ , contient toutes les parties de  $E$  qui contiennent  $F$ . Des exemples typiques d'événements croissants sont  $\{x \leftarrow y\}$ ,  $\{|C(0)| \geq 100\}$ ,  $\{0 \leftarrow \infty\}$ .

Sur un espace mesurable ordonné, il y a une notion de domination entre mesures de probabilités.

**Définition 6.2.2.** Soient  $\mathbb{P}$  et  $\mathbb{Q}$  deux mesures de probabilité sur  $(\{0,1\}^E, \mathcal{F})$ . On dit que  $\mathbb{Q}$  domine stochastiquement  $\mathbb{P}$  si pour tout événement croissant  $A$ , on a  $\mathbb{P}(A) \leq \mathbb{Q}(A)$ .

Nous allons démontrer le résultat suivant. Nous notons ici  $\lambda$  la mesure de Lebesgue sur l'intervalle  $[0, 1]$ .

**Proposition 6.2.3.** Soit  $G = (V, E)$  un graphe. Sur l'espace de probabilité  $([0, 1]^E, \mathcal{B}_{[0,1]}^{\otimes E}, \lambda^{\otimes E})$ , il existe une famille  $(\omega^p)_{p \in [0,1]}$  de variables aléatoires à valeurs dans  $(\{0, 1\}^E, \mathcal{F})$  telle que

1. pour tout  $p \in [0, 1]$ , la loi de  $\omega^p$  soit la mesure  $\mathbb{P}_p$ ,
2. pour tous  $p \leq p'$ , on ait presque sûrement  $\omega^p \leq \omega^{p'}$ .

En particulier, pour tous  $p \leq p'$ , la mesure  $\mathbb{P}_{p'}$  domine stochastiquement  $\mathbb{P}_p$ .

*Démonstration.* Un élément de l'espace de probabilité  $([0, 1]^E, \mathcal{B}_{[0,1]}^{\otimes E}, \lambda^{\otimes E})$  est une famille  $x = (x_e)_{e \in E}$  de réels compris entre 0 et 1, indexée par les arêtes de  $G$ . Pour tout  $p \in [0, 1]$  et toute arête  $e \in E$ , posons

$$\omega_e^p(x) = \begin{cases} 1 & \text{si } x_e \leq p \\ 0 & \text{sinon.} \end{cases}$$

Pour tout  $p \in [0, 1]$ , les variables aléatoires  $(\omega_e^p)_{e \in E}$  sont indépendantes et de loi  $p\delta_1 + (1-p)\delta_0$ , si bien que  $\omega^p$  a la loi  $\mathbb{P}_p$ . De plus, si  $p \leq p'$ , alors pour toute arête  $e$ , on a

$$\omega_e^p = 1 \Rightarrow x_e \leq p \Rightarrow x_e \leq p' \Rightarrow \omega_e^{p'} = 1,$$

si bien que  $\omega^p \leq \omega^{p'}$  presque sûrement.

Soit maintenant  $A$  un événement croissant. Alors on a

$$\mathbb{P}_p(A) = \lambda^{\otimes E}(\omega^p \in A) \leq \lambda^{\otimes E}(\omega^{p'} \in A) = \mathbb{P}_{p'}(A),$$

ce qui démontre la domination stochastique annoncée.  $\square$

Notons la méthode que nous avons employée pour montrer qu'une certaine mesure  $\mathbb{Q}$  domine stochastiquement une autre mesure  $\mathbb{P}$  : nous avons construit un couple de variables aléatoires  $(X, Y)$  telles que  $X$  soit de loi  $\mathbb{P}$ ,  $Y$  de loi  $\mathbb{Q}$ , et  $X \leq Y$  presque sûrement. Une fois ces variables aléatoires construites, la démonstration du fait que  $\mathbb{Q}$  domine stochastiquement  $\mathbb{P}$  prend une ligne. On peut se demander si on a eu de la chance, ou s'il existe toujours un tel couple. Un théorème dit de Strassen répond positivement à cette dernière question et assure (sous certaines hypothèses techniques) que si une mesure en domine stochastiquement une autre, alors il existe un couple de variables aléatoires dont l'une est plus grande que l'autre presque sûrement, et qui ont pour lois respectives les mesures en question.

De l'existence du couplage que nous venons de construire, ou de la domination stochastique que nous en avons déduite, nous déduisons maintenant la monotonie de la probabilité de percolation.

**Corollaire 6.2.4.** *Pour tout  $d \geq 1$ , la fonction  $\theta_d : [0, 1] \rightarrow [0, 1]$  est croissante.*

*Démonstration.* L'événement  $\{|C(0)| = \infty\}$  est croissant et pour tous  $p \leq p'$ , on a

$$\theta_d(p) = \mathbb{P}_p(|C(0)| = \infty) \leq \mathbb{P}_{p'}(|C(0)| = \infty) = \theta_d(p'),$$

ce qu'il fallait démontrer.  $\square$

Il découle de ce qui précède que l'ensemble  $\{p \in [0, 1] : \theta_d(p) = 0\}$  est un intervalle, que

$$p_c(d) = \inf\{p \in [0, 1] : \theta_d(p) > 0\}$$

et que la fonction  $\theta_d$  est nulle sur l'intervalle  $[0, p_c(d)[$  et strictement positive sur l'intervalle  $]p_c(d), 1]$ .

**Corollaire 6.2.5.** *La fonction  $\theta_d$  est continue à droite sur  $[0, 1]$ .*

*Démonstration.* Il découle de l'équation (6.1) qu'elle est la limite décroissante d'une suite de fonctions continues. Elle est donc semi-continue supérieurement. Comme elle est croissante, elle est donc continue à droite.  $\square$

Pour montrer que  $p_c$  n'est ni 0 ni 1 lorsque  $d \geq 2$ , il suffit donc de montrer qu'il existe deux paramètres  $p$  et  $p'$  tels que  $0 < p < p' < 1$  tels que  $\theta_d(p) = 0$  et  $\theta_d(p') > 0$ . C'est ce que nous allons faire dans la section suivante.

## 6.2.2 Encadrement de la probabilité critique

**Théorème 6.2.6** (Transition de phase). *Supposons  $d \geq 2$ . Alors  $0 < p_c(d) < 1$ .*

Nous allons en fait montrer que pour tout  $d \geq 2$ , on a

$$\frac{1}{2d-1} \leq p_c(d) \leq p_c(2) \leq \frac{3}{4}.$$

Cet encadrement de  $p_c(d)$  deviendra évidemment meilleur quand nous aurons déterminé la valeur de  $p_c(2)$ .

Commençons par montrer que  $0 < p_c(d)$ . Pour cela, nous allons considérer des chemins particuliers issus de l'origine, que nous appelons *auto-évitants*.

**Définition 6.2.7.** *On dit qu'un chemin  $(x_0, \dots, x_n)$  est auto-évitant si ses sommets  $x_0, \dots, x_n$  sont deux à deux distincts. On dit que c'est un lacet auto-évitant si  $x_0, \dots, x_{n-1}$  sont deux à deux distincts et  $x_n = x_0$ .*

Pour tout  $n \geq 1$ , notons  $A_n$  l'ensemble des chemins auto-évitants dans  $\mathbb{L}^d$  issus de 0.<sup>2</sup> Essayons d'estimer le nombre d'éléments de  $A_n$ . D'une part, les chemins qui ne font que des pas dans les directions croissantes pour chaque coordonnée sont auto-évitants, si bien que  $|A_n| \geq d^n$ . Par ailleurs, un chemin auto-évitant peut commencer par n'importe lequel des  $2d$  pas permis par le réseau, et il ne peut ensuite, en tout cas, jamais revenir immédiatement en arrière. Nous avons donc

$$d^n \leq |A_n| \leq 2d(2d-1)^{n-1}.$$

2. Nous notons 0 l'origine de  $\mathbb{Z}^d$ , ce qui ne devrait pas créer de confusion.

**Exercice 6.6.** Montrer que pour tous  $n, m \geq 0$ , on a  $|A_{n+m}| \leq |A_n||A_m|$ . En déduire qu'il existe une constante  $c \in [d, 2d - 1]$  telle que, lorsque  $n$  tend vers l'infini,

$$\log |A_n| \sim \log c^n.$$

Par ailleurs, l'agrégat de 0 est infini si et seulement s'il existe, partant de 0, des chemins auto-évitants ouverts arbitrairement longs :

$$\{|C(0)| = \infty\} = \bigcap_{n \geq 1} \{\exists c \in A_n, c \text{ ouvert}\}.$$

**Exercice 6.7.** Démontrer précisément cette égalité d'événements.

**Proposition 6.2.8.** On a  $p_c(d) \geq \frac{1}{2d-1}$ .

*Démonstration.* Soit  $p < \frac{1}{2d-1}$ . On a, pour tout  $n \geq 1$ ,

$$\mathbb{P}_p(\exists c \in A_n, c \text{ ouvert}) \leq |A_n|p^n = \frac{2d}{2d-1}((2d-1)p)^n.$$

Puisque la suite d'événements  $(\{\exists c \in A_n, c \text{ ouvert}\})_{n \geq 1}$  est décroissante, on en tire

$$\mathbb{P}_p(|C(0)| = \infty) = \lim_{n \rightarrow \infty} \downarrow \mathbb{P}_p(\exists c \in A_n, c \text{ ouvert}) = 0,$$

et donc  $p_c \geq p$ . □

Pour la majoration, commençons par observer que le cas  $d = 2$  est le pire. Pour cela, observons, sans introduire un formalisme trop lourd, que le réseau  $\mathbb{L}^2$  est naturellement un sous-réseau de  $\mathbb{L}^d$ , en considérant l'ensemble de sommets  $\mathbb{Z}^2 \times \{0\}^{d-2}$  et les arêtes qui les relient.

**Proposition 6.2.9.** On a  $p_c(d) \leq p_c(2)$ .

*Démonstration.* La clé de cette inégalité est le fait que l'intersection avec le sous-graphe  $\mathbb{L}^2$  d'une configuration de percolation tirée sous  $\mathbb{P}_p$  dans  $\mathbb{L}^d$  est une configuration de percolation dans  $\mathbb{L}^2$  sous  $\mathbb{P}_p$ .<sup>3</sup>

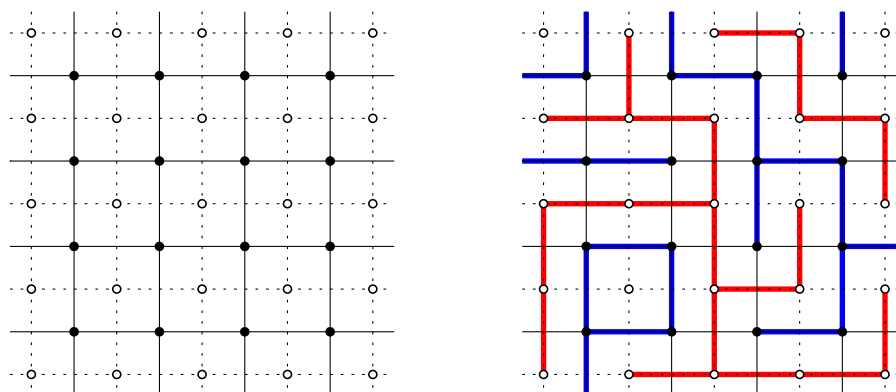
Soit donc  $p > p_c(2)$ . L'agrégat  $C(0)$  de l'origine dans  $\mathbb{L}^d$ , intersecté avec  $\mathbb{L}^2$ , a la loi de l'agrégat de l'origine dans une percolation de paramètre  $p$  dans  $\mathbb{L}^2$ . Or, par définition de  $p_c$ , on a  $\theta_2(p) > 0$ , donc avec une probabilité strictement positive,  $C(0) \cap (\mathbb{Z}^2 \times \{0\}^{d-2})$  est infini. En particulier,  $C(0)$  est infini avec une probabilité strictement positive, donc  $p \geq p_c(d)$ . Ceci entraîne l'inégalité annoncée. □

Le théorème 6.2.6 sera complètement démontré lorsque nous aurons prouvé que  $p_c(2) < 1$ . Pour démontrer cette assertion, nous allons utiliser pour la première fois une propriété très féconde du réseau  $\mathbb{L}^2$ , qui est son auto-dualité. Considérons  $\mathbb{Z}^2$  comme une partie de  $\mathbb{R}^2$ , et construisons un nouveau graphe  $\mathbb{L}_*^2 = (\mathbb{Z}_*^2, \mathbb{E}_*^2)$ , où

$$\mathbb{Z}_*^2 = (\mathbb{Z} + \frac{1}{2}) \times (\mathbb{Z} + \frac{1}{2}) \quad \text{et} \quad \mathbb{E}_*^2 = \{\{x, y\} \in \mathbb{Z}_*^2 : \|x - y\| = 1\}.$$

3. Cette observation justifie, un peu, le fait qu'on n'ait pas attaché à la mesure  $\mathbb{P}_p$  d'indice ou d'exposant qui indique dans quelle dimension on travaille.

Dans cette définition,  $\mathbb{Z} + \frac{1}{2}$  désigne l'ensemble  $\{\dots, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \dots\}$ . Un dessin vaut ici mieux qu'un long discours : la partie gauche du dessin suivant montre un morceau de  $\mathbb{L}^2$ , avec des sommets noirs et des arêtes pleines, et un morceau de  $\mathbb{L}_*^2$ , avec des sommets blancs et des arêtes pointillées.<sup>4</sup> On observe sur ce dessin un point fondamental : il y a une correspondance bijective évidente entre les arêtes de  $\mathbb{L}^2$  et les arêtes de  $\mathbb{L}_*^2$ . Dans la représentation que nous avons choisie, chaque arête d'un graphe rencontre en effet une unique arête du graphe dual. Pour toute arête  $e \in \mathbb{E}^d$ , nous noterons  $e_* \in \mathbb{E}_*^d$  l'arête correspondante.



Grâce à cette correspondance, nous définissons une bijection

$$D : \{0, 1\}^{\mathbb{E}^2} \longrightarrow \{0, 1\}^{\mathbb{E}_*^2}$$

$$\omega \longmapsto \omega^*$$

où pour toute arête  $e \in \mathbb{E}^2$ , nous avons posé

$$\omega_{e_*}^* = 1 - \omega_e.$$

Autrement dit, pour passer d'une configuration à la configuration duale, on commence par prendre la configuration complémentaire, puis on tourne toutes les arêtes ouvertes d'un quart de tour autour de leur milieu. Cette correspondance est illustrée sur la partie droite de la figure ci-dessus, où une configuration est dessinée en bleu, et la configuration duale en rouge.

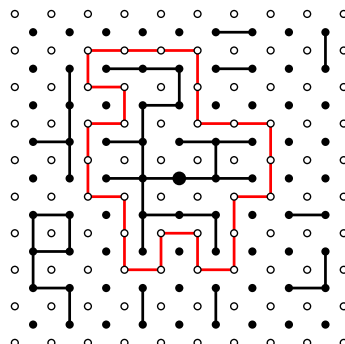
L'application  $D$  de dualité que nous venons de construire pousse manifestement la mesure de percolation de paramètre  $p$  sur  $\mathbb{L}^2$  sur la mesure de percolation de paramètre  $1 - p$  sur  $\mathbb{L}_*^2$ .

Nous allons maintenant énoncer un fait "évident", mais dont une démonstration parfaitement rigoureuse, bien qu'élémentaire, demanderait plus de temps que ce dont nous disposons. Rappelons que nous avons défini les lacets auto-évitant à la définition 6.2.7.

**Lemme 6.2.10.** *L'agrégat de 0 dans une configuration sur  $\mathbb{L}^2$  est fini si et seulement s'il existe dans la configuration duale sur  $\mathbb{L}_*^2$  un lacet auto-évitant ouvert qui encercle l'origine.*

4. À moins que ce ne soit le contraire.

*Idée de démonstration.*



□

Il ne nous reste plus qu'à montrer que si  $p$  est assez grand, la probabilité qu'il existe un lacet auto-évitant dual ouvert encerclant l'origine est strictement plus petite que 1.

**Proposition 6.2.11.** *On a  $p_c(2) \leq \frac{3}{4}$ .*

*Démonstration.* En examinant le dessin qui nous tient lieu de démonstration du lemme 6.2.10, nous voyons que la géométrie de notre problème ne nous fournit pas exactement un lacet auto-évitant, mais un *cycle* auto-évitant, que nous définissons comme un lacet auto-évitant dont on a oublié l'origine.<sup>5</sup>

Notons  $C_n$  l'ensemble des cycles auto-évitants de longueur  $n$  dans  $\mathbb{L}_*^2$  qui entourent l'origine. Il découle du lemme 6.2.10 que

$$\{|C(0)| < \infty\} = \bigcup_{n \geq 0} \{\exists c \in C_n, c \text{ ouvert}\}.$$

Nous déduisons de cette égalité d'événements que

$$\theta_2(p) \geq 1 - \sum_{n=0}^{\infty} |C_n| (1-p)^n$$

et notre but est maintenant de trouver une majoration de  $|C_n|$  qui nous assure que pour  $p$  assez proche de 1, la série qui apparaît dans le membre de droite de cette inégalité est de somme strictement plus petite que 1. On peut faire cette majoration avec plus ou moins de soin, et obtenir une plus ou moins bonne majoration de  $p_c(2)$  : le jeu est de toute façon un peu vain, car nous finirons par déterminer sa valeur exacte, mais nous allons y jouer tout de même.<sup>6</sup>

Faisons pour cela une série de remarques. Tout d'abord, un cycle auto-dual qui entoure l'origine doit traverser les quatre demi-axes de  $\mathbb{R}^2$ , et doit en particulier avoir une longueur au moins égale à 4. On peut donc commencer notre somme à 4.

Ensuite, un cycle a toujours un nombre pair d'arêtes, puisqu'à chaque pas la parité de la somme de ses coordonnées change. Nous pouvons donc restreindre notre somme aux entiers pairs.

5. Pour être précis, nous déclarons deux lacets *circulairement équivalents* s'ils diffèrent par une permutation circulaire de leurs sommets, et nous appelons cycle une classe d'équivalence circulaire de lacets.

6. Une version très simplifiée de ce qui suit permet de dire avec peu d'effort que  $|C_n| \leq n4^n$ , et de conclure (en observant tout de même que  $C_1$  est vide, ce qui améliore légèrement la minoration) que  $p_c(2) < 0.9$ .

Observons maintenant que parce qu'il traverse le demi-axe horizontal positif, un cycle auto-évitant passe nécessairement par un sommet d'abscisse positive et d'ordonnée  $\frac{1}{2}$ . Distinguons, parmi tous les sommets visités par le cycle, celui qui est de cette forme et qui a la plus grande abscisse. Notons  $x$  ce sommet, et basons notre cycle en  $x$ . On peut se convaincre en faisant un petit dessin que l'abscisse du sommet  $x$  ne peut être supérieure à  $\frac{n}{2} - 1$ . Le sommet  $x$  appartient donc l'un des  $\frac{n}{2} - 1$  sommets  $(\frac{1}{2}, \frac{1}{2}), (\frac{3}{2}, \frac{1}{2}), \dots, (\frac{n-3}{2}, \frac{1}{2})$ .

À tout élément de  $C_n$ , nous avons donc associé un lacet auto-évitant de longueur  $n$  basé en un élément d'un ensemble de  $\frac{n}{2} - 1$  sommets, d'une manière évidemment injective, puisque nous retrouvons le cycle dont nous sommes partis en oubliant l'origine du lacet. Ce lacet auto-évitant est en particulier un chemin auto-évitant, et nous avons choisi le point base de telle manière que le premier pas de ce chemin ne peut être un pas vers la droite.

De toutes ces considérations<sup>7</sup>, nous déduisons que

$$|C_n| \leq \frac{n-2}{2} 3^n,$$

puis, en posant  $n = 2m$ , que

$$\theta_2(p) \geq 1 - \sum_{m=2}^{\infty} (m-1)(9(1-p)^2)^m = \frac{1-2q}{(1-q)^2},$$

où on a posé  $q = 9(1-p)^2$ . On en déduit que  $\theta_2(p)$  est strictement positif dès que  $q < \frac{1}{2}$ , c'est-à-dire dès que  $p > 1 - \frac{1}{3\sqrt{2}} \simeq 0.764$ .

Notre majoration du cardinal de  $C_n$  est très grossière même pour de très petites valeurs de  $n$ , ainsi, nous avons utilisé  $|C_4| \leq 81$  alors que  $|C_4| = 1$  et  $|C_6| \leq 1458$  alors que  $|C_6| = 4$ . En tenant compte de ces deux valeurs, on trouve

$$\begin{aligned} \theta_2(p) &\geq 1 - (1-p)^4 - 4(1-p)^6 - \sum_{m=4}^{\infty} (m-1)(9(1-p)^2)^m \\ &= 1 - (1-p)^4 - 4(1-p)^6 + \frac{19683(1-p)^8(6(1-p)^2 - 1)}{(1 - 9(1-p)^2)^2}, \end{aligned}$$

dont on déduit

$$\theta_2\left(\frac{3}{4}\right) \geq \frac{1447}{100352} > 0.$$

Nous n'essaierons pas d'aller plus loin. □

Faisons un petit bilan de ce que nous avons démontré, en revenant au théorème 6.1.5. Nous avons démontré la première assertion, et presque entièrement la deuxième, à un point près : nous n'avons pas démontré que la fonction  $\theta_d$  était continue sur l'intervalle  $]p_c(d), 1]$ . Comme nous savons déjà qu'elle est continue à droite, il nous reste à montrer qu'elle est continue à gauche. Nous le ferons au chapitre suivant, après avoir démontré l'unicité de l'agrégat infini dans la phase sur-critique.

7. Il y a quelque chose d'un peu ridicule à faire un argument aussi long pour obtenir un résultat aussi grossier, mais je trouve cela amusant.

# 7

## Phases sous-critique et sur-critique

---

Dans ce chapitre, nous allons étudier certaines propriétés des phases sous- et sur-critique de la percolation sur  $\mathbb{L}^d$ , c'est-à-dire les cas où  $p < p_c(d)$  et  $p > p_c(d)$ . Nous allons en particulier démontrer les assertions 3 et 4 du théorème 6.1.5.

Comme au chapitre précédent, la démonstration de ce théorème sera l'occasion de rencontrer divers outils utiles dans des situations autres que celle de la percolation.

### 7.1 Ergodicité

Fixons  $d \geq 1$  et considérons la percolation sur le réseau  $\mathbb{L}^d$ . Le groupe  $\mathbb{Z}^d$  agit par translations sur l'espace des configurations de percolation, d'une manière qui préserve la mesure. Nous allons préciser ce point, et démontrer que les événements invariants par cette action sont triviaux, au sens où ils sont de probabilité 0 ou 1 : on dit que l'action de  $\mathbb{Z}^d$  par translations est *ergodique*.

Pour tout  $v \in \mathbb{Z}^d$ , on note  $\tau_v = \mathbb{Z}^d \rightarrow \mathbb{Z}^d$  l'application définie par

$$\tau_v(x) = x + v.$$

Il s'agit d'une action du groupe additif  $\mathbb{Z}^d$  sur l'espace  $\mathbb{Z}^d$ , au sens où pour tous  $v, w \in \mathbb{Z}^d$ ,

$$\tau_v \circ \tau_w = \tau_{v+w}.$$

Pour tout  $v \in \mathbb{Z}^d$ , l'application  $\tau_v$  est une bijection de  $\mathbb{Z}^d$  qui a de plus la propriété que pour tous  $x, y \in \mathbb{Z}^d$ , on a l'équivalence

$$\{x, y\} \in \mathbb{E}^d \iff \{\tau_v(x), \tau_v(y)\} \in \mathbb{E}^d.$$

On dit que  $\tau_v$  est un automorphisme du graphe  $(\mathbb{Z}^d, \mathbb{E}^d)$ . Cet automorphisme agit sur l'espace des configurations de percolation  $\{0, 1\}^{\mathbb{E}^d}$  : à toute configuration  $\omega$ , il associe la configuration  $\tau_v(\omega)$  définie, pour toute arête  $e$ , par

$$[\tau_v(\omega)]_e = \omega_{\tau_v^{-1}(e)}.$$

En français, les arêtes ouvertes dans la configuration  $\tau_v(\omega)$  sont les images par la translation de vecteur  $v$  des arêtes ouvertes dans la configuration  $\omega$ .

