

Convergence de mesures, grandes déviations, inégalités de concentration : une introduction

Kevin Tanguy
Lille, France

Version du 6 mars 2021

Table des matières

Avant-propos	9
I Convergence de mesures et grandes déviations	13
1 Introduction	15
1.1 Convergence de variables aléatoires	15
1.2 A propos de la convergence faible	17
1.3 Dualité et structure des espaces de mesures	18
1.3.1 Mesures signés et bornées, mesures de Radon	19
1.3.2 Espace dual	20
1.4 Convergence faible et convergence étroite dans $\mathbb{R}^d$	22
1.4.1 Hiérarchie des modes de convergence	22
1.4.2 Critère de convergence étroite dans $\mathbb{R}^d$	23
1.5 Convergence étroite dans $\mathbb{R}^d$ et transformation de Fourier	26
1.6 Autres caractérisations de la convergence étroite	28
1.7 Références historiques	29
2 Convergence de mesures dans des espaces polonais	33
2.1 Différences avec la dimension finie	33
2.2 Convergence étroite de mesures de probabilités	34
2.2.1 Propriétés des éléments de $\mathcal{P}(E)$	34
2.3 Quelques critères pour établir de la convergence étroite	37
2.4 Distances et convergence étroite	40
2.4.1 Distance de Lévy-Prokhorov	40
2.4.2 Distance de Kantorovich-Rubinstein	42
2.4.3 Topologie de $\mathcal{P}(E)$	42
2.5 Applications	47
2.5.1 Critère de convergence étroite dans $C^0[0, 1]$	48
2.5.2 Principe d'invariance de Donsker	50
2.6 Quelques propriétés du mouvement brownien et de la mesure de Wiener	56
2.7 Remarques historiques	59

3	Introduction aux principes de grandes déviations	61
3.1	Introduction	62
3.2	Théorème de grandes déviations de Cramér dans $\mathbb{R}$	63
3.2.1	Transformée de Fenchel-Legendre (1ère partie)	63
3.2.2	Théorème de Cramér sur $\mathbb{R}$	66
3.3	Théorème de Cramér dans $\mathbb{R}^d$ avec $d \geq 1$	75
3.3.1	Transformée de Fenchel-Legendre (2ième partie)	76
3.3.2	Application : théorème de Sanov pour des mesures empiriques	81
3.4	Théorème de Gärtner-Ellis	84
3.4.1	Transformée de Fenchel-Legendre (3ième partie)	86
3.4.2	Théorème de Gärtner-Ellis	86
3.5	Principes de grandes déviations	92
3.5.1	Fonctions de taux et principes de grandes déviations	93
3.5.2	Schéma de preuve et aspects théoriques des grandes déviations	93
3.5.3	Principes de contraction	95
3.6	Versions générales des théorèmes de Cramér, Gärtner-Ellis et Sanov	99
3.6.1	Théorèmes de Cramér et de Gärtner-Ellis	99
3.6.2	Théorème de Sanov	100
3.7	Références historiques	101
4	Mesures gaussiennes dans des espaces de Banach	103
4.1	Mesures gaussiennes sur des espaces de dimension finie	103
4.1.1	Principe de grandes déviations pour les vecteurs gaussiens	105
4.2	Propriétés de concentration des mesures gaussiennes	108
4.2.1	Inégalités de concentration	108
4.2.2	Isopérimétrie gaussienne	112
4.3	Principe de grandes déviations pour la mesure de Wiener	115
4.3.1	Formule de translation pour la mesure de Wiener	115
4.3.2	Inégalité isopérimétrique (faible) pour la mesure de Wiener	119
4.3.3	Théorème de Schilder	121
4.3.4	Extension à un cadre abstrait	123
4.4	Intégrabilité des suites gaussiennes	126
4.5	Références historiques	129
II	Concentration de la mesure	131
5	Introduction à la concentration de la mesure	133
5.1	Cadre abstrait	133
5.2	Isopérimétrie	135
5.2.1	Problème isopérimétrique dans l'espace euclidien	135
5.2.2	Isopérimétrie sur la sphère et l'espace gaussien	136
5.3	Inégalités de déviations	137
5.4	Réduction de la dimension	139
5.5	Théorème de Dvoretzky	140
5.6	Références historiques	144