L'action de  $\tau_v$  sur  $\{0, 1\}^{\mathbb{E}^d}$  est mesurable et préserve la mesure  $\mathbb{P}_p$  pour tout  $p \in [0, 1]$ .



**Définition 7.1.1.** On dit qu'un événement  $A \in \mathcal{F}$  est invariant par translations si pour tout  $v \in \mathbb{Z}^d$ , on a  $A = \tau_v(A)$ .

Pour le dire avec des mots, un événement est invariant par translations s'il est possible de savoir si cet événement est réalisé ou non en examinant la configuration de percolation *sans savoir où se trouve l'origine*.

Par exemple, les événements  $\{\exists x \in \mathbb{Z}^d, C(x) = \{x\}\}$ ,  $\{\exists x, y \in \mathbb{Z}^d, \|x - y\| \geq 50 \text{ et } x \longleftrightarrow y\}$  sont invariants par translation. L'événement  $\{0 \longleftrightarrow \infty\}$  ne l'est pas.

Notons que l'ensemble des événements invariants par translation est une sous-tribu de  $\mathcal{F}$ .

**Théorème 7.1.2** (Ergodicité de la percolation). Soit  $p \in [0, 1]$ . Soit  $A \in \mathcal{F}$  un événement invariant par translations. Alors on a  $\mathbb{P}_p(A) \in \{0, 1\}$ .

*Démonstration.* Nous allons montrer le résultat suivant. Soit  $f : \{0, 1\}^{\mathbb{E}^d} \rightarrow \mathbb{R}$  une fonction mesurable bornée. On suppose que pour tout  $v \in \mathbb{Z}^d$ , on a l'égalité  $\mathbb{P}_p$ -presque sûre  $f = f \circ \tau_v$ . Alors  $f$  est constante  $\mathbb{P}_p$ -presque sûrement. Une fois ce résultat établi, il suffira de l'appliquer à la fonction  $f = \mathbb{1}_A$  pour démontrer le théorème.

Pour démontrer notre résultat, nous allons approcher  $f$  par des fonctions qui ne dépendent de la configuration que par l'état, ouvert ou fermé, d'un nombre fini d'arêtes. Ces fonctions, étant proches de  $f$ , sont, à une petite erreur près, invariantes par translations; mais comme elles ne dépendent que de l'état d'un nombre fini d'arêtes, elles sont indépendantes de leur translation par un vecteur assez grand. Ainsi, elles sont, toujours à une petite erreur près, indépendantes d'elles-mêmes, donc constantes, et  $f$  elle-même est constante. Donnons une forme rigoureuse à cet argument.

Définissons, pour toute partie  $B$  de  $\mathbb{Z}^d$ ,  $E(B) = \{\{x, y\} : x, y \in B, x \sim y\}$  et

$$\mathcal{F}_B = \sigma((\omega \mapsto \omega_e) : e \in E(B)).$$

Rappelons que nous avons posé, pour tout  $n \geq 0$ ,

$$B_n = \{x \in \mathbb{Z}^d : \|x\|_\infty \leq n\}$$

et considérons sur  $\{0, 1\}^{\mathbb{E}^d}$  la filtration  $(\mathcal{F}_{B_n})_{n \geq 0}$ . Puisque toute partie finie de  $\mathbb{E}^d$  est contenue dans  $E(B_n)$  pour  $n$  assez grand, l'union de toutes les sous-tribus  $\mathcal{F}_{B_n}$  contient tous les événements cylindriques, et on a l'égalité

$$\sigma\left(\bigcup_{n \geq 0} \mathcal{F}_{B_n}\right) = \mathcal{F}.$$

Revenons maintenant à notre fonction  $f$ . Quitte à retrancher à  $f$  sa moyenne sous  $\mathbb{P}_p$ , ce qui n'affecte pas son invariance, nous pouvons la supposer de moyenne nulle sous  $\mathbb{P}_p$ . Nous définissons alors, pour tout  $n \geq 0$ ,<sup>1</sup>

$$f_n = \mathbb{E}_p[f | \mathcal{F}_{B_n}].$$

1. Dans l'équation qui suit,  $\mathbb{E}_p$  désigne l'espérance (en l'occurrence conditionnelle) sous la mesure  $\mathbb{P}_p$ , et non un ensemble d'arêtes.

La suite  $(f_n)_{n \geq 0}$  est une martingale par rapport à la filtration  $(\mathcal{F}_{B_n})_{n \geq 0}$ . Cette martingale est bornée dans  $L^2$ , elle converge donc dans  $L^2$  vers l'espérance conditionnelle sous  $\mathbb{P}_p$  de  $f$  sachant la tribu engendrée par l'union de tous les  $\mathcal{F}_{B_n}$ , c'est-à-dire vers  $f$ .

Pour tout  $n \geq 0$  et tout  $v \in \mathbb{Z}^d$ , nous avons

$$\|f_n - f_n \circ \tau_v\|_{L^2} \leq \|f_n - f\|_{L^2} + \|f - f \circ \tau_v\|_{L^2} + \|f \circ \tau_v - f_n \circ \tau_v\|_{L^2}.$$

Le terme du milieu est nul par l'hypothèse que  $f$  est invariante par translations, et le troisième terme est égal au premier, en vertu du fait que  $\tau_v$  préserve la mesure  $\mathbb{P}_p$ . Ainsi,

$$\|f_n - f_n \circ \tau_v\|_{L^2} \leq 2\|f_n - f\|_{L^2}.$$

Observons maintenant que la fonction  $f_n$  est mesurable par rapport à la tribu  $\mathcal{F}_{B_n}$ , si bien que la fonction  $f_n \circ \tau_v$  est mesurable par rapport à la tribu  $\mathcal{F}_{\tau_v^{-1}(B_n)}$ . Si  $\|v\|_\infty > n$ , l'égalité  $B_n \cap \tau_v^{-1}(B_n) = \emptyset$  entraîne que ces deux tribus sont indépendantes sous  $\mathbb{P}_p$ . En particulier, si  $\|v\|_\infty > n$ , alors  $f_n$  et  $f_n \circ \tau_v$ , qui sont toutes deux centrées, sont orthogonales dans  $L^2(\mathbb{P}_p)$  et en élevant la dernière équation au carré, on trouve

$$2\|f_n\|_{L^2}^2 = \|f_n - f_n \circ \tau_v\|_{L^2}^2 \leq 4\|f_n - f\|_{L^2}^2.$$

Il s'ensuit que la suite  $(f_n)_{n \geq 0}$  converge dans  $L^2$  vers 0, puis que  $f$  est nulle, ce que nous voulions démontrer.  $\square$

## 7.2 Nombre d'agrégats infinis

Nous allons nous intéresser à la famille suivante d'événements invariants par translation : pour tout  $k \in \mathbb{N} \cup \{\infty\}$ , nous posons

$$I_k = \{\text{il y a exactement } k \text{ agrégats infinis}\}.$$

Ainsi, une configuration  $\omega$  appartient à  $I_k$  si et seulement si parmi toutes les classes d'équivalence pour la relation  $\xleftrightarrow{\omega}$ , il y en a exactement  $k$  qui sont infinies.

Les événements  $(I_k)_{k \in \mathbb{N} \cup \{\infty\}}$  forment une partition dénombrable de l'espace des configurations de percolation. De cette observation et de l'ergodicité de la percolation, nous déduisons que pour tout  $p \in [0, 1]$ , il existe un unique  $k \in \mathbb{N} \cup \{\infty\}$  tel que  $\mathbb{P}_p(I_k) = 1$ .

Dans la phase sous-critique, il n'est pas difficile de démontrer que c'est  $I_0$  qui est presque sûr. Notons que nous ne faisons pas, ci-dessous, l'hypothèse  $p < p_c(d)$ , mais l'hypothèse  $\theta_d(p) = 0$ . Cette hypothèse est plus faible, car elle n'exclut pas a priori le cas  $p = p_c(d)$ .

**Proposition 7.2.1.** *Supposons  $\theta_d(p) = 0$ . Alors  $\mathbb{P}_p$ -presque sûrement, tous les agrégats sont finis.*

*Démonstration.* Il existe un agrégat infini si et seulement si il existe un sommet dont l'agrégat est infini. Or pour tout sommet, la probabilité que l'agrégat de ce sommet soit infini est égale, par invariance de la mesure par translations, à la probabilité que l'agrégat de l'origine soit infini, qui puisque nous sommes dans la phase sous-critique, est nulle. Ainsi,

$$\begin{aligned} \mathbb{P}_p(\text{il existe un agrégat infini}) &= \mathbb{P}_p(\exists x \in \mathbb{Z}^d, |C(x)| = \infty) \\ &\leq \sum_{x \in \mathbb{Z}^d} \mathbb{P}_p(|C(x)| = \infty) = \sum_{x \in \mathbb{Z}^d} \mathbb{P}_p(|C(0)| = \infty) = 0, \end{aligned}$$

ce qui démontre la proposition.  $\square$

Nous n'avons même pas eu besoin d'utiliser l'ergodicité de la percolation. Dans le cas sur-critique, l'argument est plus délicat. Commençons par éliminer toutes les possibilités sauf deux.

Ici encore, nous faisons une hypothèse sur  $\theta_d(p)$  plutôt que sur  $p$ , afin de ne pas exclure d'emblée le cas  $p = p_c(d)$ .

**Proposition 7.2.2.** *Supposons  $\theta_d(p) > 0$ . Alors  $\mathbb{P}_p(I_1 \cup I_\infty) = 1$ .*

*Démonstration.* Tout d'abord, la probabilité qu'il y ait au moins un agrégat est d'une part égale à 0 ou 1 par ergodicité, et d'autre part minorée par la probabilité que l'agrégat de l'origine soit infini, qui est strictement positive. En symboles,

$$0 < \mathbb{P}_p(0 \longleftrightarrow \infty) \leq \mathbb{P}_p(I_0^c) = 1$$

donc  $\mathbb{P}_p(I_0) = 0$ . Il est donc acquis que  $\mathbb{P}_p$ -presque sûrement, il y a au moins un agrégat infini.

Ensuite, par ergodicité et par invariance de l'événement  $I_\infty$ , on sait que  $\mathbb{P}_p(I_\infty) \in \{0, 1\}$ . Si  $\mathbb{P}_p(I_\infty) = 1$ , le résultat est démontré. Supposons donc que  $\mathbb{P}_p(I_\infty) = 0$ . Autrement dit, supposons que  $\mathbb{P}_p$ -presque sûrement, il n'y ait qu'un nombre fini d'agrégats infinis. Il nous faut montrer que  $\mathbb{P}_p(I_1) = 1$ , et par ergodicité encore, il nous suffit pour cela de montrer que  $\mathbb{P}_p(I_1) > 0$ .

Définissons une suite croissante d'événements en posant, pour tout  $n \geq 0$ ,

$$R_n = \{\text{la boîte } B_n \text{ rencontre tous les agrégats infinis}\}.$$

Sur l'événement  $R_n$ , le nombre d'agrégats est fini (plus petit, par exemple, que le nombre de sommets de  $B_n$ , ou si l'on veut finasser, que le nombre de sommets de  $\partial B_n$ ) et toute configuration qui a un nombre fini d'agrégats infinis appartient à  $R_n$  pour  $n$  assez grand. Autrement dit,

$$1 = \mathbb{P}_p(I_\infty^c) = \mathbb{P}_p\left(\bigcup_{n \geq 0} R_n\right) = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}_p(R_n).$$

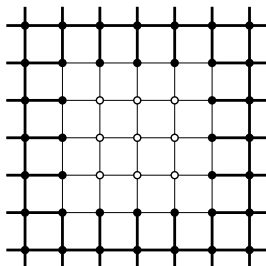
Il existe donc un entier  $n \geq 0$ , et nous en choisissons un, tel que  $\mathbb{P}_p(R_n) > 0$ .

Nous allons maintenant formaliser et exploiter l'idée qu'on peut décider si une configuration appartient à  $R_n$  sans regarder à l'intérieur de la boîte  $B_n$ . Ensuite, en ouvrant toutes les arêtes à l'intérieur de cette boîte, nous déduirons de nos hypothèses qu'il y a une probabilité strictement positive que tous les agrégats infinis de notre configuration soient connectés à l'intérieur de  $B_n$ , et ne soient donc qu'un seul et même agrégat, ce qui motrera que  $\mathbb{P}_p(I_1) > 0$ , et nous permettra de conclure.

Considérons donc le sous-graphe  $(V_n, E_n)$  défini par

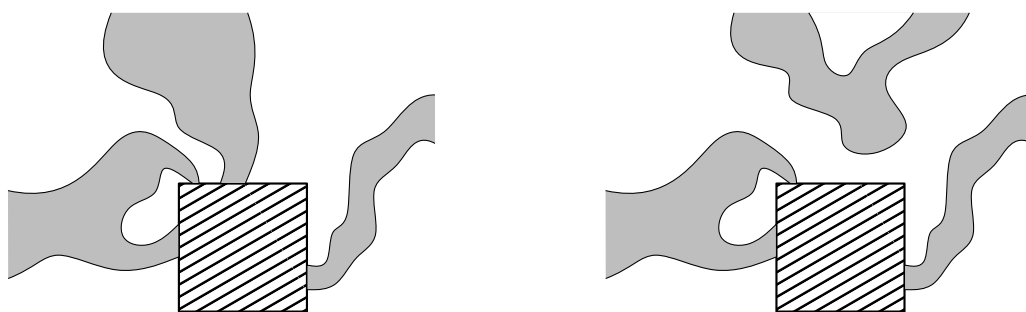
$$V_n = \mathbb{Z}^d \setminus B_{n-1} \text{ et } E_n = \mathbb{E}^d \setminus E(B_n).$$

Ce sous-graphe est représenté ci-dessous lorsque  $d = 2$  et  $n = 2$ .



Décomposons toute configuration  $\omega$  en une paire  $(\omega_{\text{int}}, \omega_{\text{ext}})$ , où  $\omega_{\text{int}}$  et  $\omega_{\text{ext}}$  sont les projections respectives de  $\omega$  sur les sous-graphes  $(B_n, E(B_n))$  et  $(V_n, E_n)$  de  $\mathbb{L}^d$ . Notons que chaque arête de  $\mathbb{L}^d$  appartient à un et un seul de ces deux sous-graphes, si bien que sous  $\mathbb{P}_p$ , les deux composantes d'une configuration de percolation sont indépendantes, de loi égale à celle de la percolation de paramètre  $p$  sur chacun de nos deux sous-graphes.

L'observation fondamentale est qu'une configuration appartient à l'événement  $R_n$  si et seulement si tous les agrégats infinis de sa restriction au sous-graphe  $(V_n, E_n)$  rencontrent  $\partial B_n$ .<sup>2</sup> Ci-dessous : à gauche, une configuration qui appartient à  $B_n$ , et à droite une configuration qui n'y appartient pas.



Notons  $S_n \subseteq \{0, 1\}^{E_n}$  l'événement qui concerne la percolation à l'extérieur de  $B_n$  et qui est défini par le fait que tous les agrégats infinis rencontrent  $\partial B_n$ . Notre discussion s'écrit, en symboles,

$$R_n = \{\omega_{\text{ext}} \in S_n\}$$

et nous permet d'affirmer que

$$\mathbb{P}_p(\omega_{\text{ext}} \in S_n) > 0.$$

Nous pouvons maintenant utiliser l'indépendance des deux projections d'une configuration pour écrire que

$$\mathbb{P}_p(\omega_{\text{ext}} \in S_n \text{ et toutes les arêtes de } \omega_{\text{int}} \text{ sont ouvertes}) = \mathbb{P}_p(\omega_{\text{ext}} \in S_n) \times p^{|E(B_n)|} > 0.$$

Or sur l'événement où  $\omega_{\text{ext}}$  appartient à  $S_n$  et toutes les arêtes de  $\omega_{\text{int}}$  sont ouvertes, tous les agrégats infinis de  $\omega_{\text{ext}}$  sont connectés par les arêtes ouvertes de  $\omega_{\text{int}}$ , et  $\omega$  n'a qu'un seul et unique agrégat infini. Ceci s'écrit

$$0 < \mathbb{P}_p(\omega_{\text{ext}} \in S_n \text{ et toutes les arêtes de } \omega_{\text{int}} \text{ sont ouvertes}) \leq \mathbb{P}_p(I_1)$$

et cette inégalité conclut la démonstration. □

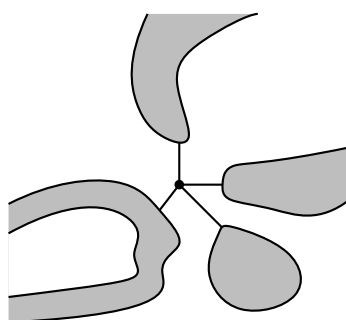
Il reste à éliminer la possibilité qu'il y ait une infinité d'agrégats infinis. La démonstration que nous allons donner utilise encore un argument géométrique, et repose sur la notion de *trifurcation*, que nous commençons par étudier.

Soient  $\omega$  une configuration et  $x$  un sommet. Notons  $\omega_x$  la configuration obtenue à partir de  $\omega$  en fermant toutes les arêtes issues de  $x$ . L'agrégat de  $x$  dans la configuration  $\omega$ , que nous

2. Notons que la restriction à  $(V_n, E_n)$  d'une configuration peut avoir strictement plus que d'agrégats infinis que la configuration initiale, si certains de ces agrégats sont connectés entre eux à l'intérieur de la boule  $B_n$ .

notons comme toujours  $C(x)$ , se scinde dans  $\omega_x$  en l'union du singleton  $\{x\}$  et d'un certain nombre (compris entre 0 et le nombre de voisins de  $x$  dans le graphe  $\mathbb{L}^d$ , en l'occurrence  $2d$ ) d'agrégats que nous appellerons les *agrégats scindés en  $x$*  de la configuration  $\omega$  (cette dénomination n'est pas standard). Ces agrégats scindés en  $x$  sont les classes d'équivalence, dans  $C(x)$ , de la relation "être égaux ou reliés par un chemin qui ne passe pas par  $x$ ". De manière équivalente, il est utile de garder en tête que deux sommets de  $C(x)$  sont dans des agrégats scindés en  $x$  distincts si et seulement si tout chemin ouvert entre ces deux sommets passe par  $x$ .

Parmi les agrégats scindés en  $x$ , certains sont finis et d'autres sont infinis. On dit que la configuration  $\omega$  a une *trifurcation* en  $x$ , ou que  $x$  est un point de trifurcation de  $\omega$ , si au moins trois des agrégats scindés en  $x$  sont infinis.



Nous allons avoir besoin du résultat suivant, purement déterministe, c'est-à-dire en l'occurrence géométrique, qui nous dit qu'une configuration ne peut avoir trop de points de trifurcation dans une boîte.

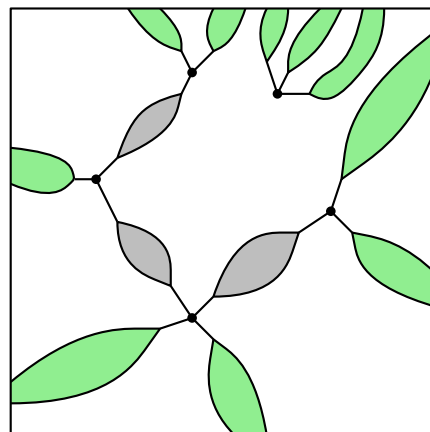
**Lemme 7.2.3.** Soit  $\omega$  une configuration de percolation sur  $\mathbb{L}^d$ . Soit  $n \geq 1$ . Alors

$$|\{x \in B_n : \omega \text{ a une trifurcation en } x\}| \leq 2|\partial B_n|.$$

Nous utiliserons de ce résultat la conséquence suivante : lorsque  $n$  tend vers l'infini, la proportion de points de la boîte  $B_n$  qui sont des points de trifurcation tend vers 0. L'idée qui sous-tend la démonstration de ce résultat est que les points de trifurcation forment les nœuds d'une sorte de réseau sous-jacent à la configuration, et que ce réseau est une forêt. Schématiquement, dans la boîte  $B_n$ , la configuration a l'aspect représenté à la figure ci-dessous.

Dans ce schéma, ne sont représentés que les points de trifurcation et les agrégats scindés infinis en ces points : il y a, en plus de ce qui est dessiné, des agrégats finis et des agrégats scindés finis en les points de trifurcation.

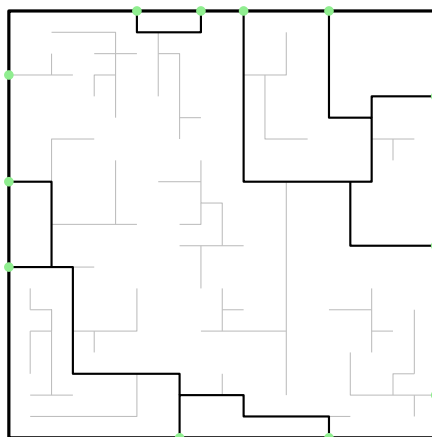
Le cœur de l'argument consiste à dire que nous voyons ici une forêt, que cette forêt a moins de feuilles (en vert) qu'il n'y a de points sur le bord de la boîte, et qu'une forêt a toujours plus de feuilles que de nœuds de degré supérieur ou égal à 3.



*Démonstration.* Considérons notre configuration  $\omega$  comme un sous-graphe de  $\mathbb{L}^d$ . Construisons un sous-graphe de  $\omega$  en remplaçant chacun de ses agrégats par un arbre couvrant de cet agrégat.<sup>3</sup> Nous obtenons de cette manière une nouvelle configuration, que nous notons  $\tau$ , qui vérifie  $\tau \leq \omega$ , et qui est une forêt, c'est-à-dire un graphe dont chaque composante connexe est un arbre.

Pour tout sommet  $x$  et tout agrégat scindé  $S$  en  $x$  (autre que  $\{x\}$ ), la forêt  $\tau$  contient un chemin de  $x$  à un point de  $S$ , et contient donc l'une des arêtes adjacentes à  $x$  qui le connectent à  $S$ . En particulier, et c'est ce qui compte pour nous, tous les points de trifurcation de  $\omega$  sont des sommets de degré supérieur ou égal à 3 de  $\tau$ .<sup>4</sup>

Fixons maintenant un entier  $n \geq 1$  et considérons la restriction de la forêt  $\tau$  à la boîte  $B_n$ , que nous notons  $\tau|_{B_n}$ . Il s'agit toujours d'une forêt, mais cette fois finie. Appelons *feuille* d'une forêt un sommet de degré 1. La forêt  $\tau|_{B_n}$  a des feuilles sur le bord de la boîte  $B_n$ , et des feuilles à l'intérieur de cette boîte, c'est-à-dire dans la boîte  $B_{n-1}$ . Nous allons rogner les feuilles de  $\tau|_{B_n}$  appartenant à  $B_{n-1}$ , c'est-à-dire fermer dans  $\tau|_{B_n}$  toutes les arêtes adjacentes à un sommet de degré 1 contenu dans  $B_{n-1}$ , et recommencer jusqu'à ce qu'il n'y en ait plus. Nous aboutissons donc à une forêt, que nous notons  $\sigma$ , dont toutes les feuilles sont situées sur le bord de la boîte.



Pour construire la forêt  $\sigma$ , nous avons enlevé des feuilles "parasites" de  $\tau|_{B_n}$ , mais il faut vérifier que nous n'avons pas, ce faisant, perdu de points de trifurcation de  $\omega$ . Pour nous en convaincre, observons d'une part que tout point de trifurcation de  $\omega$  contenu dans  $B_{n-1}$  est relié à  $\partial B_n$  par au moins trois chemins autoévitants deux à deux disjoints; et d'autre part que

3. Un arbre est un graphe connexe sans cycle. Tout graphe dénombrable et connexe admet un *arbre couvrant*, c'est-à-dire un sous-graphe qui contient tous ses sommets et qui est un arbre (il en admet en général beaucoup). On peut caractériser un arbre couvrant comme un sous-graphe qui est un arbre et qui est maximal pour l'inclusion avec cette propriété. On peut en construire un de la manière suivante. On énumère l'ensemble des sommets du graphe :  $V = \{x_1, x_2, \dots\}$  et on définit par récurrence une suite croissante de sous-graphes. On commence en posant  $T_1 = (\{x_1\}, \emptyset)$ . Ensuite,  $T_n$  étant défini, on considère le plus petit sommet du graphe qui ne soit pas dans  $T_n$  et on forme  $T_{n+1}$  en ajoutant à  $T_n$  un chemin de longueur minimale qui relie ce sommet à  $T_n$ . Tous les sous-graphes ainsi construits sont des arbres. Si le graphe est fini, ce processus se termine et le dernier arbre construit est un arbre couvrant. Sinon, l'union de la suite des arbres construits est un arbre couvrant.

4. La réciproque n'est pas vraie : notre forêt couvrante peut avoir beaucoup plus de sommets de degré supérieur ou égal à 3 qu'il n'y a de points de trifurcation de  $\omega$ . Pour obtenir une image correspondant exactement à la figure de la page précédente, il faudrait travailler un peu plus, mais ce n'est pas nécessaire pour ce que nous voulons faire.

si une arête de  $\tau_{|B_n}$  se trouve sur un chemin auto-évitant qui relie deux points de  $\partial B_n$ , alors elle ne peut être effacée au cours du processus de rognage. Ainsi, tous les points de trifurcation de  $\omega$  sont encore des sommets de degré supérieur ou égal à 3 de la forêt  $\sigma$ .

Nous utilisons maintenant le fait qu'un arbre (et donc une forêt) a moins de sommets de degré supérieur ou égal à 3 que de feuilles.<sup>5</sup>

Il découle de tout ce que nous avons dit que les points de trifurcation de  $\omega$  dans  $B_{n-1}$ , qui sont des sommets de  $\sigma$  de degré supérieur ou égal à 3, sont moins nombreux que les feuilles de  $\sigma$ , donc moins nombreux que les sommets de  $\partial B_n$ . En envisageant la possibilité, si peu plausible soit-elle, que tous les sommets de  $\partial B_n = B_n \setminus B_{n-1}$  soient des points de trifurcation, on arrive à l'inégalité cherchée.  $\square$

Ce résultat étant acquis, nous pouvons exclure la possibilité d'une infinité d'agrégats infinis dans le régime sur-critique.

**Proposition 7.2.4.** *Supposons  $\theta_d(p) > 0$ . Alors  $\mathbb{P}_p(I_1) = 1$  et  $\mathbb{P}_p(I_\infty) = 0$ .*

*Démonstration.* Pour chaque sommet  $x$ , notons  $T_x$  l'événement où  $x$  est un point de trifurcation. Par invariance par translations de la percolation, la probabilité  $\mathbb{P}_p(T_x)$  ne dépend pas de  $x$ .

Pour tout  $n \geq 0$ , notons  $N_n$  la variable aléatoire qui compte le nombre de points de trifurcation dans la boîte  $B_n$ . On a

$$\mathbb{E}_p[N_n] = \sum_{x \in B_n} \mathbb{P}_p(T_x) = |B_n| \mathbb{P}_p(T_0).$$

Par ailleurs, le lemme précédent nous assure que  $N_n \leq 2|\partial B_n|$ . On en déduit que

$$\mathbb{P}_p(T_0) \leq \frac{2|\partial B_n|}{|B_n|}$$

et en faisant tendre  $n$  vers l'infini, on voit que  $\mathbb{P}_p(T_0) = 0$ . Autrement dit, il n'y a,  $\mathbb{P}_p$ -presque sûrement, pas de trifurcation en 0, ni ailleurs.<sup>6</sup>

Nous allons maintenant montrer que si on avait  $\mathbb{P}_p(I_\infty) > 0$ , c'est-à-dire  $\mathbb{P}_p(I_\infty) = 1$ , alors l'origine aurait une probabilité strictement positive d'être une trifurcation.

Supposons donc  $\mathbb{P}_p(I_\infty) = 1$ . Alors, par un raisonnement analogue à celui que nous avons fait au début de la démonstration de la proposition 7.2.2, nous pouvons établir qu'il existerait un entier  $n$  tel que la boîte  $B_n$  ait une probabilité strictement positive de rencontrer au moins trois agrégats infinis.

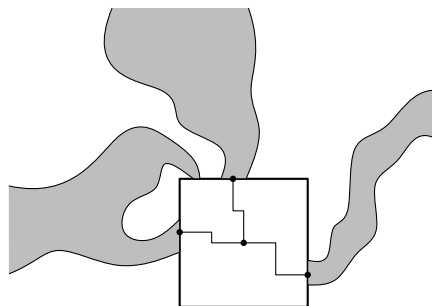
Pour exploiter ce fait, reprenons la notation  $\omega = (\omega_{\text{int}}, \omega_{\text{ext}})$  introduite dans cette démonstration. Sur l'événement où  $B_n$  rencontre au moins trois agrégats infinis, il y a dans la configuration  $\omega_{\text{ext}}$  au moins trois agrégats infinis qui rencontrent  $\partial B_n$ . Pour tous  $x, y, z \in \partial B_n$  deux à deux distincts, notons

$$J_{x,y,z} = \{\omega_{\text{ext}} \text{ a trois agrégats infinis distincts qui contiennent respectivement } x, y, z\}.$$

5. Pour être précis, si un arbre a  $f$  feuilles et  $s$  sommets de degré supérieur ou égal à 3, alors  $f \geq s + 2$ . Ce fait peut se démontrer par récurrence sur le nombre de sommets de l'arbre.

6. Dire que c'est pour en arriver à cette conclusion que nous les avons étudiées! Plus sérieusement, nous venons, assez discrètement, d'utiliser une propriété du groupe  $\mathbb{Z}^d$  qui est sa moyennabilité : l'infimum des  $|\partial F|/|F|$ , où  $F$  parcourt l'ensemble des parties finies de  $\mathbb{Z}^d$ , est nul.

Pour tous  $x, y, z \in \partial B_n$  encore, choisissons une configuration sur  $B_n$  dans laquelle toutes les arêtes sont fermées, sauf celles de trois chemins deux à deux disjoints reliant respectivement  $x$ ,  $y$  et  $z$  à 0. Notons  $\eta_{x,y,z}$  cette configuration.



Il suffit maintenant d'observer que dans la figure ci-dessus, il y a une trifurcation en 0, c'est-à-dire qu'on a l'inclusion d'événements

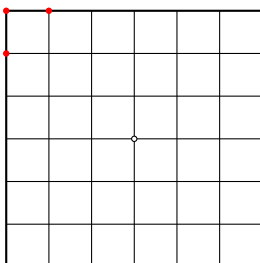
$$\bigcup_{\{x,y,z\} \subset \partial B_n} \{\omega_{\text{ext}} \in J_{x,y,z}\} \cap \{\omega_{\text{int}} = \eta_{x,y,z}\} \subseteq T_0.$$

L'union est ici prise sur toutes les parties à trois éléments de  $\partial B_n$ , qu'on étiquette de manière arbitraire  $\{x, y, z\}$ . Cette union est disjointe, parce que les configurations  $\eta_{x,y,z}$  sont distinctes, et on a donc

$$\mathbb{P}(T_0) \geq \min_{\{x,y,z\} \subset \partial B_n} \mathbb{P}_p(\omega_{\text{int}} = \eta_{x,y,z}) \sum_{\{x,y,z\} \subset \partial B_n} \mathbb{P}_p(\omega_{\text{ext}} \in J_{x,y,z}).$$

Le minimum est strictement positif, et la somme est plus grande que la probabilité, strictement positive aussi, que la boîte  $B_n$  rencontre au moins trois agrégats infinis. Ainsi, nous arrivons à la conclusion que  $\mathbb{P}(T_0) > 0$ , ce qui est contradictoire. Notre hypothèse selon laquelle  $\mathbb{P}_p(I_\infty) > 0$  était donc fausse.  $\square$

J'ai un peu triché dans la preuve précédente : il n'est pas vrai que le minimum dans le membre de droite de la dernière inégalité soit strictement positif, car il est possible, en dimension 2, de placer trois points sur le bord d'un carré de telle manière qu'on ne puisse les joindre à l'origine par des chemins disjoints qui restent à l'intérieur de ce carré :



En dimension supérieure ou égale à 3, je pense que ce problème ne se pose pas. Pour le résoudre en dimension 2, on peut faire l'une des deux choses suivantes : soit observer que si la boîte  $B_n$  rencontre trois agrégats infinis, et même si l'on est dans la situation défavorable



ci-dessus, alors la boîte  $B_{n+1}$  rencontre encore trois agrégats infinis, en trois points qu'il est possible de relier à l'origine par des chemins disjoints; soit choisir l'entier  $n$  tel que la boîte  $B_n$  ait une probabilité strictement positive de rencontrer au moins *quatre* agrégats infinis, et observer que parmi quatre points donnés sur le bord de la boîte  $B_n$ , et quelle que soit la dimension, il est possible d'en choisir trois qui peuvent être reliés à l'origine par des chemins deux à deux disjoints.

### 7.3 Régularité de $\theta_d$

Le fait maintenant établi qu'il y a, dans la phase sur-critique, un unique agrégat infini, va nous permettre d'achever la démonstration de la deuxième assertion du théorème 6.1.5.

**Proposition 7.3.1.** *La fonction  $\theta_d$  est continue sur l'intervalle  $]p_c(d), 1]$ .*

*Démonstration.* Donnons-nous, dans l'esprit de la démonstration de la proposition 6.2.3, et sur un espace de probabilité  $(\Omega, \mathcal{A}, \mathbf{P})$ , une famille  $(U_e)_{e \in \mathbb{E}^d}$  de variables aléatoires i.i.d. uniformes sur  $[0, 1]$ . Pour tout  $p > p_c$ , notons  $\omega^p$  la configuration donnée par  $\omega_e^p = \mathbb{1}_{\{U_e \leq p\}}$ . Elle est sous  $\mathbf{P}$  de loi  $\mathbb{P}_p$ , et puisque  $p > p_c$ , elle admet, d'après les résultats de la section précédente, un unique agrégat infini que nous notons  $C(p)$ . On a en particulier

$$\theta_d(p) = \mathbf{P}(0 \in C(p)).$$

Si  $p \leq p_c$ , nous notons  $C(p) = \emptyset$ . Notons que par construction, l'application  $p \mapsto C(p)$  est croissante.

Fixons  $p > p_c$ . Alors on a

$$\lim_{n \rightarrow \infty} \theta_d(p - \frac{1}{n}) = \lim_{n \rightarrow \infty} \mathbf{P}(0 \in C(p - \frac{1}{n})) = \mathbf{P}(\bigcup_{n \geq 1} \{0 \in C(p - \frac{1}{n})\}) = \mathbf{P}(0 \in \bigcup_{n \geq 1} C(p - \frac{1}{n})).$$

Notons  $\theta_d(p^-)$  la limite à gauche en  $p$  de la fonction  $\theta_d$ , et posons

$$C(p^-) = \bigcup_{q < p} C(q).$$

Alors nous avons établi que

$$\theta_d(p^-) = \mathbf{P}(0 \in C(p^-)).$$

En particulier, pour montrer que  $\theta_d$  est continue à gauche en  $p$ , il nous faut (et il nous suffit de) montrer que  $\mathbf{P}(0 \in C(p) \setminus C(p^-)) = 0$ , car alors nous aurons

$$\theta_d(p) - \theta_d(p^-) = \mathbf{P}(0 \in C(p)) - \mathbf{P}(0 \in C(p^-)) = \mathbf{P}(0 \in C(p) \setminus C(p^-)) = 0.$$

Nous allons en fait vérifier que  $\mathbf{P}$ -presque sûrement,  $C(p^-) = C(p)$ . En effet, pour que ces deux agrégats diffèrent, il faut qu'au moins une arête soit fermée dans toutes les configurations  $\omega^q$  pour  $q < p$ , et ouverte dans  $\omega^p$ . Pour qu'une arête  $e$  ait cette propriété, il faut avoir  $U_e = p$ , ce qui arrive avec probabilité nulle. En symboles,

$$\mathbf{P}(C(p^-) \neq C(p)) \leq \mathbf{P}(\exists e \in \mathbb{E}^d : U_e = p) \leq \sum_{e \in \mathbb{E}^d} \mathbf{P}(U_e = p) = 0,$$

et la preuve est complète.  $\square$

## 7.4 Deux inégalités classiques

Notre prochain objectif est de démontrer la troisième assertion du théorème 6.1.5, qui concerne la taille de l'agrégat de l'origine dans le régime sous-critique. Par définition du régime sous-critique, nous savons que cet agrégat est fini presque sûrement. Montrer que la taille de cet agrégat est d'espérance finie est moins facile que l'on ne pourrait s'y attendre. Nous allons montrer mieux, à savoir que la queue de la distribution de sa taille décroît exponentiellement.

Avant de nous lancer dans cette étude, nous allons démontrer deux inégalités très utiles, et très jolies.

### 7.4.1 L'inégalité FKG

Rappelons que l'espace des configurations de percolation est ordonné, par l'inclusion, et que nous avons défini la notion d'événements croissants, qui sont les événements dont la fonction indicatrice est croissante : de tels événements sont "d'autant plus facilement réalisés qu'il y a d'arêtes ouvertes". L'inégalité FKG (des noms de Fortuin, Kasteleyn et Ginibre) exprime que deux événements croissants sont positivement corrélés : si l'un est réalisé, alors on peut penser que beaucoup d'arêtes sont ouvertes, et cela favorise l'autre.

L'inégalité FKG est plutôt une famille d'inégalités, qui s'appliquent dans différents contextes de physique statistique, et dans le cas de la percolation, elle est due à Harris. Elle est valable sur n'importe quel graphe (dénombrable)  $G = (V, E)$ .

**Proposition 7.4.1** (Inégalité FKG – Inégalité de Harris). *Soient  $f$  et  $g$  deux fonctions mesurables bornées croissantes sur l'espace  $\{0, 1\}^E$ . Pour tout  $p \in [0, 1]$ , on a*

$$\mathbb{E}_p[fg] \geq \mathbb{E}_p[f]\mathbb{E}_p[g].$$

En particulier, si  $A$  et  $B$  sont des événements croissants, alors

$$\mathbb{P}_p(A \cap B) \geq \mathbb{P}_p(A)\mathbb{P}_p(B).$$

*Démonstration.* Énumérons les arêtes de notre graphe :  $E = \{e_1, e_2, \dots\}$  et pour tout  $n \geq 1$ , notons  $\mathcal{F}_n$  la tribu engendrée sur  $\{0, 1\}^E$  par les applications  $(\omega \mapsto \omega_{e_i} : i \in \{1, \dots, n\})$ .

En utilisant un argument de conditionnement et d'approximation analogue à celui que nous avons utilisé pour démontrer le théorème 7.1.2, nous voyons qu'il suffit de démontrer l'inégalité lorsque  $f$  et  $g$  sont mesurables par rapport à la tribu  $\mathcal{F}_n$  pour un certain  $n \geq 1$ , c'est-à-dire lorsque les valeurs de  $f$  et  $g$  sur une configuration ne dépendent que de l'état, dans cette configuration, des arêtes  $e_1, \dots, e_n$ . C'est ce que nous allons faire, par récurrence sur  $n$ . Par commodité, nous écrivons  $\omega_i = \omega_{e_i}$ , pour toute configuration  $\omega$  et tout  $i \geq 1$ .

Supposons donc que  $f$  et  $g$  ne dépendent que de  $\omega_1$ . Puisque l'inégalité à démontrer n'est pas affectée par l'ajout d'une constante à  $f$  ou à  $g$ , nous pouvons supposer que  $f(0) = g(0) = 0$ . Alors puisque  $f$  et  $g$  sont croissantes, on a  $f(1) \geq 0$  et  $g(1) \geq 0$ , et

$$\mathbb{E}_p[fg] = pf(1)g(1) \geq p^2f(1)g(1) = \mathbb{E}_p[f]\mathbb{E}_p[g].$$

Considérons maintenant  $n \geq 2$  et supposons l'inégalité démontrée pour toute paire de fonctions mesurables par rapport à  $\mathcal{F}_{n-1}$ . Soient  $f$  et  $g$  mesurables par rapport à  $\mathcal{F}_n$ . Pour toute

configuration  $\omega$ , notons  $\omega_{<n} = (\omega_1, \dots, \omega_{n-1})$  et autorisons-nous à écrire par exemple  $f(\omega) = f(\omega_{<n}, \omega_n)$ .

Pour toute fonction  $h$  bornée mesurable par rapport à  $\mathcal{F}_n$ , on a

$$\mathbb{E}_p[h|\mathcal{F}_{n-1}] = \sum_{\eta \in \{0,1\}^{n-1}} \mathbb{1}_{\{\omega_{<n}=\eta\}} (ph(\eta, 1) + (1-p)h(\eta, 0)) = \sum_{\eta \in \{0,1\}^{n-1}} \mathbb{1}_{\{\omega_{<n}=\eta\}} \mathbb{E}_p[h(\eta, \cdot)],$$

où la dernière espérance ne porte que sur la variables  $\omega_n$ . En appliquant cette égalité à  $h = fg$  et en utilisant l'inégalité FKG pour les fonctions qui ne dépendent que d'une arête, on trouve

$$\mathbb{E}[fg|\mathcal{F}_{n-1}] \geq \sum_{\eta \in \{0,1\}^{n-1}} \mathbb{1}_{\{\omega_{<n}=\eta\}} \mathbb{E}_p[f(\eta, \cdot)] \mathbb{E}_p[g(\eta, \cdot)] = \mathbb{E}_p[f|\mathcal{F}_{n-1}] \mathbb{E}_p[g|\mathcal{F}_{n-1}].$$

En appliquant maintenant l'inégalité FKG pour les fonctions  $\mathcal{F}_{n-1}$ -mesurables, on trouve

$$\begin{aligned} \mathbb{E}_p[fg] &= \mathbb{E}_p[\mathbb{E}_p[fg|\mathcal{F}_{n-1}]] \\ &\geq \mathbb{E}_p[\mathbb{E}_p[f|\mathcal{F}_{n-1}] \mathbb{E}_p[g|\mathcal{F}_{n-1}]] \\ &\geq \mathbb{E}_p[\mathbb{E}_p[f|\mathcal{F}_{n-1}]] \mathbb{E}_p[\mathbb{E}_p[g|\mathcal{F}_{n-1}]] \\ &= \mathbb{E}_p[f] \mathbb{E}_p[g] \end{aligned}$$

et l'inégalité est démontrée. □

### 7.4.2 L'inégalité BK

L'inégalité BK (des noms de van den Berg et Kesten) va, en quelque sorte, dans le sens opposé à l'inégalité FKG. Il en existe différentes versions, et celle que nous allons énoncer concerne des événements croissants pour la percolation sur un graphe fini (mais nous pourrions l'appliquer à un sous-graphe fini de notre graphe favori  $\mathbb{L}^d$ ).

L'inégalité BK fait intervenir une notion nouvelle, qui est l'*occurrence disjointe* de deux événements. Soit  $G = (V, E)$  un graphe fini. Soit  $\omega$  une configuration de percolation sur  $G$ . Pour toute partie  $F$  de  $E$ , on note  $\omega^F$  la configuration obtenue en partant de  $\omega$  et en fermant toutes les arêtes de  $E \setminus F$ . Autrement dit, pour tout  $e \in E$ ,

$$\omega_e^F = \omega_e \mathbb{1}_F(e).$$

Soit  $A$  un événement pour la percolation sur  $G$ . Soit  $\omega$  une configuration, et  $F$  une partie de  $E$ .

Soient maintenant  $A$  et  $B$  deux événements croissants. On note

$$A \circ B = \{\omega \in \{0,1\}^E : \exists F \subseteq E, \omega^F \in A \text{ et } \omega^{E \setminus F} \in B\}$$

et on dit que  $A \circ B$  est l'événement où  $A$  et  $B$  sont réalisés de manière disjointe.

En français,  $\omega$  appartient à  $A \circ B$  si et seulement si on peut trouver dans les arêtes ouvertes de  $\omega$  un sous-ensemble dont le fait qu'il soit formé d'arêtes ouvertes suffit à assurer que  $\omega$  réalise  $A$ , et un autre sous-ensemble d'arêtes ouvertes dans  $\omega$ , disjoint du premier, dont le fait qu'il soit formé d'arêtes ouvertes suffit à assurer que  $\omega$  réalise  $B$ . On a en particulier l'inclusion  $A \circ B \subseteq A \cap B$ .

L'exemple classique est celui de  $\{w \longleftrightarrow x\} \circ \{y \longleftrightarrow z\}$  qui est l'événement où il existe deux chemins ouverts disjoints, l'un de  $w$  à  $x$  et l'autre de  $y$  à  $z$ .

Il est plausible que si  $A$  et  $B$  sont croissants, alors conditionnellement au fait que  $A$  est réalisé, l'événement  $A \circ B$  est moins probable que  $B$  : car pour réaliser  $A \circ B$ , on ne peut pas utiliser les arêtes utiles à la réalisation de  $A$ .<sup>7</sup> Autrement dit, on peut s'attendre à ce que  $\mathbb{P}_p(A \circ B|A) \leq \mathbb{P}_p(B)$ . C'est ce qu'affirme l'inégalité BK.

**Proposition 7.4.2** (Inégalité BK). *Soit  $G = (V, E)$  un graphe fini. Soient  $A$  et  $B$  deux événements croissants de percolation sur  $G$ . Soit  $p \in [0, 1]$ . Alors on a*

$$\mathbb{P}_p(A \circ B) \leq \mathbb{P}_p(A)\mathbb{P}_p(B).$$

La démonstration de l'inégalité BK est élémentaire, mais pas très courte, et nécessite d'introduire un peu de notation. Pour l'instant, nous ne l'incluons pas dans ces notes.

## 7.5 Petitesse exponentielle des agrégats sous-critiques

Dans la phase sous-critique, c'est-à-dire pour  $p < p_c$ , nous savons (c'est la définition de cette phase) que l'agrégat de l'origine est fini presque sûrement. Il semble que la démonstration du fait que sa taille est *intégrable* ait été un jalon important du développement de la théorie. Nous allons montrer, au prix d'un certain travail, le résultat plus fort suivant.

**Proposition 7.5.1.** *Supposons  $p < p_c(d)$ . Alors il existe un réel strictement positif  $\alpha > 0$ , qui dépend de  $p$ , tel qu'on ait pour tout  $n \geq 1$  l'inégalité*

$$\mathbb{P}_p(0 \longleftrightarrow \partial B_n) \leq e^{-\alpha n}.$$

Ce résultat décrit la queue de distribution du diamètre de l'agrégat de 0, c'est-à-dire de la distance maximale entre deux points de cet agrégat. En effet, en mesurant les distances en norme  $\|\cdot\|_\infty$ , on a les inclusions

$$\{\text{diam}(C(0)) > 2n - 2\} \subseteq \{C(0) \longleftrightarrow \partial B_n\} \subseteq \{\text{diam}(C(0)) \geq n\},$$

si bien qu'on a, pour un autre réel  $\alpha' > 0$  et pour tout entier  $n \geq 1$ ,

$$\mathbb{P}_p(\text{diam}(C(0)) \geq n) \leq e^{-\alpha' n}.$$

On peut aussi déduire de ce résultat une majoration de la queue de distribution du nombre de sommets de  $C(0)$ . En effet, on a l'inclusion

$$\{|C(0)| > (2n - 1)^d\} \subseteq \{C(0) \longleftrightarrow \partial B_n\},$$

si bien qu'il existe un réel  $\alpha'' > 0$  (le réel  $\alpha/2$  convient) tel qu'on ait pour tout  $n$  l'inégalité

$$\mathbb{P}_p(|C(0)| \geq n) \leq e^{-\alpha'' n^{\frac{1}{d}}}.$$

7. Attention, ce genre de phrases au sens vague peut créer plus de confusion que de clarté. Il n'y a par exemple pas, en général, pour un événement  $A$  donné, un ensemble particulier d'arêtes qui doivent absolument être ouvertes pour que  $A$  soit réalisé.

Cette inégalité entraîne l'intégrabilité de  $|C(0)|$ , mais une inégalité plus forte est vraie : il est possible de démontrer qu'il existe une constante  $\alpha > 0$  (une autre que dans notre énoncé) telle qu'on ait pour tout  $n$  l'inégalité

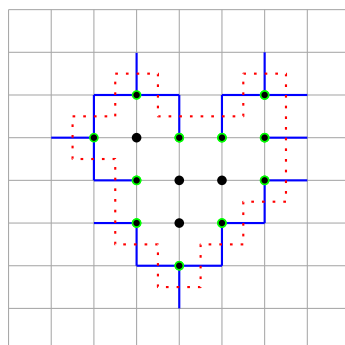
$$\mathbb{P}_p(|C(0)| \geq n) \leq e^{-\alpha n}.$$

Nous ne démontrerons pas cette inégalité, je laisse les plus curieuses et curieux aller en chercher la démonstration dans la littérature.

La démonstration de la proposition 7.5.1 n'est pas tout à fait simple. Celle que nous allons présenter se fait en plusieurs étapes et repose sur une quantité un peu mystérieuse au premier abord, que nous définirons dans un instant. Rappelons tout d'abord que nous avons défini, pour toute partie  $A$  de  $\mathbb{Z}^d$ , le bord de  $A$  comme l'ensemble  $\partial A = \{x \in A : \exists y \notin A, \{x, y\} \in \mathbb{E}^d\}$ . Nous définissons une autre notion de bord en posant

$$\delta A = \{\{x, y\} \in E : x \in A, y \notin A\}.$$

Alors que  $\partial A$  est un ensemble de sommets,  $\delta A$  est un ensemble d'arêtes. Dans la figure ci-dessous, les sommets de  $A$  sont représentés en noir, ceux de  $\partial A$  sont encadrés de vert, et les arêtes de  $\delta A$  sont bleues. Le circuit dual rouge est représenté à des fins décoratives.



Pour tout ensemble  $A$  de sommets et pour tous sommets  $x, y$ , on note  $\{x \xrightarrow{A} y\}$  l'événement où  $x$  et  $y$  sont connectés par un chemin ouvert qui ne passe que par des sommets de  $A$ . On définit, pour tout ensemble fini de sommets  $A$ , et tout  $p \in [0, 1]$ ,

$$\varphi_p(A) = p \sum_{\substack{e \in \delta A \\ e = \{x, y\}, x \in A}} \mathbb{P}_p(0 \xrightarrow{A} x).$$

La somme a exactement un terme pour chaque arête de  $\delta A$ , que nous notons  $e = \{x, y\}$ , où  $x$  est le sommet de l'arête qui appartient à  $A$ , et  $y$  celui qui n'y appartient pas. En considérant l'arête  $e$  comme un ensemble de deux sommets (ce qu'elle est par définition), on pourrait donc

écrire<sup>8</sup>

$$\varphi_p(A) = p \sum_{e \in \delta A} \mathbb{P}_p(0 \xleftarrow{A} e).$$

Remarquons dès maintenant que si l'origine de  $\mathbb{Z}^d$  n'appartient pas à  $A$ , alors  $\varphi_p(A) = 0$ .

La signification de la quantité  $\varphi_p(A)$  est encore assez obscure à ce stade. Son utilité dans la démonstration de la proposition 7.5.1 est résumée par les énoncés qui suivent.

**Lemme 7.5.2.** *Soit  $p \in [0, 1]$ . S'il existe une partie finie  $A$  de  $\mathbb{Z}^d$  contenant l'origine et telle que  $\varphi_p(A) < 1$ , alors la conclusion de la proposition 7.5.1 est vraie.*

Pour démontrer la proposition 7.5.1, il suffit donc de montrer que pour  $p < p_c$ , il existe une telle partie  $A$ . Notons que pour toute partie  $A$ , la fonction  $p \mapsto \varphi_p(A)$  est croissante. Posons

$$\tilde{p}_c = \sup\{p \in [0, 1] : \exists A \subset \mathbb{Z}^d, A \text{ finie}, 0 \in A, \text{ et } \varphi_p(A) < 1\}.$$

Il suffit donc de montrer que  $p_c \leq \tilde{p}_c$ , de manière à ce que l'hypothèse  $p < p_c$  entraîne  $p < \tilde{p}_c$ . Pour montrer cette inégalité, nous allons démontrer le lemme suivant.

**Lemme 7.5.3.** *Soit  $p \in [0, 1]$ . Si  $p > \tilde{p}_c$ , alors  $\theta_d(p) > 0$ .*

En particulier, si  $\tilde{p}_c < p$ , alors  $p_c \leq p$ . Il s'ensuit que  $p_c \leq \tilde{p}_c$ , et la proposition 7.5.1 est démontrée.

Avant de nous lancer dans la démonstration des deux lemmes, faisons deux remarques. Tout d'abord, une fois les lemmes démontrés, nous saurons que  $\tilde{p}_c = p_c$ . En effet, nous venons de dire que le second entraînait l'inégalité  $p_c \leq \tilde{p}_c$ . Par ailleurs, si  $p < \tilde{p}_c$ , alors d'après le lemme 7.5.2, la conclusion de la proposition 7.5.1 est vraie, et l'agrégat de 0 est fini  $\mathbb{P}_p$ -presque sûrement, donc  $p \leq p_c$ . L'implication  $p < \tilde{p}_c \Rightarrow p \leq p_c$  est donc vraie, et on en déduit que  $\tilde{p}_c \leq p_c$ .

Ensuite, soulignons que l'hypothèse de la proposition 7.5.1 est bien  $p < p_c$ , et non  $\theta_d(p) = 0$ . En dimension  $d = 2$ , les agrégats de la phase critique (c'est-à-dire pour  $p = p_c(2)$ ) sont beaucoup plus gros que des agrégats sous-critiques, et la queue de la distribution de leur diamètre ne décroît pas exponentiellement, bien que la probabilité de percolation soit nulle au point critique (dans le cas  $d = 2$ ), comme nous le verrons.

**Jeu** — Il découle de ce qui précède (et de ce qui suivra) que pour tout  $p < \frac{1}{2}$ , il existe une partie finie  $A$  de  $\mathbb{Z}^2$  contenant l'origine et telle que  $\varphi_p(A) < 1$ . Le jeu consiste à trouver, pour un  $p$  le plus grand possible, une telle partie  $A$ . Par exemple, pour  $p < \frac{1}{4}$ , la partie  $A = \{0\}$  convient, puisque  $\varphi_p(A) = 4p$ . La partie  $B$  constituée de l'origine et de ses quatre plus proches voisins vérifie  $\varphi_p(B) = 12p^2$  et convient donc pour  $p < \frac{1}{2\sqrt{3}} \simeq 0,288$ .

Venons-en maintenant à la démonstration du lemme 7.5.2, démonstration dans laquelle nous utiliserons une fois l'inégalité BK.

8. Bien que cette forme ne nous soit pas utile, mais pour essayer de mieux comprendre la définition, notons que  $\varphi_p(A)$  peut aussi s'écrire comme une somme sur l'ensemble de sommets  $\partial A$ , pourvu qu'on pondère chaque sommet par le nombre d'arêtes de  $\delta A$  qui en est issu :

$$\varphi_p(A) = p \sum_{x \in \partial A} |\{e \in \delta A : x \in e\}| \mathbb{P}_p(0 \xleftarrow{A} x).$$

*Démonstration du lemme 7.5.2.* Soit  $A$  une partie finie contenant l'origine telle que  $\varphi_p(A) < 1$ . Soit  $N$  un entier suffisamment grand pour que  $A$  soit incluse dans  $B_{N-1}$ . Nous allons démontrer que pour tout entier  $k \geq 2$ , on a

$$\mathbb{P}_p(0 \longleftrightarrow \partial B_{kN}) \leq \varphi_p(A) \mathbb{P}_p(0 \longleftrightarrow \partial B_{(k-1)N}).$$

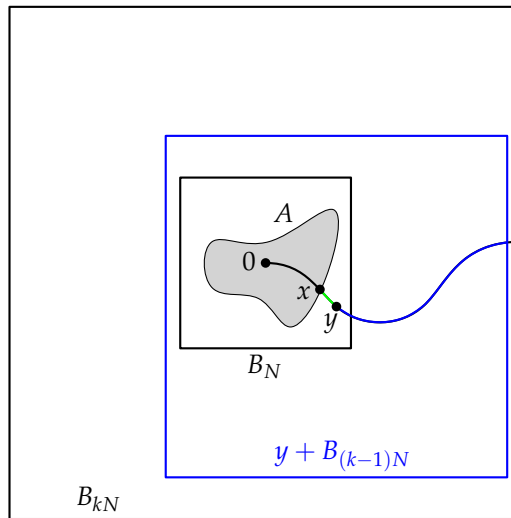
Pour cela, considérons une configuration  $\omega$  dans l'événement  $\{0 \longleftrightarrow \partial B_{kN}\}$ . Alors il existe un chemin ouvert de  $0$  à  $\partial B_{kN}$ , et même un chemin ouvert auto-évitant, c'est-à-dire qui ne passe plus d'une fois en aucun sommet : en effet, un chemin de longueur minimale parmi les chemins ouverts de  $0$  à  $\partial B_{kN}$  est nécessairement auto-évitant, sinon, on pourrait le raccourcir en lui enlevant la portion qu'il parcourt entre deux passages au même sommet.

Parcourons ce chemin depuis l'origine vers le bord de la boîte  $B_{kN}$  et appelons  $y$  le premier sommet que nous rencontrons qui ne soit pas dans  $A$ . Appelons  $x$  le sommet visité immédiatement avant  $y$ , et remarquons que l'arête  $\{x, y\}$  appartient à  $\delta A$ . Alors dans notre configuration,  $0$  est relié à  $x$  par un chemin ouvert qui ne visite que des sommets de  $A$ , l'arête  $\{x, y\}$  est ouverte, et  $y$  est relié à  $\partial B_{kN}$  par un chemin ouvert. De plus, puisque notre chemin est auto-évitant, les événements  $\{0 \xrightarrow{A} x\} \cap \{\omega_{\{x,y\}} = 1\}$  et  $\{y \longleftrightarrow \partial B_{kN}\}$  sont réalisés de manière disjointe au sens de la section 7.4.2. Enfin, l'événement  $\{0 \xrightarrow{A} x\}$  est mesurable par rapport à la tribu engendrée par les arêtes reliant deux sommets de  $A$ , et donc indépendante de  $\omega_{\{x,y\}}$ . Finalement, nous venons d'établir l'inclusion

$$\{0 \longleftrightarrow \partial B_{kN}\} \subseteq \bigcup_{\substack{e \in \delta A \\ e = \{x,y\}, x \in A}} (\{0 \xrightarrow{A} x\} \cap \{\omega_{\{x,y\}} = 1\}) \circ \{y \longleftrightarrow \partial B_{kN}\}$$

dont nous tirons, en intégrant par rapport à  $\mathbb{P}_p$ , en utilisant l'inégalité BK et l'indépendance que nous venons d'observer, l'inégalité

$$\mathbb{P}_p(0 \longleftrightarrow \partial B_{kN}) \leq p \sum_{\substack{e \in \delta A \\ e = \{x,y\}, x \in A}} \mathbb{P}_p(0 \xrightarrow{A} x) \mathbb{P}_p(y \longleftrightarrow \partial B_{kN}).$$



Puisque  $A$  est inclus dans  $B_{N-1}$ , le sommet  $y$  appartient à  $B_N$ , si bien que la boîte translaturée  $y + B_{(k-1)N}$  est incluse dans  $B_{kN}$ . En particulier,

$$\{y \longleftrightarrow \partial B_{kN}\} \subseteq \{y \longleftrightarrow \partial(y + B_{(k-1)N})\},$$

donc, en utilisant l'invariance par translation du modèle,

$$\mathbb{P}_p(y \longleftrightarrow \partial B_N) \leq \mathbb{P}_p(y \longleftrightarrow \partial(y + B_{(k-1)N})) = \mathbb{P}_p(0 \longleftrightarrow \partial B_{(k-1)N}).$$

En remplaçant dans l'inégalité précédente et en comparant avec la définition de  $\varphi_p(A)$ , on obtient la relation annoncée :

$$\mathbb{P}_p(0 \longleftrightarrow \partial B_{kN}) \leq \varphi_p(A) \mathbb{P}_p(0 \longleftrightarrow \partial B_{(k-1)N}).$$

Observons que le raisonnement reste valide pour  $k = 1$ , si bien que  $\mathbb{P}_p(0 \longleftrightarrow \partial B_N) \leq \varphi_p(A)$  et que pour tout  $k \geq 0$ ,

$$\mathbb{P}_p(0 \longleftrightarrow \partial B_{kN}) \leq \varphi_p(A)^k.$$

La conclusion de la proposition 7.5.1 s'en déduit de manière élémentaire.  $\square$

**Exercice 7.1.** Soit  $(a_n)_{n \geq 1}$  une suite décroissante d'éléments de  $[0, 1[$ . On suppose qu'il existe un entier  $N \geq 1$  et un réel  $\varphi \in ]0, 1[$  tels que pour tout  $k \geq 1$ , on ait  $a_{kN} \leq \varphi^k$ . Montrer qu'il existe  $\alpha > 0$  tel que pour tout  $n \geq 1$  on ait  $a_n \leq e^{-\alpha n}$ .

Il nous reste maintenant démontrer le lemme 7.5.3.

*Démonstration du lemme 7.5.3.* L'entier  $d \geq 2$  étant fixé, posons, pour tout  $n \geq 1$  et tout  $p \in [0, 1]$ ,

$$\theta_n(p) = \mathbb{P}_p(0 \longleftrightarrow \partial B_n).$$

La première partie de la preuve consiste à calculer la dérivée de la fonction  $\theta_n$ . Pour cela, nous allons utiliser le couplage croissant des mesures  $\mathbb{P}_p$ ,  $p \in [0, 1]$ , comme nous l'avons fait dans la démonstration de la proposition 7.3.1 — et avec les mêmes notations.

Donnons-nous  $p \in ]0, 1[$  et  $\varepsilon > 0$ . Nous avons

$$\theta_n(p + \varepsilon) - \theta_n(p) = \mathbf{E}[\mathbb{1}_{\{0 \longleftrightarrow \partial B_n\}}(\omega^{p+\varepsilon}) - \mathbb{1}_{\{0 \longleftrightarrow \partial B_n\}}(\omega^p)].$$

Identifions les configurations à des parties de l'ensemble  $E(B_n)$  des arêtes contenues dans la boîte  $B_n$ . La fonction dont nous calculons l'espérance vaut 0, sauf si l'origine est reliée à  $\partial B_n$  dans la configuration  $\omega^{p+\varepsilon}$  mais pas dans  $\omega^p$ . Ainsi, en écrivant les choses très explicitement,

$$\theta_n(p + \varepsilon) - \theta_n(p) = \sum_{\substack{F, G \subseteq E(B_n), F \cap G = \emptyset \\ F \notin \{0 \longleftrightarrow \partial B_n\}, F \cup G \in \{0 \longleftrightarrow \partial B_n\}}} \mathbf{P}(\omega^p = F, \omega^{p+\varepsilon} = F \cup G).$$

Les parties  $F$  et  $G$  étant fixées, l'événement dont nous calculons la probabilité se réalise si et seulement si pour toute arête  $e \in E(B_n)$ , on a  $U_e \leq p$  si  $e \in F$ ,  $U_e \in ]p, p + \varepsilon]$  si  $e \in G$ , et  $U_e > p + \varepsilon$  sinon. Ainsi,

$$\theta_n(p + \varepsilon) - \theta_n(p) = \sum_{F, G} p^{|F|} \varepsilon^{|G|} (1 - p - \varepsilon)^{|E(B_n)| - |F| - |G|},$$



la somme étant exactement sur les mêmes couples  $(F, G)$  que la précédente. Dans la somme, seuls des termes pour lesquels l'ensemble  $G$  est non vide peuvent contribuer, et il n'y a donc pas de terme d'ordre 0 en  $\varepsilon$ . Au premier ordre en  $\varepsilon$  ne contribuent que les termes où  $G$  contient une unique arête :

$$\theta_n(p + \varepsilon) - \theta_n(p) = \frac{\varepsilon}{1-p} \sum_{F,e} p^{|F|} (1-p)^{|E(B_n)|-|F|} + O(\varepsilon^2).$$

Pour être tout à fait explicite, la somme porte ici sur les parties  $F$  de  $E(B_n)$  et les arêtes  $e \in E(B_n) \setminus F$  telles que  $F \notin \{0 \longleftrightarrow \partial B_n\}$  et  $F \cup \{e\} \in \{0 \longleftrightarrow \partial B_n\}$ . De plus, la contribution à la somme du couple  $(F, e)$  est simplement la probabilité sous  $\mathbb{P}_p$  de la configuration  $F$ .

Pour toute arête  $e$ , notons  $\{0 \leftarrow e \rightarrow \partial B_n\}$  l'événement où 0 n'est pas connecté à  $\partial B_n$  mais le devient si l'on ouvre  $e$ . Du calcul que nous venons de faire, il découle donc que

$$\theta'_n(p) = \frac{1}{1-p} \sum_{e \in E(B_n)} \mathbb{P}_p(0 \leftarrow e \rightarrow \partial B_n).$$

Introduisons l'ensemble aléatoire

$$D_n = D_n(\omega) = \{x \in B_n : x \not\leftrightarrow \partial B_n\}$$

des sommets de  $B_n$  qui sont disconnectés du bord de la boîte. L'observation importante est la suivante : pour toute arête  $e$ , une configuration est dans l'événement  $\{0 \leftarrow e \rightarrow \partial B_n\}$  si et seulement si 0 est connecté à  $e$  dans  $D_n$  et  $e$  appartient à  $\delta D_n$ . Ainsi,

$$\theta'_n(p) = \frac{1}{1-p} \mathbb{E}_p \left[ \sum_{e \in E(B_n)} \mathbb{1}_{\{e \in \delta D_n\}} \mathbb{1}_{\{0 \xleftrightarrow{D_n} e\}} \right] = \frac{1}{1-p} \mathbb{E}_p \left[ \sum_{e \in \delta D_n} \mathbb{1}_{\{0 \xleftrightarrow{D_n} e\}} \right],$$

ce qui nous amène à l'expression

$$\theta'_n(p) = \frac{1}{p(1-p)} \mathbb{E}_p[\varphi_p(D_n)]$$

et conclut la première partie de la démonstration.

Cette expression étant établie, nous pouvons maintenant utiliser l'hypothèse que  $p > \tilde{p}_c$ . Sous cette hypothèse, et par définition de  $\tilde{p}_c$ , si  $A$  est une partie finie de  $\mathbb{Z}^d$ , le nombre  $\varphi_p(A)$  est supérieur ou égal à 1 pourvu que 0 appartienne à  $A$ , et nul sinon. Autrement dit, pour toute partie finie  $A$  de  $\mathbb{Z}^d$ , on a

$$\varphi_p(A) \geq \mathbb{1}_A(0).$$

Puisque l'origine appartient à  $D_n$  si et seulement si elle n'est pas connectée à  $\partial B_n$ , l'expression que nous avons trouvée de  $\theta'_n(p)$  nous permet d'écrire

$$\theta'_n(p) \geq \frac{1}{p(1-p)} (1 - \theta_n(p)).$$

Il ne nous reste qu'à résoudre cette inéquation différentielle. Tout d'abord, notons que puisque  $p < 1$ ,  $\theta_n(p) < 1$ . Nous pouvons donc écrire

$$\frac{\theta'_n(p)}{1 - \theta_n(p)} \geq \frac{1}{p(1-p)}.$$

On reconnaît à gauche la dérivée par rapport à  $p$  de  $\log \frac{1}{1-\theta_n(p)}$  et à droite celle de  $\log \frac{p}{1-p}$ . On en déduit donc que pour tout  $p > \tilde{p}_c$ ,

$$\frac{1 - \theta_n(\tilde{p}_c)}{1 - \theta_n(p)} \geq \frac{p(1 - \tilde{p}_c)}{(1-p)\tilde{p}_c}.$$

Notons que nous savons que  $\tilde{p}_c$  n'est égal ni à 0 ni à 1. En effet, comme nous l'avons observé après l'énoncé des deux lemmes, le premier d'entre eux, le lemme 7.5.2, entraîne  $\tilde{p}_c \leq p_c < 1$ . Par ailleurs, la partie  $\{0\}$  vérifie  $\varphi_p(\{0\}) = 2dp$ , si bien que  $\tilde{p}_c \geq \frac{1}{2d} > 0$ .

En écrivant maintenant simplement que  $1 - \theta_n(\tilde{p}_c) \leq 1$ , on trouve

$$\theta_n(p) \geq 1 - \frac{(1-p)\tilde{p}_c}{p(1-\tilde{p}_c)} = \frac{p - \tilde{p}_c}{p(1-\tilde{p}_c)}.$$

Puisque cette minoration ne dépend pas de  $n$ , on en déduit que

$$\theta(p) = \lim_{n \rightarrow \infty} \theta_n(p) \geq \frac{p - \tilde{p}_c}{p(1-\tilde{p}_c)} > 0$$

et la preuve est achevée. □

# 8

## Le théorème de Kesten

Dans ce bref chapitre, nous allons utiliser les informations que nous avons accumulées sur les phases sur- et sous-critiques pour enfin démontrer l’assertion 5 du théorème 6.1.5, et ainsi conclure l’étude de la percolation. Rappelons, pour mémoire, l’encadrement de  $p_c(2)$  que nous avons obtenu à la section 6.2.2 : c’était l’encadrement  $\frac{1}{3} \leq p_c(2) \leq \frac{3}{4}$ .

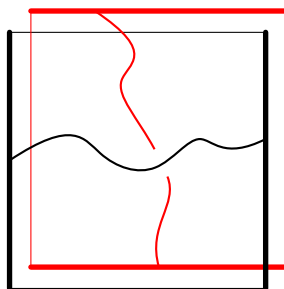
Nous nous plaçons donc désormais dans  $\mathbb{Z}^2$ . Nous allons faire usage de la propriété d’auto-dualité de  $\mathbb{Z}^2$  expliquée en préambule du lemme 6.2.10, et de considérations géométriques spécifiques au cas de la dimension 2, et dans une certaine mesure “évidentes”. Il ne faut néanmoins pas perdre de vue que ces propriétés nécessiteraient des démonstrations rigoureuses, que nous ne fournirons pas ici.

### 8.1 Une inégalité...

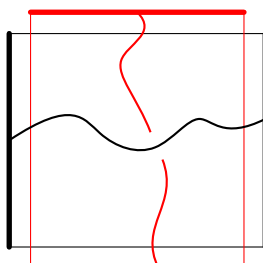
La première chose que nous allons démontrer est la suivante. Considérons une boîte  $B_n$  comme nous l’avons souvent fait. Notons  $H_n$  l’événement “le bord gauche de la boîte  $B_n$  est connecté au bord droit”. Considérons aussi, sans trop chercher à préciser,  $B_n^*$  la boîte duale de  $B_n$ , et notons  $V_n^*$  l’événement “le bord supérieur de la boîte  $B_n^*$  est connecté au bord inférieur”. Un raisonnement approximatif consiste à dire que  $B_n$  et  $V_n^*$  sont des événements complémentaires, à cause de la géométrie de la dualité ; à se souvenir ensuite que la configuration duale est sous  $\mathbb{P}_p$  une percolation de paramètre  $1 - p$  ; et à dire enfin qu’à rotation d’un quart de tour près, les événements  $H_n$  et  $V_n^*$  sont les mêmes, l’un dans le graphe primal et l’autre dans le graphe dual, si bien que

$$\mathbb{P}_p(H_n) = 1 - \mathbb{P}_p(V_n^*) = 1 - \mathbb{P}_{1-p}(H_n).$$

En particulier, pour  $p = \frac{1}{2}$ , on trouve  $\mathbb{P}_{\frac{1}{2}}(H_n) = \frac{1}{2}$ . Ce raisonnement est résumé par le dessin suivant :



qui malheureusement est un peu trop simple pour être correct, puisque rien n'interdit au carré dual d'être entièrement ouvert, et au bord inférieur du carré primal de l'être aussi, auquel cas  $H_n$  et  $V_n^*$  sont simultanément réalisés. Le dessin correct est le suivant :



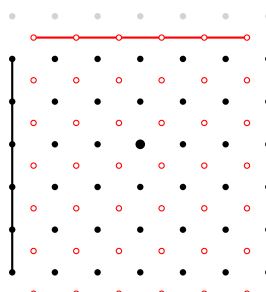
et pour que les boîtes primale et duale aient la même forme à une rotation d'un quart de tour près, elles ne peuvent pas être tout à fait carrées. Définissons donc, pour tout  $n \geq 1$ ,

$$\tilde{B}_n = B_n \setminus (\{-n, \dots, n\} \times \{n\}),$$

la boîte carrée  $B_n$  dont on a enlevé la ligne supérieure. On note  $H_n$  l'événement "le bord gauche de  $\tilde{B}_n$  est connecté à son bord droit". L'énoncé correct est le suivant.

**Lemme 8.1.1.**  $\mathbb{P}_{\frac{1}{2}}(H_n) = \frac{1}{2}$ .

*Démonstration.* Nous nous contenterons du dessin suivant, qui est une représentation précise du cas  $n = 3$  :



Dans ce dessin, l'origine est représentée par un point plus gros que les autres ; les sommets primaux sont comme toujours en noir, et les sommets duaux en rouge. Les sommets gris sont ceux de  $B_n \setminus \tilde{B}_n$ .

L'argument consiste à mettre ensemble les deux observations suivantes :

1. les deux segments noirs sont reliés dans le graphe primal à l'intérieur du rectangle qu'ils délimitent (et qui n'est autre que  $\tilde{B}_n$ ) si et seulement si les deux segments rouges ne sont pas reliés dans le graphe dual à l'intérieur du rectangle qu'ils délimitent ;
2. à  $p = \frac{1}{2}$ , les deux segments noirs ont exactement la même probabilité d'être reliés dans le graphe primal à l'intérieur du rectangle qu'ils délimitent que les deux segments rouges d'être reliés dans le graphe dual à l'intérieur du rectangle qu'ils délimitent.

L'événement  $H_n$  a donc sous  $\mathbb{P}_{\frac{1}{2}}$  la même probabilité que son complémentaire. □

Le fait que le diamètre des agrégats sous-critique ait une queue de distribution exponentielle va nous permettre de montrer que dans la phase sous-critique, l'événement  $H_n$  n'est pas très probable quand  $n$  est grand.

**Lemme 8.1.2.** *Si  $p < p_c(2)$ , alors  $\mathbb{P}_p(H_n) \xrightarrow{n \rightarrow \infty} 0$ .*

*Démonstration.* Sur l'événement  $H_n$ , au moins l'un des sommets de  $\tilde{B}_n$  situés sur l'axe vertical est relié aux bords gauche et droit de  $\tilde{B}_n$ , donc au bord de la boîte  $B_n$  translatée pour être centrée en ce sommet. Ainsi,

$$\mathbb{P}_p(H_n) \leq \sum_{x \in \{0\} \times \{-n, n-1\}} \mathbb{P}_p(x \longleftrightarrow \partial(x + B_n)) = 2n\mathbb{P}_p(0 \longleftrightarrow \partial B_n).$$

Si  $p < p_c(2)$ , il existe donc, d'après la proposition 7.5.1 une constante  $\alpha$  telle que pour tout  $n$  on ait

$$\mathbb{P}_p(H_n) \leq 2ne^{-\alpha n}.$$

Le résultat s'ensuit immédiatement. □

La conséquence logique de ce lemme, au vu du précédent, est que si  $p < p_c(2)$ , alors  $p$  n'est pas égal à  $\frac{1}{2}$ , et de ceci nous déduisons que  $p_c(2) \leq \frac{1}{2}$ .

## 8.2 ...et l'autre

Nous allons maintenant démontrer qu'à  $p = \frac{1}{2}$ , la probabilité de percolation est nulle. Nous allons pour cela utiliser l'inégalité FKG, et l'unicité de l'agrégat infini lorsque  $\theta(p) > 0$ . Rappelons à cette occasion que l'hypothèse de la proposition 7.2.4 est bien  $\theta(p) > 0$ , et non l'hypothèse plus faible  $p > p_c$ .

**Proposition 8.2.1.**  $\theta(\frac{1}{2}) = 0$ .

*Démonstration.* Nous allons raisonner par contradiction. Supposons que  $\theta(\frac{1}{2}) > 0$  et considérons la percolation de paramètre  $\frac{1}{2}$ .

Considérons, pour tout  $n \geq 1$ , les événements  $A_n^h, A_n^b, A_n^g$  et  $A_n^d$  définis respectivement par le fait que le bord haut (resp. bas, gauche, droit) de la boîte  $B_n$  est relié à l'infini à l'extérieur de la boîte  $B_n$ . Autrement dit, par exemple, une configuration appartient à  $A_n^h$  s'il existe un sommet situé sur le bord haut de  $B_n$  et dont l'agrégat est infini dans la configuration obtenue à partir de  $\omega$  en fermant toutes les arêtes de  $E(B_n)$ .

Pour tout  $n \geq 1$ , les quatre événements que nous venons de définir sont croissants et équi-probables. Grâce à l'inégalité FKG, on a donc

$$\mathbb{P}_{\frac{1}{2}}((A_n^h \cup A_n^b \cup A_n^g \cup A_n^d)^c) \geq (1 - \mathbb{P}_{\frac{1}{2}}(A_n^h))^4.$$

L'événement dont le membre de gauche est la probabilité est l'événement où la boîte  $B_n$  ne rencontre pas l'agrégat infini. Puisque nous avons supposé  $\theta(\frac{1}{2}) > 0$ , lorsque  $n$  tend vers l'infini,

cette probabilité tend vers 0. Il s'ensuit que lorsque  $n$  tend vers l'infini, la probabilité commune des quatre événements que nous avons définis tend vers 1.

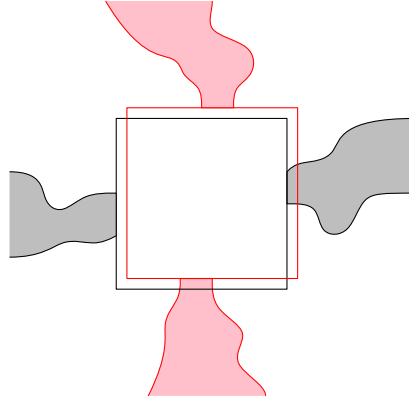
Choisissons  $n$  tel que cette probabilité soit strictement plus grande que  $\frac{4}{5}$ .

Considérons maintenant dans le réseau dual les événements  $A_n^{*h}$ ,  $A_n^{*b}$ ,  $A_n^{*g}$  et  $A_n^{*d}$  définis respectivement par le fait que le bord haut (resp. bas, gauche, droit) de la boîte duale  $B_n^* = B_n + (\frac{1}{2}, \frac{1}{2})$  est relié à l'infini à l'extérieur de la boîte  $B_n^*$ . Par le même raisonnement que ce qui précède, et avec le même choix de  $n$ , ces événements ont une probabilité strictement plus grande que  $\frac{4}{5}$ .

Il découle de ce qui précède que

$$\mathbb{P}_{\frac{1}{2}}(A_n^g \cap A_n^d \cap A_n^{*h} \cap A_n^{*b}) > 1 - 4(1 - \frac{4}{5}) > 0.$$

Avec une probabilité strictement positive, on est donc dans la situation suivante :



Nous savons, par notre hypothèse  $\theta(\frac{1}{2}) > 0$  et d'après la proposition 7.2.4, qu'il y a dans la configuration primale et dans la configuration duale, presque sûrement, un unique agrégat infini. Lorsque nous sommes dans l'événement que nous venons de représenter, cela impose d'une part que les agrégats de la configuration primale qui touchent les bords gauche et droit de  $B_n$  soient reliés dans  $B_n$ , et d'autre part que les agrégats de la configuration duale qui touchent les bords haut et bas de  $B_n^*$  soient reliés dans  $B_n^*$ .

Or ceci n'est pas permis par la géométrie de  $\mathbb{Z}^2$ , et nous aboutissons à la contradiction cherchée. □

Pour la forme, nous allons maintenant conclure la démonstration du théorème 6.1.5 avec la proposition suivante.

**Proposition 8.2.2.**  $p_c(2) = \frac{1}{2}$

*Démonstration.* Nous avons déjà expliqué comment l'inégalité  $p_c \leq \frac{1}{2}$  se déduisait des lemmes 8.1.1 et 8.1.2. La proposition 8.2.1, quant à elle, entraîne immédiatement, par définition de  $p_c$ , l'inégalité  $p_c(2) \geq \frac{1}{2}$ . L'égalité est donc démontrée. □

**Quatrième partie**

**Compléments**

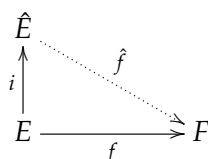
# A

## Complété d'un espace métrique

Nous allons démontrer que tout espace métrique est isométrique à une partie dense d'un espace métrique complet. Cette construction peut être décrite comme la résolution d'un problème universel.

**Définition A.1.** On appelle complété d'un espace métrique  $(E, d)$  la donnée d'un espace métrique complet  $(\hat{E}, \hat{d})$  et d'une injection isométrique  $i : E \rightarrow \hat{E}$  tels que pour tout espace métrique complet  $(F, \delta)$  et toute application uniformément continue  $f : E \rightarrow F$ , il existe une unique application uniformément continue  $\hat{f} : \hat{E} \rightarrow F$  telle que  $\hat{f} \circ i = f$ .

Le diagramme qui représente le problème universel dont le complété est la solution est le suivant :

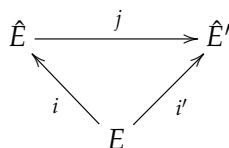


Si l'on dispose d'un espace métrique complet  $G$  et d'une injection isométrique  $i : E \rightarrow G$ , alors  $i : E \rightarrow \hat{E} = \overline{i(E)}$  est un complété de  $E$ . Affirmer, comme nous allons le faire tout de suite, que tout espace métrique admet un complété revient donc à affirmer que tout espace métrique est isométrique à une partie d'un espace métrique complet.

**Théorème A.2.** Soit  $E$  un espace métrique.

1. Il existe un complété de  $E$ .
2. Pour tout complété  $i : E \rightarrow \hat{E}$  de  $E$ , la partie  $i(E)$  est dense dans  $\hat{E}$ .
3. Soient  $i : E \rightarrow \hat{E}$  et  $i' : E \rightarrow \hat{E}'$  deux complétés de  $E$ . Alors il existe une isométrie bijective  $j : \hat{E} \rightarrow \hat{E}'$  telle que  $j \circ i = i'$ .

Le diagramme qui illustre la troisième assertion est le suivant :





*Démonstration.* 1. Notons  $\mathcal{S}$  l'ensemble des suites de Cauchy de  $(E, d)$ . On dit que deux suites de Cauchy  $(x_n)_{n \geq 0}$  et  $(y_n)_{n \geq 0}$  sont *confluentes* si  $\lim_{n \rightarrow \infty} d(x_n, y_n) = 0$ .

La relation de confluence est une relation d'équivalence sur  $\mathcal{S}$  : elle est réflexive, symétrique, et transitive. On note  $\hat{E}$  l'ensemble quotient. On note  $[(x_n)_{n \geq 0}]$  la classe de confluence d'une suite  $(x_n)_{n \geq 0}$ , c'est-à-dire l'ensemble des suites qui lui sont confluentes.

Étant donné deux éléments de  $\mathcal{S}$ , c'est-à-dire deux suites  $(x_n)_{n \geq 0}$  et  $(y_n)_{n \geq 0}$  de Cauchy sur  $E$ , confluentes ou non, la suite de réels  $(d(x_n, y_n))_{n \geq 0}$  est de Cauchy. En effet, pour tous  $n$  et  $m$  entiers, on a

$$|d(x_n, y_n) - d(x_m, y_m)| \leq |d(x_n, y_n) - d(x_m, y_n)| + |d(x_m, y_n) - d(x_m, y_m)| \leq d(x_n, x_m) + d(y_n, y_m)$$

qui tend vers 0 lorsque  $n$  et  $m$  tendent vers l'infini. Ainsi, la suite  $(d(x_n, y_n))_{n \geq 0}$  admet une limite. De plus, si les suites  $(x'_n)_{n \geq 0}$  et  $(y'_n)_{n \geq 0}$  sont respectivement confluentes à  $(x_n)_{n \geq 0}$  et  $(y_n)_{n \geq 0}$ , alors l'inégalité

$$|d(x_n, y_n) - d(x'_n, y'_n)| \leq d(x_n, x'_n) + d(y_n, y'_n)$$

montre que les suites  $(d(x_n, y_n))_{n \geq 0}$  et  $(d(x'_n, y'_n))_{n \geq 0}$  ont la même limite. On peut donc poser sans ambiguïté

$$\hat{d}([(x_n)_{n \geq 0}], [(y_n)_{n \geq 0}]) = \lim_{n \rightarrow \infty} d(x_n, y_n).$$

On vérifie sans difficulté que la fonction  $\hat{d} : \hat{E} \times \hat{E} \rightarrow \mathbb{R}_+$  est une distance sur  $\hat{E}$ .

L'application de  $i : E \rightarrow \hat{E}$  qui envoie un élément  $x$  de  $E$  sur la classe de la suite constante égale à  $x$ , c'est-à-dire sur l'ensemble des suites de  $E$  qui convergent vers  $x$ , est isométrique, puisque pour tous  $x, y \in E$ , on a  $\hat{d}(i(x), i(y)) = d(x, y)$ . Elle est en particulier injective.

Montrons maintenant que l'espace  $(\hat{E}, \hat{d})$  est complet. Soit

$$(\mathbf{x}^p)_{p \geq 0} = ((x_n^p)_{n \geq 0})_{p \geq 0}$$

une suite de Cauchy de  $\hat{E}$ . C'est donc une suite (ici indexée par  $p$ ) de classes d'équivalences de suites de Cauchy de  $E$  (indexées par  $n$ ), et la notation suggère que nous avons choisi, pour chaque  $p \geq 0$ , un représentant  $(x_n^p)_{n \geq 0}$  du terme  $\mathbf{x}^p$  de cette suite.

Nous allons chercher à définir une suite  $(y_n)_{n \geq 0}$  d'éléments de  $E$  en définissant une application strictement croissante  $p : \mathbb{N} \rightarrow \mathbb{N}$  et en posant, pour tout  $n \geq 0$ ,

$$y_n = x_n^{p(n)}. \tag{A.1}$$

Nous voulons choisir l'application  $p$  de telle sorte que cette suite soit de Cauchy, et que la suite  $(\mathbf{x}^p)_{p \geq 0}$  converge vers  $[(y_n)_{n \geq 0}]$ . Définissons pour cela par récurrence une suite strictement croissante  $(N_p)_{p \geq 0}$  d'entiers, comme suit. Posons  $N_0 = 0$  puis, pour tout  $p \geq 1$ , les entiers  $N_0, \dots, N_{p-1}$  étant définis, choisissons  $N_p > N_{p-1}$  tel que pour tous  $n, m \geq N_p$  on ait  $d(x_n^p, x_m^p) < 2^{-p}$ . La suite  $(N_p)_{p \geq 0}$  étant construite, définissons, pour tout entier  $n \geq 0$ ,  $p(n)$  comme l'unique entier tel que  $N_{p(n)} \leq n < N_{p(n)+1}$ . Enfin, pour tout  $n \geq 0$ , définissons  $y_n$  par (A.1).

Montrons que la suite  $(y_n)_{n \geq 0}$  est de Cauchy. Donnons-nous  $\varepsilon > 0$ . Soit  $p \geq 0$  tel que  $2^{-p} < \varepsilon/4$  et tel que pour tous  $q, r \geq p$ , on ait  $\hat{d}(\mathbf{x}^q, \mathbf{x}^r) < \varepsilon/4$ . Pour tous entiers  $n, m$  tels que  $m \geq n \geq N_p$ , on a, pour tout  $k \geq m$ ,

$$\begin{aligned} d(y_n, y_m) &= d(x_n^{p(n)}, x_m^{p(m)}) \\ &\leq d(x_n^{p(n)}, x_m^{p(n)}) + d(x_m^{p(n)}, x_m^{p(m)}) \\ &\leq d(x_n^{p(n)}, x_m^{p(n)}) + d(x_m^{p(n)}, x_k^{p(n)}) + d(x_k^{p(n)}, x_k^{p(m)}) + d(x_k^{p(m)}, x_m^{p(m)}) \\ &\leq 2^{-p(n)} + 2^{-p(n)} + d(x_k^{p(n)}, x_k^{p(m)}) + 2^{-p(m)}. \end{aligned}$$

En faisant tendre  $k$  vers l'infini dans le membre de droite, on trouve

$$d(y_n, y_m) \leq \frac{3\varepsilon}{4} + \hat{d}(\mathbf{x}^{p(n)}, \mathbf{x}^{p(m)}) \leq \varepsilon.$$

Montrons maintenant que la suite  $(\mathbf{x}^p)_{p \geq 0}$  converge vers  $\mathbf{y} = [(y_n)_{n \geq 0}]$ . Donnons-nous à nouveau  $\varepsilon > 0$ . Soit comme précédemment  $p \geq 0$  tel que  $2^{-p} < \varepsilon/3$  et tel que pour tous  $q, r \geq p$ , on ait  $\hat{d}(\mathbf{x}^q, \mathbf{x}^r) < \varepsilon/3$ . Soit  $q \geq p$ . Alors pour tout  $n \geq N_q$  et tout  $m \geq n$ , on a

$$d(x_n^q, y_n) = d(x_n^q, x_n^{p(n)}) \leq d(x_n^q, x_m^q) + d(x_m^q, x_m^{p(n)}) + d(x_m^{p(n)}, x_n^{p(n)}) \leq \varepsilon,$$

si bien que  $\hat{d}(\mathbf{x}^q, \mathbf{y}) \leq \varepsilon$ .

Vérifions maintenant que  $E$  est dense dans  $\hat{E}$ . Soit  $\mathbf{x} = [(x_n)_{n \geq 0}]$  un élément de  $\hat{E}$ . Pour tout  $n \geq 0$ , rappelons que nous notons  $i(x_n)$  la classe de la suite constante égale à  $x_n$ . Nous affirmons que la suite  $(i(x_n))_{n \geq 0}$  converge, dans  $\hat{E}$ , vers  $\mathbf{x}$ . En effet, soit  $\varepsilon > 0$ . Soit  $n_0 \geq 0$  tel que pour tout  $n, m \geq n_0$ , on ait  $d(x_n, x_m) < \varepsilon$ . Alors pour tout  $n \geq n_0$ ,

$$\hat{d}(i(x_n), \mathbf{x}) = \lim_{m \rightarrow \infty} d(x_n, x_m) \leq \varepsilon.$$

Finalement, étant donné un espace complet  $(F, \delta)$  et une application uniformément continue  $f : E \rightarrow F$ , le théorème d'extension des applications uniformément continues assure l'existence d'un prolongement de  $f$  à  $\hat{E}$ , et ce prolongement est unique par densité de  $E$  dans  $\hat{E}$ .

2. Soit  $i : E \rightarrow \hat{E}$  un complété de  $E$ . Considérons  $F = \mathbb{R}$  et l'application  $f : E \rightarrow \mathbb{R}$  identiquement nulle. Les applications  $\hat{f} = 0$  et  $\hat{f}'(\cdot) = \hat{d}(\cdot, i(E))$  sont deux solutions du problème universel qui définit  $i : E \rightarrow \hat{E}$ . Elles sont donc égales, ce qui signifie que la fonction  $\hat{d}(\cdot, i(E))$  est identiquement nulle sur  $\hat{E}$ , c'est-à-dire que  $i(E)$  est dense dans  $\hat{E}$ .

3. Étant donné deux complétés  $i : E \rightarrow \hat{E}$  et  $i' : E \rightarrow \hat{E}'$  de  $E$ , on déduit de la propriété universelle de  $i : E \rightarrow \hat{E}$  appliquée à  $F = \hat{E}'$  et  $f = i'$  l'existence d'une application uniformément continue  $\hat{i}' : \hat{E} \rightarrow \hat{E}'$  telle que  $\hat{i}' \circ i = i'$ . Nous allons montrer que  $j = \hat{i}' : \hat{E} \rightarrow \hat{E}'$  est une bijection isométrique.

Tout d'abord, pour tous  $x, y \in E$ , la suite d'égalités

$$\hat{d}'(j(i(x), i(y))) = \hat{d}'(i'(x), i'(y)) = d(x, y) = \hat{d}(i(x), i(y))$$

montre que  $j$  est isométrique sur  $i(E)$ . Puisque  $j$  est continue et  $i(E)$  est dense dans  $\hat{E}$ , elle est isométrique sur  $\hat{E}$  tout entier. Elle est en particulier injective. Il ne reste qu'à montrer qu'elle est surjective.

---

Pour cela, on applique la propriété universelle de  $i' : E \rightarrow \hat{E}'$  à  $F = \hat{E}$  et  $f = i$  pour obtenir l'existence d'une application uniformément continue  $\hat{i} : \hat{E}' \rightarrow \hat{E}$  telle que  $\hat{i} \circ i' = i$ . On a donc  $\hat{i}' \circ \hat{i} \circ i' = i'$ . Or d'après la propriété universelle de  $i' : E \rightarrow \hat{E}'$  appliquée à  $F = \hat{E}'$  et  $f = i' : E \rightarrow \hat{E}'$ , l'unique application uniformément continue  $\hat{f} : \hat{E}' \rightarrow \hat{E}'$  telle que  $\hat{f} \circ i' = i'$  est l'identité de  $\hat{E}'$ . Ainsi, l'égalité  $\hat{i}' \circ \hat{i} \circ i' = i'$  entraîne  $\hat{i}' \circ \hat{i} = \text{id}_{\hat{E}'}$ . On a donc  $j \circ \hat{i} = \text{id}_E$ , égalité qui impose que  $j$  soit surjective. <sup>1</sup>  $\square$

---

1. Si  $f \circ g$  est injective, alors  $g$  est injective. Si  $f \circ g$  est surjective, alors  $f$  est surjective.

# B

## Limite inférieure, limite supérieure

Ce qui suit est un résumé, sous forme d'exercice, ou plutôt de texte à trous, des définitions et de quelques propriétés des limites inférieure et supérieure d'une suite de réels.

Soit  $(x_n)_{n \geq 0}$  une suite d'éléments de  $[-\infty, +\infty]$ . On définit, pour tout  $n \geq 0$ ,

$$c_n = \inf\{x_p : p \geq n\} \text{ et } d_n = \sup\{x_p : p \geq n\}.$$

La suite  $(c_n)_{n \geq 0}$  est croissante, la suite  $(d_n)_{n \geq 0}$  est décroissante et on a pour tout  $n \geq 0$  les inégalités

$$c_n \leq x_n \leq d_n.$$

Les suites  $(c_n)_{n \geq 0}$  et  $(d_n)_{n \geq 0}$  sont respectivement la plus grande suite croissante et la plus petite suite décroissante qui encadrent la suite  $(x_n)_{n \geq 0}$ , c'est-à-dire telle que ces inégalités aient lieu.

Les suites  $(c_n)_{n \geq 0}$  et  $(d_n)_{n \geq 0}$ , puisqu'elles sont monotones, sont convergentes (insistons sur le fait que nous travaillons dans  $[-\infty, +\infty]$ ). On appelle *limite supérieure* de la suite  $(x_n)_{n \geq 0}$  la limite de la suite  $(d_n)_{n \geq 0}$ , et *limite inférieure* de la suite  $(x_n)_{n \geq 0}$  la limite de la suite  $(c_n)_{n \geq 0}$ . On note

$$\underline{\lim}_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \uparrow c_n \text{ et } \overline{\lim}_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \downarrow d_n.$$

Les limites inférieure et supérieure de  $(x_n)_{n \geq 0}$  sont respectivement la plus petite et la plus grande valeur d'adhérence de cette suite, c'est-à-dire la plus petite et la plus grande limite possible d'une suite extraite de la suite  $(x_n)_{n \geq 0}$ .

**Exercice B.1.** Déterminer l'ensemble des valeurs d'adhérence de la suite  $(x_n)_{n \geq 1}$  dans les cas suivants.

1.  $x_n = (-1)^n$

2. Pour tout  $n \geq 1$ , on écrit  $n = 2^m + k$  avec  $m \geq 0$  et  $k \in \{0, \dots, 2^m - 1\}$ , et on pose  $x_n = k/2^m$ . Les premiers termes de cette suite sont :

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, \dots) = \left(\frac{0}{1}, \frac{0}{2}, \frac{1}{2}, \frac{0}{4}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{0}{8}, \frac{1}{8}, \dots\right) = \left(0, 0, \frac{1}{2}, 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 0, \frac{1}{8}, \dots\right).$$

3.  $x_n = \sin n$  (on pourra étudier la suite  $(e^{in})_{n \geq 1}$ ).

La suite  $(x_n)_{n \geq 0}$  converge si et seulement si ses limites inférieure et supérieure sont égales. Dans ce cas, elle converge vers leur valeur commune.

Jouez au jeu des limites inférieure et supérieure : donnez-vous un réel  $\ell$  et exprimez les assertions

$$\begin{aligned} \underline{\lim}_{n \rightarrow \infty} x_n \leq \ell, \quad \underline{\lim}_{n \rightarrow \infty} x_n < \ell, \quad \underline{\lim}_{n \rightarrow \infty} x_n \geq \ell, \quad \underline{\lim}_{n \rightarrow \infty} x_n > \ell, \\ \overline{\lim}_{n \rightarrow \infty} x_n \leq \ell, \quad \overline{\lim}_{n \rightarrow \infty} x_n < \ell, \quad \overline{\lim}_{n \rightarrow \infty} x_n \geq \ell, \quad \overline{\lim}_{n \rightarrow \infty} x_n > \ell, \end{aligned}$$

sous la forme

$$\begin{aligned} \forall \varepsilon > 0 \mid \forall n \geq 0, \exists p \geq n \mid x_p \leq \ell + \varepsilon \\ \exists \varepsilon > 0 \mid \exists n \geq 0, \forall p \geq n \mid x_p \geq \ell - \varepsilon \end{aligned}$$

Ainsi, par exemple,

$$\underline{\lim}_{n \rightarrow \infty} x_n \leq \ell \iff \forall \varepsilon > 0, \forall n \geq 0, \exists p \geq n, x_p \leq \ell + \varepsilon.$$

Les notions de limite inférieure et supérieure sont des outils puissants, qui permettent souvent de clarifier des énoncés ou des raisonnements. Une propriété simple, mais très utile, est la suivante : si deux suites  $(x_n)_{n \geq 0}$  et  $(y_n)_{n \geq 0}$  vérifient, pour tout  $n \geq 0$ , ou pour tout  $n$  assez grand,  $x_n \leq y_n$ , alors  $\underline{\lim}_{n \rightarrow \infty} x_n \leq \underline{\lim}_{n \rightarrow \infty} y_n$  et  $\overline{\lim}_{n \rightarrow \infty} x_n \leq \overline{\lim}_{n \rightarrow \infty} y_n$ . Ceci est particulièrement utile lorsque l'une des deux suites converge.

Rappelons, pour mémoire et à titre d'illustration, l'énoncé et la démonstration du lemme de Fatou à partir du théorème de convergence monotone.

**Théorème B.1** (Lemme de Fatou). Soit  $(X, \mathcal{A}, \mu)$  un espace mesuré. Soit  $(f_n)_{n \geq 0}$  une suite de fonctions mesurables positives sur cet espace. Alors

$$\int_X \underline{\lim}_{n \rightarrow \infty} f_n \, d\mu \leq \underline{\lim}_{n \rightarrow \infty} \int_X f_n \, d\mu.$$

La limite inférieure d'une suite de fonctions est ici à comprendre point par point : pour tout  $x \in X$ , on a

$$(\underline{\lim}_{n \rightarrow \infty} f_n)(x) = \underline{\lim}_{n \rightarrow \infty} f_n(x).$$

*Démonstration.* Posons, pour tout  $n \geq 0$ ,

$$g_n = \inf\{f_p : p \geq n\},$$

cette définition étant également à comprendre point par point. La suite de fonctions  $(g_n)_{n \geq 0}$  est une suite de fonctions mesurables positives qui converge simplement en croissant vers  $\underline{\lim}_{n \rightarrow \infty} f_n$ . Le théorème de convergence monotone assure donc que

$$\lim_{n \rightarrow \infty} \int_X g_n \, d\mu = \int_X \underline{\lim}_{n \rightarrow \infty} f_n \, d\mu.$$

Par ailleurs, pour tout  $n \geq 0$ , on a  $g_n \leq f_n$ , si bien que

$$\int_X g_n \, d\mu \leq \int_X f_n \, d\mu.$$

On conclut en prenant la limite inférieure des deux membres. □



## Fonctions semi-continues

**Définition C.1** (Fonctions semi-continues). Soit  $(E, d)$  un espace métrique. Soit  $f : E \rightarrow \mathbb{R}$  une fonction.

1. On dit que  $f$  est semi-continue inférieurement (s.c.i.) si pour tout réel  $a$ , l'ensemble

$$\{f \leq a\} = \{x \in E : f(x) \leq a\}$$

est fermé.

2. On dit que  $f$  est semi-continue supérieurement (s.c.s.) si pour tout réel  $a$ , l'ensemble

$$\{f \geq a\} = \{x \in E : f(x) \geq a\}$$

est fermé.

On vérifie, par exemple, que l'indicatrice d'un fermé est semi-continue supérieurement et que l'indicatrice d'un ouvert est semi-continue inférieurement.

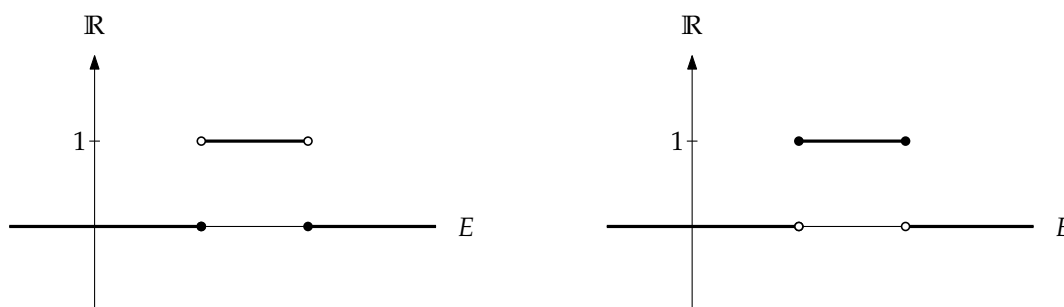


FIGURE C.1 – Les indicatrices d'un ouvert et d'un fermé de  $E$ .

Il découle de la définition précédente (mais cela ne veut pas dire qu'il ne faut pas prendre le temps de le vérifier) qu'une fonction est continue si et seulement si elle est à la fois semi-continue inférieurement et semi-continue supérieurement. Notons également qu'un énoncé sur les fonction s.c.i. peut être converti en un énoncé sur les fonctions s.c.s. (et réciproquement) en observant qu'une fonction  $f$  est s.c.i. si et seulement si la fonction  $-f$  est s.c.s.

On peut caractériser séquentiellement les fonctions s.c.i. (et donc les fonctions s.c.s.), comme suit.

**Proposition C.2.** Soit  $(E, d)$  un espace métrique. Soit  $f : E \rightarrow \mathbb{R}$  une fonction.

1. La fonction  $f$  est s.c.i. si et seulement si pour tout  $x \in E$  et toute suite  $(x_n)_{n \geq 0}$  d'éléments de  $E$  qui converge vers  $x$ , on a

$$f(x) \leq \underline{\lim}_{n \rightarrow \infty} f(x_n). \quad (\text{C.1})$$

2. La fonction  $f$  est s.c.s. si et seulement si pour tout  $x \in E$  et toute suite  $(x_n)_{n \geq 0}$  d'éléments de  $E$  qui converge vers  $x$ , on a

$$\overline{\lim}_{n \rightarrow \infty} f(x_n) \leq f(x).$$

*Démonstration.* Il suffit de démontrer la première assertion. Supposons  $f$  s.c.i. Montrons que pour tout réel  $a$  tel que  $\underline{\lim}_{n \rightarrow \infty} f(x_n) < a$ , on a  $f(x) \leq a$  : ceci entraînera en effet l'inégalité voulue.<sup>1</sup> Or sous l'hypothèse  $\underline{\lim}_{n \rightarrow \infty} f(x_n) < a$ , il existe une sous-suite  $(x_{n_k})_{k \geq 0}$  de la suite  $(x_n)_{n \geq 0}$  telle que pour tout  $k \geq 0$  on ait  $f(x_{n_k}) < a$ , en particulier  $f(x_{n_k}) \leq a$ . Autrement dit, l'ensemble  $\{f \leq a\}$  contient l'ensemble  $\{x_{n_k} : k \geq 0\}$ . Comme  $f$  est s.c.i., l'ensemble  $\{f \leq a\}$  est fermé, et contient donc l'adhérence de l'ensemble  $\{x_{n_k} : k \geq 0\}$ , donc le point  $x$ , si bien que  $f(x) \leq a$ .

Réciproquement, supposons que  $f$  satisfasse (C.1). Soit  $a$  un réel. Si l'ensemble  $\{f \leq a\}$  est vide, il est fermé. Supposons-le non vide et considérons un point  $x$  adhérent à cet ensemble. Il existe une suite  $(x_n)_{n \geq 0}$  d'éléments de  $\{f \leq a\}$  qui converge vers  $x$ . Alors  $\underline{\lim} f(x_n) \leq a$ , donc  $f(x) \leq a$ . Ainsi, l'ensemble  $\{f \leq a\}$  contient le point  $x$ . Il contient donc tous les points qui lui sont adhérents : il est fermé.  $\square$

Dans l'exercice suivant, nous énonçons une propriété importante (mais facile à vérifier en se servant de leur définition) des fonctions semi-continues.

**Exercice C.1.** Une limite simple croissante de fonctions s.c.i. est s.c.i. Plus généralement, le sup d'une famille de fonctions s.c.i. est s.c.i.

Une limite simple décroissante de fonctions s.c.s. est s.c.s. Plus généralement, l'inf d'une famille de fonctions s.c.s. est s.c.s.

Vous avez peut-être remarqué un problème avec l'exercice précédent (pourtant annoncé comme étant facile) : le sup d'une famille de fonctions s.c.i. peut prendre la valeur  $+\infty$ , auquel cas ce sup ne serait plus s.c.i. au sens strict de notre définition. Plutôt que de devoir écarter ce cas dans l'énoncé précédent, on autorise la valeur  $+\infty$  aux fonctions s.c.i. (et, naturellement, la valeur  $-\infty$  aux fonctions s.c.s.). Ainsi, l'énoncé démontré dans l'exercice est vrai, sans restriction.

Nous allons maintenant montrer que dans un espace métrique, toute fonction s.c.i. est la limite simple d'une suite croissante de fonctions continues (ainsi que l'énoncé qui s'en déduit immédiatement pour les fonctions s.c.s.).

Commençons par un exercice d'échauffement. Rappelons que si  $k$  est un réel positif, on dit qu'une fonction  $f : (E, d) \rightarrow \mathbb{R}$  est  $k$ -lipschitzienne si pour tous  $x, y \in E$  on a l'inégalité

$$|f(x) - f(y)| \leq kd(x, y).$$

1. En effet, si on avait  $\underline{\lim} f(x_n) < f(x)$ , il existerait un réel  $a$  tel que  $\underline{\lim} f(x_n) < a < f(x)$ .

Puisque nous autorisons des valeurs infinies à nos fonctions, accordons-nous sur le fait que les fonctions constantes  $+\infty$  et  $-\infty$  sont lipschitziennes. En adoptant l'arithmétique usuelle des infinis, nous voyons que ce sont les seules fonctions lipschitziennes qui prennent, ne serait-ce qu'un en un point, une valeur infinie.

**Exercice C.2.** Soit  $k$  un réel positif. Montrer que le sup (resp. l'inf) d'une famille de fonctions  $k$ -lipschitziennes est  $k$ -lipschitzienne.

Considérons une fonction  $f : E \rightarrow \mathbb{R}$  bornée inférieurement, disons par une certaine constante  $m$ . Donnons-nous une constante positive  $k$ . La fonction constante identiquement égale à  $m$  est inférieure à  $f$ , et  $k$ -lipschitzienne. La famille des fonctions  $k$ -lipschitziennes inférieures à  $f$  est donc non vide et nous pouvons considérer son sup. Il s'agit d'une fonction  $k$ -lipschitzienne, inférieure à  $f$ , et c'est la plus grande. S'il fallait la nommer, on pourrait appeler cette fonction l'*enveloppe  $k$ -lipschitzienne inférieure* de  $f$  (mais cette notion n'est pas d'usage suffisamment courant pour avoir reçu un nom). Notons la  $L_k(f)$ .

Si  $f$  est bornée supérieurement, on définit de même son enveloppe  $k$ -lipschitzienne supérieure, qu'on note  $L^k(f)$ .

**Exercice C.3.** Déterminer l'enveloppe 1-lipschitzienne inférieure de la fonction  $x \mapsto -x^2$  de  $\mathbb{R}$  dans  $\mathbb{R}$ .

**Lemme C.3.** Soit  $(E, d)$  un espace métrique. Soit  $f : E \rightarrow \mathbb{R}$  une fonction bornée inférieurement. Soit  $k \geq 0$  un réel. Pour tout  $x \in E$ ,

$$L_k(f)(x) = \inf\{f(y) + kd(x, y) : y \in E\}. \quad (\text{C.2})$$

Autrement dit, le membre de droite de (C.2) est la plus grande fonction  $k$ -lipschitzienne inférieure à  $f$ .

*Démonstration.* Pour tous  $x, y \in E$ , notons  $g_y(x) = f(y) + kd(x, y)$ . Notons  $m$  une constante qui minore  $f$ . Pour tout  $y$ , la fonction  $g_y$  est minorée par  $m$ . D'après l'exercice précédent, l'inf de la famille  $\{g_y : y \in E\}$  est  $k$ -lipschitzienne. Par ailleurs, pour tout  $x$ , l'égalité  $g_x(x) = f(x)$  montre que cet inf est inférieur à  $f$ . Le membre de droite de (C.2) est donc une fonction  $k$ -lipschitzienne inférieure à  $f$ . Il reste à vérifier que c'est la plus grande.

Soit  $h$  une fonction  $k$ -lipschitzienne inférieure à  $f$ . Pour tous  $x, y \in E$ , on a

$$h(x) \leq h(y) + kd(x, y) \leq f(y) + kd(x, y) = g_y(x).$$

Autrement dit, pour tout  $y \in E$ , la fonction  $h$  est inférieure à la fonction  $g_y$ . La fonction  $h$  est donc inférieure à l'inf de la famille  $\{g_y : y \in E\}$ , ce qui conclut la preuve.  $\square$

Nous pouvons maintenant presque démontrer le résultat annoncé.

**Lemme C.0.1.** Soit  $(E, d)$  un espace métrique. Soit  $f : E \rightarrow \mathbb{R}$  une fonction semi-continue inférieurement et bornée inférieurement. La suite  $(L_n(f))_{n \geq 0}$  est une suite croissante de fonctions continues qui converge simplement vers  $f$ .



*Démonstration.* Par construction, la suite  $(L_n(f))_{n \geq 0}$  est une suite de fonctions continues inférieures à  $f$ . De plus, pour tout  $n \geq 0$ , la fonction  $L_n(f)$  est  $(n+1)$ -lipschitzienne et inférieure à  $f$ , donc elle est inférieure à  $L_{n+1}(f)$ . Ainsi, la suite  $(L_n(f))_{n \geq 0}$  est croissante.

Il reste à montrer que la limite simple de cette suite (qui existe puisqu'elle est croissante) est bien  $f$ . Donnons-nous un point  $x$  de  $E$  et un réel  $\varepsilon > 0$ . L'ensemble  $\{f > f(x) - \varepsilon\}$  est ouvert et contient  $x$ . Il contient donc la boule ouverte  $B(x, r)$  pour un certain  $r > 0$ . Soit  $N$  un entier tel que  $m + Nr \geq f(x)$ . Considérons un entier  $n \geq N$  et estimons  $L_n(f)(x)$ . Pour cela, donnons-nous  $y \in E$ . De deux choses l'une. Soit  $y$  appartient à la boule  $B(x, r)$ , auquel cas

$$f(y) + nd(x, y) \geq f(y) \geq f(x) - \varepsilon,$$

soit  $y$  n'appartient pas à cette boule, ce qui signifie que  $d(x, y) \geq r$ , si bien que

$$f(y) + nd(x, y) \geq m + Nr \geq f(x) \geq f(x) - \varepsilon.$$

En prenant l'inf sur  $y$ , on trouve  $L_n(f)(x) \geq f(x) - \varepsilon$ . Le résultat s'ensuit.  $\square$

**Exercice C.4.** Le fait que la suite  $(L_n(f))_{n \geq 0}$  converge simplement ne dépend pas du fait que  $f$  soit s.c.i., seulement du fait qu'elle soit bornée inférieurement. Quelle est sa limite en général?

Pour résumer, nous avons démontré le résultat suivant.

**Proposition C.4.** *Sur un espace métrique quelconque, toute fonction semi-continue inférieurement et bornée inférieurement est la limite simple d'une suite croissante de fonctions uniformément continues et bornées.*

On peut se demander si l'hypothèse, dans cet énoncé, la que la fonction est bornée inférieurement est essentielle, ou si c'est une hypothèse technique. D'une part, l'exemple de l'exercice C.3 montre que notre stratégie de preuve, telle quelle, ne fonctionne pas en général. L'idée de tronquer une fonction  $f$  en la remplaçant par  $\max(f, -M)$  pour  $M$  une grande constante positive ne semble pas fonctionner facilement.

L'exercice suivant donne toutefois une propriété importante des fonctions s.c.i.

**Exercice C.5.** Montrer que sur un espace métrique compact, une fonction semi-continue inférieurement est bornée inférieurement et atteint son infimum.



## Théorème de représentation de Riesz

---

Sur un espace topologique quelconque, la donnée d'une mesure borélienne finie détermine une forme linéaire sur l'espace des fonctions continues bornées. Les théorèmes de représentation de Riesz, ou de Riesz–Markov, sont une famille d'énoncés qui affirment réciproquement que sur un bon espace topologique, une bonne forme linéaire sur un bon espace de fonctions continues se représente par une bonne mesure borélienne. Nous allons démontrer l'énoncé le plus simple de cette famille.

### D.1 Un théorème de Riesz

**Théorème D.1.1.** Soit  $(E, d)$  un espace métrique compact. Soit  $\Lambda : \mathcal{C}(E) \rightarrow \mathbb{R}$  une forme linéaire. On fait les deux hypothèses suivantes :

1. si  $f \in \mathcal{C}(E)$  est une fonction positive, alors  $\Lambda(f)$  est un réel positif,
2.  $\Lambda(\mathbb{1}_E) = 1$ .

Alors il existe sur  $E$  une unique mesure borélienne de probabilité  $\mu$  telle que pour tout  $f \in \mathcal{C}(E)$  on ait

$$\Lambda(f) = \int_E f \, d\mu.$$

Les hypothèses de ce théorème sont remarquablement faibles : si l'on oublie la condition  $\Lambda(\mathbb{1}_E) = 1$ , à laquelle on peut toujours se ramener, pourvu que  $\Lambda$  ne soit pas identiquement nulle, en divisant  $\Lambda$  par  $\Lambda(\mathbb{1}_E)$  (qui est un nombre positif par la première hypothèse), on ne demande, précisément, que la positivité de  $\Lambda$ , c'est-à-dire que  $f \leq g$  entraîne  $\Lambda(f) \leq \Lambda(g)$ .

Notons tout de même que les hypothèses du théorème, avec le fait que  $(E, d)$  est compact, entraînent la continuité de  $\Lambda$ . En effet, pour tout  $f \in \mathcal{C}(E)$ , on a  $\min f \leq \Lambda(f) \leq \max f$ , si bien que  $|\Lambda(f)| \leq \|f\|$ .

L'unicité de la mesure  $\mu$  est une conséquence de la proposition 1.2.6. En revanche, la démonstration de son existence n'est pas tout à fait simple. Plus précisément, sa difficulté ne réside pas dans les notions auxquelles elle fait appel, qui sont assez élémentaires, mais à sa longueur : il ne semble pas en exister de version très courte. Ce n'est pas si étonnant lorsqu'on pense que ce théorème, sous la forme où nous l'avons énoncé, permet par exemple de construire la mesure de Lebesgue sur le segment  $[0, 1]$  à partir de l'intégrale de Riemann. Or il est notoire que la construction de la mesure de Lebesgue non plus n'est pas tout à fait simple.

La stratégie de la preuve est la suivante. Rappelons que le support d'une fonction est l'adhérence de l'ensemble des points où cette fonction n'est pas nulle. On note  $\text{Supp}(f)$  le support

d'une fonction  $f$ . À partir de  $\Lambda$ , on définit une fonction  $\mu^*$  sur la classe des ouverts de  $E$  en posant, pour tout ouvert  $O$ ,

$$\mu^*(O) = \sup\{\Lambda(f) : f \in \mathcal{C}(E), 0 \leq f \leq 1, \text{Supp}(f) \subseteq O\},$$

puis sur toutes les parties de  $E$  en posant, pour tout  $A \subseteq E$ ,

$$\mu^*(A) = \inf\{\mu^*(O) : O \text{ ouvert}, A \subseteq O\}.$$

On démontre que  $\mu^* : \mathcal{P}(E) \rightarrow [0, 1]$  est une *mesure extérieure*, c'est-à-dire une fonction nulle sur l'ensemble vide, croissante, et sous- $\sigma$ -additive.

À une mesure extérieure sur un ensemble est associée une classe de parties, dites *mesurables* : on dit qu'une partie  $A$  est mesurable si pour toute partie  $F$ , on a

$$\mu^*(F) = \mu^*(F \cap A) + \mu^*(F \cap A^c).$$

Un théorème de Carathéodory affirme que la classe des parties mesurables d'une mesure extérieure est une tribu, et que la restriction de la mesure extérieure à cette tribu est une mesure.

En appliquant ce théorème, on dispose donc d'une mesure, notée  $\mu$ , définie sur la tribu des parties mesurables de  $\mu^*$ . On montre alors que tous les ouverts de  $E$  sont mesurables, si bien que la tribu sur laquelle la mesure  $\mu$  est définie contient la tribu borélienne.

La dernière étape consiste à reconnaître que l'intégrale par rapport à la mesure  $\mu$  coïncide avec la forme linéaire  $\Lambda$ .

Nous allons commencer par l'étude des mesures extérieures.

## D.2 Mesures extérieures et le théorème de Carathéodory

**Définition D.2.1.** Soit  $E$  un ensemble. On appelle mesure extérieure sur  $E$  une fonction

$$\mu^* : \mathcal{P}(E) \rightarrow [0, +\infty]$$

qui vérifie les propriétés suivantes :

1.  $\mu^*(\emptyset) = 0$ ,
2. si  $A \subseteq B$ , alors  $\mu^*(A) \leq \mu^*(B)$ ,
3. si  $(A_n)_{n \geq 0}$  est une suite de parties de  $E$ , alors  $\mu^*(\bigcup_{n=0}^{\infty} A_n) \leq \sum_{n=0}^{\infty} \mu^*(A_n)$ .

À une mesure extérieure, on associe la classe des parties mesurables.

**Définition D.2.2.** Soit  $E$  un ensemble et  $\mu^*$  une mesure extérieure sur  $E$ . On dit qu'une partie  $A$  de  $E$  est mesurable pour  $\mu^*$  si pour toute partie  $F$  de  $E$ , on a

$$\mu^*(F) = \mu^*(F \cap A) + \mu^*(F \cap A^c).$$

Le théorème de Carathéodory est le suivant.

**Théorème D.2.3 (Carathéodory).** Soit  $E$  un ensemble et  $\mu^*$  une mesure extérieure sur  $E$ . La classe des parties de  $E$  mesurables pour  $\mu^*$  est une tribu, et la restriction de  $\mu^*$  à cette tribu est une mesure.

Notons

$$\mathcal{M} = \{A \subseteq E : \forall F \subseteq E, \mu^*(F) = \mu^*(F \cap A) + \mu^*(F \cap A^c)\}$$

la classe des parties mesurables de  $E$ . Nous allons démontrer le théorème en plusieurs étapes.

**Lemme D.2.4.** *La classe  $\mathcal{M}$  est une algèbre.*

*Démonstration.* Il est immédiat que  $\mathcal{M}$  contient l'ensemble vide et que  $\mathcal{M}$  est stable par passage au complémentaire. Montrons que  $\mathcal{M}$  est stable par intersections finies. Pour cela considérons  $A, B \in \mathcal{M}$  et  $F \subseteq E$ .

Alors en observant que

$$(A \cap B)^c \cap A = A \cap B^c \text{ et } (A \cap B)^c \cap A^c = A^c,$$

on trouve

$$\mu^*(F \cap (A \cap B)^c) = \mu^*(F \cap A \cap B^c) + \mu^*(F \cap A^c),$$

d'où on tire

$$\begin{aligned} \mu^*(F \cap (A \cap B)) + \mu^*(F \cap (A \cap B)^c) &= \mu^*(F \cap A \cap B) + \mu^*(F \cap A \cap B^c) + \mu^*(F \cap A^c) \\ &= \mu^*(F \cap A) + \mu^*(F \cap A^c) \\ &= \mu^*(F), \end{aligned}$$

ce qui montre que  $A \cap B$  appartient à  $\mathcal{M}$ . Nous avons donc montré que  $\mathcal{M}$  était une algèbre de parties de  $E$ .  $\square$

Avant de montrer que  $\mathcal{M}$  est une  $\sigma$ -algèbre, nous allons démontrer une assertion qui entraîne en particulier que  $\mu^*$  est  $\sigma$ -additive sur  $\mathcal{M}$ .

**Lemme D.2.5.** *Soit  $(A_n)_{n \geq 0}$  une suite d'éléments deux à deux disjoints de  $\mathcal{M}$  et soit  $F$  une partie de  $E$ . On a l'égalité*

$$\mu^*(F \cap \bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} \mu^*(F \cap A_n).$$

*Démonstration.* Soient  $A$  et  $B$  deux éléments disjoints de  $\mathcal{M}$ . On a

$$\begin{aligned} \mu^*(F \cap (A \cup B)) &= \mu^*(F \cap (A \cup B) \cap A) + \mu^*(F \cap (A \cup B) \cap A^c) \\ &= \mu^*(F \cap A) + \mu^*(F \cap B). \end{aligned}$$

On en déduit par récurrence que pour tout  $N \geq 0$ ,

$$\mu^*(F \cap \bigcup_{n=0}^N A_n) = \sum_{n=0}^N \mu^*(F \cap A_n).$$

Lorsque  $N$  tend vers l'infini, le membre de droite tend vers  $\sum_{n=0}^{\infty} \mu^*(F \cap A_n)$ . La croissance de la fonction  $\mu^*$  entraîne que le membre de gauche est, pour tout  $N$ , inférieur à  $\mu^*(F \cap \bigcup_{n=0}^{\infty} A_n)$ . On en déduit que

$$\mu^*(F \cap \bigcup_{n=0}^{\infty} A_n) \geq \sum_{n=0}^{\infty} \mu^*(F \cap A_n).$$

La  $\sigma$ -sous-additivité de  $\mu^*$  donne l'inégalité inverse.  $\square$

Nous pouvons maintenant montrer que  $\mathcal{M}$  est une tribu, ce qui terminera la démonstration du théorème de Carathéodory.

**Lemme D.2.6.** *La classe  $\mathcal{M}$  est une  $\sigma$ -algèbre.*

*Démonstration.* Puisque nous savons que c'est une algèbre, il suffit par exemple de montrer qu'elle contient l'union de toute suite dont les termes sont deux à deux disjoints. Soit donc  $(A_n)_{n \geq 0}$  une suite d'éléments deux à deux disjoints de  $\mathcal{M}$ . Soit  $F$  une partie de  $E$ . Pour tout  $N \geq 0$ , le fait que  $\mathcal{M}$  soit une algèbre et la croissance de  $\mu^*$  entraînent que

$$\begin{aligned} \mu^*(F) &= \mu^*\left(F \cap \bigcup_{n=1}^N A_n\right) + \mu^*\left(F \cap \bigcap_{n=0}^N A_n^c\right) \\ &\geq \mu^*\left(F \cap \bigcup_{n=1}^N A_n\right) + \mu^*\left(F \cap \bigcap_{n=0}^{\infty} A_n^c\right). \end{aligned}$$

Lorsque  $N$  tend vers l'infini, il découle du lemme précédent que le premier terme du membre de droite tend vers  $\mu^*(F \cap \bigcup_{n=1}^{\infty} A_n)$ , si bien que

$$\mu^*(F) \geq \mu^*\left(F \cap \bigcup_{n=1}^{\infty} A_n\right) + \mu^*\left(F \cap \left(\bigcup_{n=1}^{\infty} A_n\right)^c\right).$$

La sous-additivité de  $\mu^*$  (que nous déduisons de sa sous- $\sigma$ -additivité et du fait que  $\mu^*(\emptyset) = 0$ ) donne l'inégalité inverse, et termine la preuve.  $\square$

### D.3 Démonstration du théorème de Riesz

Considérons maintenant un espace métrique compact  $(E, d)$  et  $\Lambda : \mathcal{C}(E) \rightarrow \mathbb{R}$  une forme linéaire satisfaisant les hypothèses du théorème D.1.1. Définissons, comme nous l'avons annoncé, pour tout ouvert  $O$  de  $E$ ,

$$\mu^*(O) = \sup\{\Lambda(f) : f \in \mathcal{C}(E), 0 \leq f \leq 1, \text{Supp}(f) \subseteq O\}$$

puis, pour toute partie  $A$  de  $E$ ,

$$\mu^*(A) = \inf\{\mu^*(O) : O \text{ ouvert}, A \subseteq O\}.$$

Si  $O$  est une partie ouverte de  $E$ , nous avons deux définitions concurrentes de  $\mu^*(O)$ , mais on vérifie aisément qu'elles coïncident, le point crucial étant que la première définition fait de  $\mu^*$  une fonction croissante sur la classe des ouverts.

Insistons sur le fait que les fonctions qui interviennent dans la définition de  $\mu^*(O)$  sont à support compact inclus dans  $O$ .

Au cours de la preuve, nous aurons besoin d'utiliser deux fois le lemme suivant, ce qui justifie de l'énoncer et de le démontrer séparément.

**Lemme D.3.1.** *Soit  $(K, d)$  un espace métrique compact. Soit  $K = U_1 \cup \dots \cup U_n$  un recouvrement de  $K$  par des ouverts. Il existe des fonctions  $\rho_1, \dots, \rho_n$  continues, positives, à supports respectivement inclus dans  $U_1, \dots, U_n$ , et telles que  $\rho_1 + \dots + \rho_n = 1$ .*

*Démonstration.* Soit  $\varepsilon > 0$  un nombre de Lebesgue du recouvrement : ceci signifie que pour tout  $x \in E$ , il existe  $i \in \{1, \dots, n\}$  tel que  $B(x, \varepsilon) \subseteq U_i$ , si bien qu'en particulier  $d(x, U_i^c) \geq \varepsilon$ . Pour tout  $i \in \{1, \dots, n\}$ , posons

$$F_i = \{x \in E : d(x, U_i^c) \geq \varepsilon\}.$$

L'union des fermés  $F_1, \dots, F_n$  est égale à  $K$ . Définissons, pour tout  $i \in \{1, \dots, n\}$ , une fonction  $f_i$  en posant, pour tout  $x \in K$ ,

$$f_i(x) = \max(1 - \frac{2}{\varepsilon}d(x, F_i), 0).$$

La fonction  $f_i$  est continue, positive, et à support inclus dans  $U_i$ . De plus,  $f_i$  vaut 1 sur  $F_i$ , si bien que la fonction  $f_1 + \dots + f_n$  est strictement positive sur  $K$ . Les fonctions

$$\rho_i = \frac{f_i}{f_1 + \dots + f_n}, \quad i \in \{1, \dots, n\}$$

ont toutes les propriétés voulues. □

Nous pouvons maintenant commencer à étudier  $\mu^*$ .

**Lemme D.3.2.** *La fonction  $\mu^*$  est une mesure extérieure sur  $E$ .*

*Démonstration.* Il est immédiat que  $\mu^*(\emptyset) = 0$  et que la fonction  $\mu^*$  est croissante sur  $\mathcal{P}(E)$ . Montrons qu'elle est sous- $\sigma$ -additive. Commençons par montrer qu'elle est sous-additive sur les ouverts.

**Sous-additivité sur les ouverts** — Donnons-nous donc deux ouverts  $U$  et  $V$  de  $E$  et montrons que

$$\mu^*(U \cup V) \leq \mu^*(U) + \mu^*(V). \quad (\text{D.1})$$

Pour cela, choisissons une fonction  $h \in \mathcal{C}(E)$  comprise entre 0 et 1 et dont le support est inclus dans  $U \cup V$ . Appliquons le lemme D.3.1 au recouvrement du compact  $\text{Supp}(h)$  donné par

$$\text{Supp}(h) = (\text{Supp}(h) \cap U) \cup (\text{Supp}(h) \cap V),$$

ce qui nous fournit deux fonctions  $\rho_U$  et  $\rho_V$ . Posons

$$h_U(x) = \rho_U(x)h(x) \text{ si } x \in \text{Supp}(h) \text{ et } h_U(x) = 0 \text{ sinon.}$$

Considérons un point  $x$  de  $E$ . Si  $x$  n'appartient pas au support de  $h$ , alors la fonction  $h_U$  est nulle dans un voisinage de  $x$ , donc continue en  $x$ . Supposons maintenant que  $x$  appartient au support de  $h$ . Si  $h(x) > 0$ , alors  $x$  appartient à l'intérieur du support de  $h$ , donc  $h_U$ , qui est égale à  $\rho_U h$  sur un voisinage de  $x$ , est continue en  $x$ . Si au contraire  $h(x) = 0$ , alors l'inégalité  $0 \leq h_U \leq h$ , qui est valable en tout point, montre que  $h_U$  est continue en  $x$ .

On définit et on étudie de même la fonction  $h_V$ , et on a partout l'égalité  $h = h_U + h_V$ . Ainsi,

$$\Lambda(h) = \Lambda(h_U) + \Lambda(h_V) \leq \mu^*(U) + \mu^*(V).$$

En prenant le supremum sur  $h$ , on trouve l'inégalité (D.1).

**Sous- $\sigma$ -additivité sur les ouverts** — Montrons maintenant que  $\mu^*$  est sous- $\sigma$ -additive sur les ouverts. Soit donc  $(O_n)_{n \geq 0}$  une suite d'ouverts de  $E$ . Posons  $O = \bigcup_{n=0}^{\infty} O_n$ . Soit  $h$  une fonction continue positive et dont le support est inclus dans  $O$ . Par compacité de  $E$ , et donc du support de  $h$ , il existe  $N \geq 0$  tel que

$$\text{Supp}(h) \subset \bigcup_{n=0}^N O_n.$$

Alors, en utilisant la sous-additivité montrée plus haut,

$$\Lambda(h) \leq \mu^*\left(\bigcup_{n=0}^N O_n\right) \leq \sum_{n=0}^N \mu^*(O_n) \leq \sum_{n=0}^{\infty} \mu^*(O_n).$$

En passant au sup sur  $h$ , on obtient

$$\mu^*\left(\bigcup_{n=0}^{\infty} O_n\right) \leq \sum_{n=0}^{\infty} \mu^*(O_n).$$

**Sous- $\sigma$ -additivité** — Soit finalement  $(A_n)_{n \geq 0}$  une suite de parties de  $E$ . Donnons-nous un réel  $\varepsilon > 0$ . Pour tout  $n \geq 0$ , choisissons un ouvert  $O_n$  contenant  $A_n$  et tel que

$$\mu^*(O_n) \leq \mu^*(A_n) + 2^{-n}\varepsilon.$$

En utilisant la croissance de  $\mu^*$  et sa sous- $\sigma$ -additivité sur les ouverts, on trouve

$$\mu^*\left(\bigcup_{n=0}^{\infty} A_n\right) \leq \mu^*\left(\bigcup_{n=0}^{\infty} O_n\right) \leq \sum_{n=0}^{\infty} \mu^*(O_n) \leq \sum_{n=0}^{\infty} \mu^*(A_n) + 2\varepsilon.$$

Puisque ceci a lieu pour tout  $\varepsilon > 0$ , on a bien démontré la sous- $\sigma$ -additivité.  $\square$

Le théorème de Carathéodory (théorème D.2.3) nous assure que  $\mu^*$  définit une mesure sur la classe des parties  $\mu^*$ -mesurables, c'est-à-dire sur la tribu

$$\mathcal{M} = \{A \subseteq E : \forall F \subseteq E, \mu^*(F) = \mu^*(F \cap A) + \mu^*(F \cap A^c)\}.$$

Notre prochain objectif est de montrer que  $\mathcal{M}$  contient tous les ouverts de  $E$ , donc la tribu borélienne de  $E$ .

**Lemme D.3.3.** *Tout ouvert de  $E$  est  $\mu^*$ -mesurable.*

*Démonstration.* Soit  $O$  un ouvert de  $E$ . Nous voulons montrer que pour toute partie  $F$  de  $E$ ,

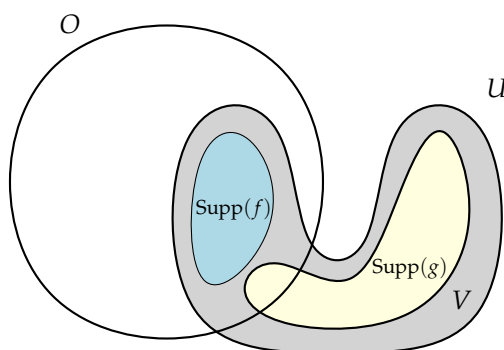
$$\mu^*(F \cap O) + \mu^*(F \cap O^c) = \mu^*(F).$$

Le fait que le membre de gauche soit supérieur au membre de droite est une instance de la sous-additivité de  $\mu^*$ . Il nous reste donc à montrer que

$$\mu^*(F \cap O) + \mu^*(F \cap O^c) \leq \mu^*(F).$$

Nous allons commencer par le faire lorsque  $F$  est un ouvert. Considérons donc un ouvert  $U$ . Soit  $f$  une fonction continue comprise entre 0 et 1 et dont le support est inclus dans  $U \cap O$ . Notons  $V = U \cap \text{Supp}(f)^c$  et considérons une fonction  $g$  continue, comprise entre 0 et 1 et dont le support est inclus dans  $V$ . Alors  $f$  et  $g$  ont des supports disjoints, si bien que  $f + g$  est continue, comprise entre 0 et 1, et à support inclus dans  $U$ . Ainsi,

$$\mu^*(U) \geq \Lambda(f + g) = \Lambda(f) + \Lambda(g).$$



En passant au sup sur  $g$  et en utilisant le fait que  $U \cap O^c \subseteq V$ , on obtient

$$\mu^*(U) \geq \Lambda(f) + \mu^*(V) \geq \Lambda(f) + \mu^*(U \cap O^c).$$

En passant au sup sur  $f$ , on trouve finalement l'inégalité

$$\mu^*(U) \geq \mu^*(U \cap O) + \mu^*(U \cap O^c).$$

Revenons maintenant au cas d'une partie  $F$  arbitraire de  $E$ . Donnons-nous  $\varepsilon > 0$  et choisissons un ouvert  $U$  contenant  $F$  tel que  $\mu^*(U) \leq \mu^*(F) + \varepsilon$ . Alors

$$\mu^*(F \cap O) + \mu^*(F \cap O^c) \leq \mu^*(U \cap O) + \mu^*(U \cap O^c) \leq \mu^*(U) \leq \mu^*(F) + \varepsilon.$$

Puisque  $\varepsilon$  est arbitraire, on a établi l'inégalité annoncée. □

Arrivés à ce point, nous pouvons appliquer le théorème de Carathéodory (théorème D.2.3) qui nous assure que la restriction de  $\mu^*$  à la tribu borélienne de  $E$  est une mesure, que nous noterons désormais  $\mu$ . Il découle directement de la définition de  $\mu^*$  que  $\mu(E) = 1$ , ainsi  $\mu$  est bien une mesure de probabilité.

La dernière étape de la preuve consiste à démontrer que la mesure de probabilité  $\mu$  représente bien la forme linéaire  $\Lambda$ .

**Lemme D.3.4.** *Pour toute fonction  $f \in \mathcal{C}(E)$ , on a  $\int_E f \, d\mu = \Lambda(f)$ .*

*Démonstration.* Soit  $f \in \mathcal{C}(E)$  une fonction continue. Donnons-nous  $\varepsilon > 0$ . Nous allons montrer que

$$\Lambda(f) \leq \int_E f \, d\mu + 2\varepsilon.$$



Puisque  $\varepsilon$  est arbitraire, ceci montrera que  $\Lambda(f) \leq \int_E f \, d\mu$ , et en appliquant cette inégalité à la fonction  $-f$ , on aura le résultat.

Soit  $\delta > 0$  un module d' $\varepsilon$ -uniforme continuité de  $f$  : sur une partie de  $E$  de diamètre inférieur à  $\delta$ , le sup et l'inf de  $f$  diffèrent donc de moins de  $\varepsilon$ .

Soit  $E = B_1 \cup \dots \cup B_n$  une partition de  $E$  en boréliens de diamètre inférieur à  $\delta/3$ . Pour construire une telle partition, on peut commencer par recouvrir  $E$  par toutes ses boules ouvertes de rayon  $\delta/3$ , extraire de ce recouvrement un recouvrement fini  $D_1 \cup \dots \cup D_n$ , puis poser

$$B_1 = D_1, B_2 = D_2 \setminus B_1, \dots, B_k = D_k \setminus (B_1 \cup \dots \cup B_{k-1}), \dots, B_n = D_n \setminus (B_1 \cup \dots \cup B_{n-1}).$$

Pour chaque  $i \in \{1, \dots, n\}$ , considérons un ouvert  $U_i$  contenant  $B_i$  et tel que

$$\mu(U_i) \leq \mu(B_i) + \frac{\varepsilon}{n\|f\|}.$$

Quitte à remplacer  $U_i$  par son intersection avec l'ouvert  $\{x \in E : d(x, B_i) < \delta/3\}$ , ce qui ne peut que diminuer sa mesure, on peut supposer, et nous allons supposer, que le diamètre de  $U_i$  est inférieur à  $\delta$ . Notant  $s_i$  le sup de  $f$  sur  $U_i$ , nous avons donc, sur  $U_i$ , l'inégalité  $s_i - \varepsilon \leq f \leq s_i$ .

Soit  $1 = \rho_1 + \dots + \rho_n$  une partition de l'unité associée au recouvrement de  $E$  par les ouverts  $U_1, \dots, U_n$ , donnée par le lemme D.3.1. Calculons  $\Lambda(f)$ . Commençons par utiliser la linéarité de  $\Lambda$ , puis comparons  $f$  à son sup sur chaque ouvert  $U_i$  : nous obtenons

$$\Lambda(f) = \sum_{i=1}^n \Lambda(f\rho_i) \leq \sum_{i=1}^n s_i \Lambda(\rho_i).$$

Par définition de  $\mu(U_i)$ , puis en utilisant le fait que  $\mu(U_i)$  n'excède pas de beaucoup  $\mu(B_i)$ , nous trouvons

$$\Lambda(f) \leq \sum_{i=1}^n s_i \mu(U_i) \leq \varepsilon + \sum_{i=1}^n s_i \mu(B_i) = \varepsilon + \int_E \sum_{i=1}^n s_i \mathbb{1}_{B_i} \, d\mu.$$

En utilisant maintenant le fait la fonction  $f$  ne diffère pas de plus de  $\varepsilon$ , en norme uniforme, de la fonction  $\sum_{i=1}^n s_i \mathbb{1}_{B_i}$ , nous trouvons

$$\Lambda(f) \leq 2\varepsilon + \int_E f \, d\mu.$$

C'est l'inégalité annoncée, qui conclut la preuve de ce lemme, et du théorème de Riesz.  $\square$

# E

## La mesure de Wiener

Nous allons donner une construction explicite de la mesure de Wiener comme loi de la somme d'une série presque sûrement normalement convergente de fonctions continues aléatoires.

Pour tout  $t \geq 0$ , notons  $\gamma_t$  la loi normale centrée de variance  $t$  sur  $\mathbb{R}$ , étant entendu que  $\gamma_0 = \delta_0$ . Nous allons démontrer le théorème suivant.

**Théorème E.1.** *Il existe sur  $\mathcal{C}([0, 1], \mathbb{R})$  une unique mesure borélienne de probabilité  $\mu$  telle que pour tout  $n \geq 1$  et tous réels  $0 \leq t_1 \leq \dots \leq t_n \leq 1$ , on ait, pour toute fonction borélienne bornée  $F : \mathbb{R}^n \rightarrow \mathbb{R}$ ,*

$$\int_{\mathcal{C}([0,1],\mathbb{R})} F(x_{t_1}, \dots, x_{t_n}) \, d\mu(x) = \int_{\mathbb{R}^n} F(y_1, y_1 + y_2, \dots, y_1 + \dots + y_n) \, d\gamma_{t_1}(y_1) \, d\gamma_{t_2-t_1}(y_2) \dots \, d\gamma_{t_n-t_{n-1}}(y_n).$$

*Démonstration.* Nous allons commencer par définir une base orthonormée de l'espace de Hilbert  $L^2([0, 1])$ . Posons

$$u_{0,0} = \mathbb{1}_{[0,1]}$$

puis, pour tout  $n \geq 1$  et tout  $k \in \{0, \dots, 2^{n-1} - 1\}$

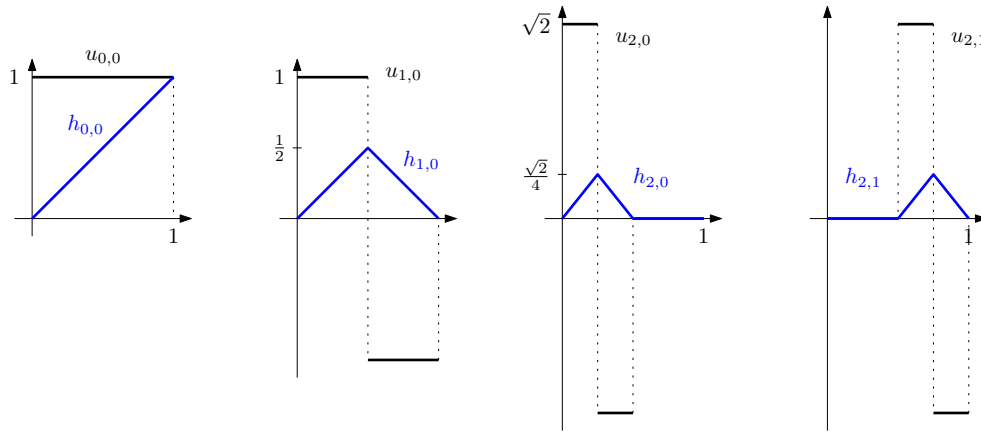
$$u_{n,k} = 2^{\frac{n-1}{2}} \left( \mathbb{1}_{\left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right]} - \mathbb{1}_{\left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right]} \right).$$

Pour tous  $n$  et  $k$  pour lesquels nous avons défini la fonction  $u_{n,k}$ , posons, pour tout  $t \in [0, 1]$ ,

$$h_{n,k}(t) = \int_0^t u_{n,k}(s) \, ds.$$

Les fonctions  $h_{n,k}$  (dont les premières sont représentées plus bas) sont affines par morceaux, positives et, à part  $h_{0,0}$ , « en chapeau ». De plus,  $n$  étant fixé, pour  $k$  et  $l$  distincts, les fonctions  $h_{n,k}$  et  $h_{n,l}$  sont strictement positives sur des ensembles disjoints. Plus précisément, pour  $n \geq 1$  et  $k \in \{0, \dots, 2^{n-1} - 1\}$ , la fonction  $h_{n,k}$  est non nulle uniquement dans l'intervalle  $[k2^{-(n-1)}, (k+1)2^{-(n-1)}]$ . Elle s'annule à l'extrémité gauche de cet intervalle, croît linéairement jusqu'à sa valeur maximale  $2^{-\frac{n+1}{2}}$  qu'elle atteint au milieu de cet intervalle, puis décroît linéairement pour s'annuler à nouveau à l'extrémité droite de l'intervalle. Retenons que

$$\text{supp}(h_{n,k}) = \left[ \frac{k}{2^{n-1}}, \frac{k+1}{2^{n-1}} \right] \quad \text{et} \quad \|h_{n,k}\|_\infty = 2^{-\frac{n+1}{2}}. \quad (\text{E.1})$$



Donnons-nous maintenant, sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ , une famille  $(\zeta_{n,k})_{n,k \geq 0}$  de variables aléatoires i.i.d.  $\mathcal{N}(0, 1)$ . Nous allons montrer que la série de fonctions aléatoires

$$\zeta_{0,0}h_{0,0} + \sum_{n=1}^{\infty} \sum_{k=0}^{2^{n-1}-1} \zeta_{n,k}h_{n,k}$$

converge en un sens que nous allons préciser.

Posons  $f_0 = \zeta_{0,0}h_{0,0}$  et, pour tout  $n \geq 1$ ,

$$f_n = \sum_{k=0}^{2^{n-1}-1} \zeta_{n,k}h_{n,k}.$$

Il s'agit de fonctions de  $\omega \in \Omega$  et de  $t \in [0, 1]$ , plus précisément de variables aléatoires sur  $(\Omega, \mathcal{F}, \mathbb{P})$  à valeurs dans  $\mathcal{C}([0, 1], \mathbb{R})$ . En particulier, pour tout  $n \geq 0$ ,  $\|f_n\|$  est une variable aléatoire réelle positive. Nous allons démontrer que

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \|f_n\| < \infty\right) = 1.$$

La preuve repose sur trois arguments. Le premier est un calcul exact de la loi de  $\|f_n\|$ , pour tout  $n \geq 1$ . En effet,  $f_n$  est une somme de  $2^{n-1}$  fonctions de supports disjoints.<sup>1</sup> Ainsi, la norme de cette somme est le maximum des normes des termes de la somme, c'est-à-dire, vu (E.1),

$$\|f_n\| = 2^{-\frac{n+1}{2}} \max\{|\zeta_{n,k}| : k = 0, \dots, 2^{n-1} - 1\}.$$

Ainsi, si notant  $\zeta$  une variable aléatoire de loi  $\mathbb{N}(0, 1)$ , nous avons, pour tout réel  $a \geq 0$ ,

$$\mathbb{P}(\|f_n\| > a) = \mathbb{P}(\max\{|\zeta_{n,k}| : k = 0, \dots, 2^{n-1} - 1\} > 2^{\frac{n+1}{2}} a) \leq 2^{n-1} \mathbb{P}(|\zeta| > 2^{\frac{n+1}{2}} a).$$

1. Plus précisément, ces fonctions sont strictement positives sur des parties disjointes de  $[0, 1]$ . Leurs supports, au sens topologique, c'est-à-dire les adhérences des ensembles où elles sont strictement positives, ne sont pas disjoints.

Le deuxième argument est une estimation classique de la queue d'une loi gaussienne : pour tout  $u > 0$ , on a

$$\mathbb{P}(|\zeta| > u) \leq \frac{2}{\sqrt{2\pi}u} e^{-\frac{u^2}{2}} \leq \frac{1}{u} e^{-\frac{u^2}{2}}. \quad (\text{E.2})$$

On en déduit donc, pour  $a > 0$ , que

$$\mathbb{P}(\|f_n\| > a) \leq 2^{\frac{n-3}{2}} a^{-1} e^{-2^n a^2}.$$

Le troisième argument est le lemme de Borel–Cantelli : nous allons choisir  $a$  dépendant de  $n$  de manière appropriée. Posons, pour tout  $n \geq 1$ ,  $a_n = 2^{-\frac{n}{4}}$ . Alors

$$\mathbb{P}(\|f_n\| > a_n) \leq 2^{\frac{3n}{4}} e^{-2^{\frac{n}{2}}} = \left( (2^{\frac{n}{4}})^4 e^{-(2^{\frac{n}{4}})^2} \right) 2^{-\frac{n}{4}} = o(2^{-\frac{n}{4}})$$

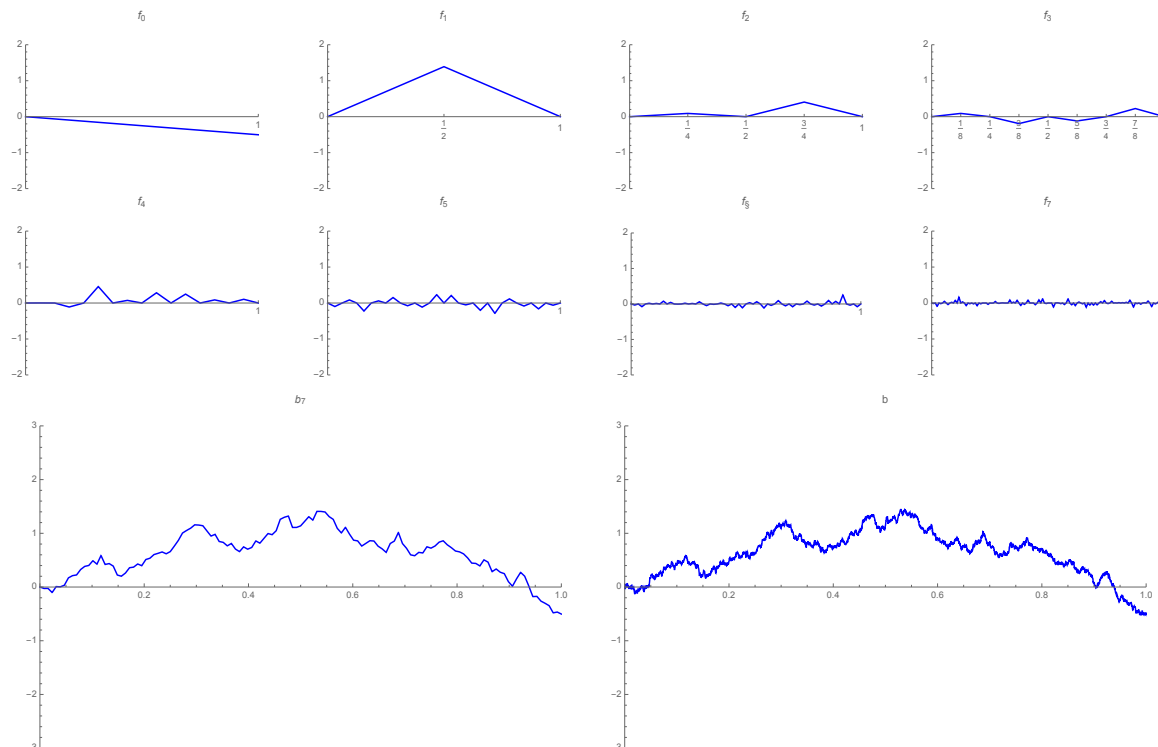
qui est donc sommable en  $n$ .

Ainsi, presque sûrement, on a  $\|f_n\| \leq 2^{-\frac{n}{4}}$  pour  $n$  assez grand, si bien que la série de terme général  $f_n$  est normalement convergente.

Pour tout  $n \geq 1$ , posons

$$b_n = f_0 + \dots + f_n.$$

On voit ci-dessous un tirage des fonctions  $f_0, \dots, f_7$ , de leur somme  $b_7$ , et une approximation de la somme de tous les  $f_n$ .



Nous avons maintenant démontré que la somme

$$b = \sum_{n=0}^{\infty} f_n$$

définissait une variable aléatoire à valeurs dans  $\mathcal{C}([0, 1], \mathbb{R})$ . Montrons que sa loi a les propriétés souhaitées. Nous allons montrer que le processus  $b$  est un processus gaussien, de covariance

$$\mathbb{E}[b_s b_t] = \min(s, t).$$

Pour montrer que le processus est gaussien, une observation cruciale est le fait que si  $t \in [0, 1]$  est un multiple de  $2^{-n}$  pour un certain entier  $n \geq 1$ , alors

$$b_t = f_{0,t} + \dots + f_{n,t}$$

est une combinaison linéaire d'un nombre fini des variables aléatoires  $\xi_{n,k}$ . Il s'ensuit que pour tous  $t_1, \dots, t_r$  nombres dyadiques entre 0 et 1, le vecteur  $(b_{t_1}, \dots, b_{t_r})$  est une image linéaire d'un vecteur formé par un nombre fini des variables aléatoires  $\xi_{n,k}$ . C'est donc un vecteur gaussien.

Pour tous  $t_1, \dots, t_r$  réels entre 0 et 1, on peut trouver des suites  $(t_{m,1})_{m \geq 0}, \dots, (t_{m,r})_{m \geq 0}$  de nombres dyadiques qui convergent respectivement vers  $t_1, \dots, t_r$ . On a alors la convergence presque sûre

$$(b_{t_{m,1}}, \dots, b_{t_{m,r}}) \xrightarrow[n \rightarrow \infty]{} (b_{t_1}, \dots, b_{t_r})$$

qui entraîne une convergence en loi, et le fait que  $(b_{t_1}, \dots, b_{t_r})$  est un vecteur gaussien. Ainsi, le processus  $b$  est gaussien.

Pour calculer sa covariance, on peut considérer  $s, t \in [0, 1]$  et écrire

$$\mathbb{E}[b_s b_t] = h_{0,0}(s)h_{0,0}(t) + \sum_{n=1}^{\infty} \sum_{k=0}^{2^{n-1}-1} h_{n,k}(s)h_{n,k}(t).$$

Revenant à la définition des fonctions  $h_{n,k}$  comme primitives nulles en 0 des fonctions  $u_{n,k}$ , et en utilisant le fait que les fonctions  $u_{n,k}$  forment une base orthonormée de  $L^2([0, 1])$ , et l'égalité de Parseval, on trouve

$$E[b_s b_t] = \sum_{n,k} \langle u_{n,k}, \mathbb{1}_{[0,s]} \rangle_{L^2([0,1])} \langle u_{n,k}, \mathbb{1}_{[0,t]} \rangle_{L^2([0,1])} = \langle \mathbb{1}_{[0,s]}, \mathbb{1}_{[0,t]} \rangle_{L^2([0,1])} = \min(s, t).$$

À partir de là, on peut si on le souhaite vérifier que les accroissements de  $b$  sont indépendants, gaussiens et de la bonne variance.  $\square$

**Exercice E.1.** En utilisant le lemme 2.4.5, montrer que les trajectoires du mouvement brownien sont presque sûrement  $\alpha$ -höldériennes pour tout  $\alpha < \frac{1}{2}$ . Plus précisément, montrer que pour tout  $\alpha \in [0, \frac{1}{2}[$ , on a

$$\mu(\{x \in \mathcal{C}([0, 1], \mathbb{R}) : N_\alpha(x) < \infty\}) = 1.$$



# Transformée de Legendre

## F.1 Définition

Soit  $V$  un espace vectoriel réel. On dit qu'une fonction  $g : V \rightarrow \mathbb{R}$  est *affine* si elle est de la forme  $x \mapsto (\zeta, x) + c$  avec  $\zeta \in V^*$  et  $c \in \mathbb{R}$ . La forme linéaire  $\zeta$  s'appelle la *pente* de  $g$ . Notons que le sup d'une famille quelconque de fonctions affines de même pente  $\zeta$  est

- soit  $+\infty$ ,
- soit une fonction affine de pente  $\zeta$ ,
- soit  $-\infty$  (ceci se produit si la famille est vide).

Nous conviendrons donc, par abus de langage, que les deux fonctions respectivement identiquement égale à  $+\infty$  et à  $-\infty$  sont affines, de toute pente.

Considérons maintenant une fonction  $f : V \rightarrow [-\infty, +\infty]$ , sur laquelle nous ne faisons pour l'instant absolument aucune hypothèse de régularité. Soit  $\zeta$  un élément de  $V^*$ . Appelons *fonction affine d'appui de pente  $\zeta$  pour  $f$*  la fonction

$$g_\zeta = \sup\{g : g \text{ affine de pente } \zeta, g \leq f\}.$$

Posons ensuite

$$f^*(\zeta) = -g_\zeta(0),$$

de telle sorte qu'on ait pour tout  $x \in V$

$$g_\zeta(x) = (\zeta, x) - f^*(\zeta),$$

les cas infinis ne posant pas de problème dans cette écriture.

**Lemme F.1.1.** *Pour tout  $\zeta \in V^*$ , on a*

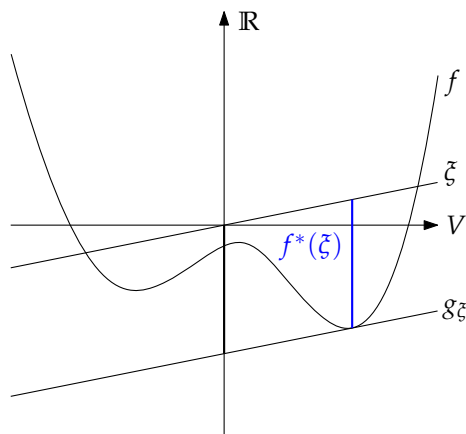
$$f^*(\zeta) = \sup\{(\zeta, x) - f(x) : x \in V\}. \tag{F.1}$$

*Démonstration.* Par notre définition,  $-f^*(\zeta)$  est la plus grande (ou plus exactement la borne supérieure de l'ensemble des) ordonnée(s) à l'origine d'une fonction affine de pente  $\zeta$  inférieure à  $f$ . Autrement dit,

$$-f^*(\zeta) = \sup\{c \in \mathbb{R} : \forall x \in V, \zeta(x) + c \leq f(x)\} = \inf\{f(x) - (\zeta, x) : x \in V\},$$

ce qui à un changement de signe près est l'égalité annoncée. □

La figure suivante montre le lien entre les deux définitions de  $f^*(\zeta)$ .



**Définition F.1.2.** Soit  $f$  une fonction sur  $V$ . La fonction  $f^*$  définie sur  $V^*$  par (F.1) s'appelle transformée de Legendre de  $f$ .

Notons que si  $f = +\infty$ , alors  $f^* = -\infty$ . En revanche, s'il existe un point  $x \in V$  tel que  $f(x) < +\infty$ , alors  $f^*$  ne prend pas la valeur  $-\infty$ . Dans tout les cas, on a  $f^* \geq -f(0)$ .

Notons aussi que si  $f$  prend la valeur  $-\infty$ , alors  $f^*$  est identiquement égale à  $+\infty$ , mais l'exemple de  $V = \mathbb{R}$  et  $f : x \mapsto -x^2$  montre qu'on peut avoir  $f^*$  identiquement égale à  $+\infty$  pour une fonction ne prenant que des valeurs finies.

Dans le cadre des grandes déviations, le fait suivant est important.

**Lemme F.1.3.** La fonction  $f^*$  est semi-continue inférieurement.

*Démonstration.* En effet, elle est le sup d'une famille de fonctions affines, donc continues. □

**Exemple F.1.4.** Prenons  $V = \mathbb{R}$ , donnons-nous un réel  $p > 1$  et une constante  $c > 0$ . Posons

$$f(x) = c \frac{|x|^p}{p}.$$

La fonction  $f$  étant paire, la fonction  $f^*$  l'est aussi. Pour tout  $\zeta \in \mathbb{R}_+$ , on cherche le point  $x_*$  qui réalise le maximum dans (F.1) et on trouve

$$x_* = \left( \frac{\zeta}{c} \right)^{\frac{1}{p-1}}.$$

En remplaçant, et en introduisant le réel  $q$  tel que  $\frac{1}{p} + \frac{1}{q} = 1$ , on trouve

$$f^*(\zeta) = \frac{1}{c^{\frac{q}{p}}} \frac{|\zeta|^q}{q}.$$

L'égalité  $(\zeta, x) \leq f(x) + f^*(\zeta)$ , valable en toute généralité, nous donne ici, pour tous  $x, y \geq 0$ , l'inégalité  $xy \leq \frac{x^p}{p} + \frac{y^q}{q}$ , qui permet par exemple de démontrer l'inégalité de Hölder.

On pourra retenir que pour tout  $t > 0$ , les fonctions  $x \mapsto \frac{x^2}{2t}$  et  $x \mapsto \frac{tx^2}{2}$  sont transformées de Legendre l'une de l'autre.

**Exemple F.1.5.** Toujours avec  $V = \mathbb{R}$ , prenons  $f(x) = x \log x$  pour  $x \in [0, 1]$  et  $x = +\infty$  pour  $x \notin [0, 1]$ . On trouve

$$f^*(\zeta) = e^{\zeta-1} \text{ pour } \zeta \leq 1 \text{ et } f^*(\zeta) = \zeta \text{ pour } \zeta \geq 1.$$

**Exemple F.1.6.** Prenons un exemple différent. Considérons l'ensemble  $\Sigma = \{a_1, \dots, a_N\}$  du chapitre précédent, une mesure  $\mu$  sur  $\Sigma$  de support égal à  $\Sigma$  et, sur l'espace  $\mathcal{M}(\Sigma)$ , la fonction  $f : \nu \mapsto f(\nu) = H(\nu|\mu)$ . Voyons  $\mathcal{M}(\Sigma)$  comme une partie de  $\mathbb{R}^N$ , en l'occurrence l'enveloppe convexe de la base canonique, et prolongeons  $f$  par  $+\infty$  en dehors de  $\mathcal{M}(\Sigma)$ . Nous avons donc

$$f(v_1, \dots, v_N) = \sum_{j=1}^N v_j \log \frac{v_j}{\mu_j}$$

si  $v_1, \dots, v_N$  sont positifs et de somme 1, et  $+\infty$  sinon.

Déterminons la transformée de Legendre de  $f$ . Pour  $g = (g_1, \dots, g_N) \in (\mathbb{R}^N)^*$ , on a

$$f^*(g) = \sup \left\{ \int_{\Sigma} g \, d\nu - H(\nu|\mu) : \nu \in \mathcal{M}(\Sigma) \right\} = \sup \left\{ \sum_{j=1}^N g_j v_j - \sum_{j=1}^N v_j \log \frac{v_j}{\mu_j} : \nu \in \mathcal{M}(\Sigma) \right\}.$$

Le gradient de la fonction  $\nu \mapsto \sum_{j=1}^N v_j (g_j - \log \frac{v_j}{\mu_j})$  vaut  $(g_j + \log \mu_j - \log v_j - 1)_{j=1 \dots N}$  et il est orthogonal à l'hyperplan de  $\mathbb{R}^N$  qui contient  $\mathcal{M}(\Sigma)$ , c'est-à-dire que toutes ses composantes sont égales, s'il existe une constante  $C$  telle que  $\nu = C e^g \mu$ . La constante  $C$  ne peut être que l'inverse de  $\int_{\Sigma} e^g \, d\mu$ . Pour ce  $\nu$ , on trouve

$$\begin{aligned} f^*(g) &= \int_{\Sigma} g \, d\nu - H(\nu|\mu) = \frac{\int_{\Sigma} g e^g \, d\mu}{\int_{\Sigma} e^g \, d\mu} - \int_{\Sigma} \frac{e^g}{\int_{\Sigma} e^g \, d\mu} (g - \log \int_{\Sigma} e^g \, d\mu) \, d\mu \\ &= \log \int_{\Sigma} e^g \, d\mu. \end{aligned}$$

## F.2 Transformée de Legendre et convexité

**Lemme F.2.1.** La fonction  $f^*$  est convexe.

Notons qu'il n'y a pas de problème de définition de la convexité, puisque  $f^*$  est soit identiquement égale à  $-\infty$ , soit à valeurs dans  $] -\infty, +\infty]$ .

*Démonstration.* L'équation (F.1) montre que  $f^*$  est un sup de fonctions affines, donc de fonctions convexes : elle est elle-même convexe.  $\square$

Supposons  $V$  de dimension finie et identifions l'espace bidual  $(V^*)^*$  de  $V$  avec  $V$ . Calculons la transformée de Legendre de  $f^*$ , que nous notons  $f^{**}$ . Pour tout  $x \in V$ , on a

$$\begin{aligned} f^{**}(x) &= \sup \{ (x, \zeta) - f^*(\zeta) : \zeta \in V^* \} \\ &= \sup \{ g_{\zeta}(x) : \zeta \in V^* \} \\ &= \sup \{ \sup \{ g : g \text{ affine de pente } \zeta, g \leq f \} : \zeta \in V^* \} \\ &= \sup \{ g : g \text{ affine, } g \leq f \} \end{aligned}$$

La fonction  $f^*$  est donc le sup de toutes les fonctions affines inférieures ou égales à  $f$ .

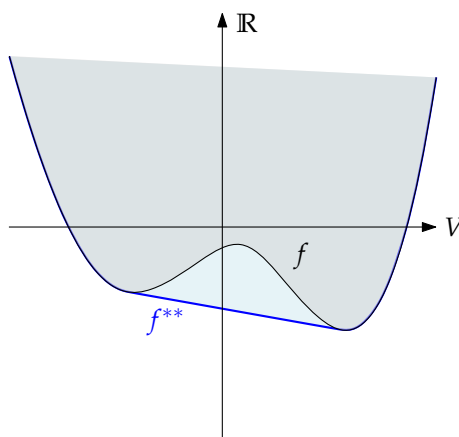


**Lemme F.2.2.** La fonction  $f^{**}$  est la plus grande fonction convexe inférieure ou égale à  $f$ . En particulier, si  $f$  ne prend pas la valeur  $-\infty$  et est convexe, alors  $f^{**} = f$ .

*Démonstration.* La fonction  $f^{**}$  est convexe et, d'après le calcul que nous venons de faire, inférieure à  $f$ . Soit  $h$  une fonction convexe telle que  $h \leq f$ . Toute fonction affine  $g$  telle que  $g \leq h$  vérifie  $g \leq f$ . Or  $h$ , puisqu'elle est convexe, est le sup de la famille des fonctions affines qui lui sont inférieures. Ainsi,

$$h = \sup\{g : g \text{ affine}, g \leq h\} \leq \{g : g \text{ affine}, g \leq f\} = f^{**},$$

ce qui montre que  $h \leq f^{**}$ . □



En termes géométriques, l'épigraphe de  $f^{**}$ , c'est-à-dire  $\{(t, x) \in \mathbb{R} \times V : t \geq f^{**}(x)\}$ , est l'enveloppe convexe de l'épigraphe de  $f$ .

La transformée de Legendre est donc une correspondance bijective entre l'espace des fonctions convexes sur  $V$  et sur  $V^*$ . Lorsque  $V$  est de dimension 1, on peut décrire cette correspondance plus concrètement.

Soit  $f : \mathbb{R} \rightarrow ]-\infty, +\infty]$  une fonction convexe. Notons  $D_f = \{x \in \mathbb{R} : f(x) < \infty\}$ , qui est un intervalle. Sur l'intérieur de  $D_f$ , la fonction  $f$  est dérivable à droite et à gauche en tout point, et on a pour tous  $x < y$  dans l'intérieur de  $D_f$  l'inégalité  $f'_d(x) \leq f'_g(y)$ .

Dans le cas simple où  $D_f = \mathbb{R}$ , où  $f$  est  $C^1$  et où sa dérivée est une bijection de  $\mathbb{R}$  dans  $\mathbb{R}$ , on peut calculer explicitement la dérivée de  $f^*$ . En effet, pour tout  $\xi$  réel, le sup qui définit  $f^*(\xi)$  est atteint en l'unique  $x$  tel que  $f'(x) = \xi$ . On a donc dans ce cas  $f^*(\xi) = \xi(f')^{-1}(\xi) - f((f')^{-1}(\xi))$ . En dérivant cette expression, on voit que la dérivée de  $f^*$  est la bijection réciproque de la dérivée de  $f$ .