6	Principes de base et inégalités exponentielles classiques	147
6.1	Inégalité de Markov et ses conséquences	148
6.2	Méthode de Cramér-Chernoff	149
6.3	Variables aléatoires sous-gaussiennes	151
6.4	Inégalités de Bernstein et de Bennett	157
6.5	Références historiques	158
7	Contrôle de la variance	161
7.1	Inégalité d'Efron-Stein	161
7.1.1	Différences bornées	163
7.2	Variation autour de l'inégalité d'Efron-Stein	166
7.3	Inégalité de Poincaré	168
7.4	Références historiques	171
8	Inégalités fonctionnelles et modèle gaussien	173
8.1	Inégalités fonctionnelles	173
8.1.1	Inégalité de Poincaré	174
8.1.2	Inégalité de Sobolev logarithmique	175
8.2	Modèle gaussien	178
8.3	Application en statistiques	180
8.3.1	Modèles linéaire gaussien généralisé	180
8.3.2	Choix de modèle	181
8.3.3	Méthode du LASSO	183
8.4	Démonstrations par semi-groupes et méthodes d'interpolation	186
8.4.1	Notations	187
8.4.2	Inégalités fonctionnelles par interpolation	189
8.4.3	Aparté : convergence vers la mesure invariante et entropie relative	192
8.4.4	Hypercontractivité	194
8.5	Inégalités de courbure-dimension	196
8.5.1	Introduction	197
8.5.2	Courbure-dimension et inégalité de Poincaré	198
8.5.3	Courbure dimension et inégalités de Sobolev logarithmiques	201
8.5.4	Autres inégalités fonctionnelles	202
8.6	Références historiques	204
9	Entropie	207
9.1	Argument de Herbst	207
9.2	Concentration sur les espaces produits	208
9.2.1	Différences bornées	208
9.2.2	Résultat de Talagrand	211
9.2.3	Approche entropique	213
9.3	Etude de processus empiriques	215
9.3.1	Approche finie dimensionnelle (partie 1)	216
9.3.2	Etude de l'inégalité différentielle (partie 1)	218
9.3.3	Approche finie dimensionnelle (partie 2)	220
9.3.4	Etude de l'inégalité différentielle (partie 2)	221
9.4	Références historiques	222

10 Transport optimal	223
10.1 Introduction	223
10.1.1 Problème de Monge-Kantorovich	223
10.1.2 Problème de Monge-Kantorovich, résultats qualitatifs	225
10.1.3 Dualité de Kantorovich	225
10.1.4 Théorème de Brenier	227
10.1.5 Equation de Monge Ampère	227
10.2 Problème de Monge-Kantorovich, aspect métrique	228
10.2.1 Lien avec la topologie faible	231
10.2.2 Applications	231
10.3 Inégalité de Transport et concentration	233
10.3.1 Transport et concentration, argument de Marton	234
10.3.2 Inégalité de concentration indépendante de la dimension et $T_2(C)$	236
10.4 Transport optimal et inégalités fonctionnelles	236
10.4.1 Inégalité de Sobolev logarithmique	238
10.4.2 Inégalités de Brunn-Minkowski et de Prekopa-Leidler	240
10.5 Semi-groupe d'Hamilton-Jacobi et hypercontractivité	243
10.6 Hiérarchie des inégalités fonctionnelles	246
10.7 Références historiques	251
 III Superconcentration	 255
11 Introduction	257
11.1 Phénomène de superconcentration	258
12 Modèles	263
12.1 Théorie des extrêmes	263
12.1.1 Famille de variables aléatoires indépendantes et de même loi	264
12.1.2 Famille de variables aléatoires gaussiennes	265
12.2 Matrices aléatoires	265
12.3 Marche aléatoire branchante	268
12.4 Percolation	269
12.5 Verres de spin	271
12.6 Références bibliographiques	274
13 Superconcentration : étude de la variance	283
13.1 Inégalité de Talagrand	283
13.1.1 Marche aléatoire branchante	287
13.1.2 Percolation	288
13.1.3 Remarques sur l'inégalité de Talagrand	290
13.2 Inégalité de Chatterjee	290
13.2.1 Exemple d'application	294
13.3 Transport monotone	295
13.3.1 Introduction	295
13.3.2 Notations et rappels de transport optimal	296
13.3.3 Inégalités de Poincaré à poids	297

13.3.4 Applications	298
13.4 Remarques sur les inégalités de Poincaré à poids	302
13.5 Références historiques	303
14 Inégalités de superconcentration	307
14.1 Version exponentielle du Théorème de Chatterjee	308
14.1.1 Suite gaussienne stationnaire	311
14.1.2 Application en test statistique	313
14.2 Inégalités de déviation et inégalités de Poincaré à poids	315
14.2.1 Applications	316
14.3 Convexité et concentration	317
14.3.1 Inégalité de Nguyen Tien	318
14.3.2 Extension	320

Avant-propos

De nombreux ouvrages spécialisés, de différentes factures, sont sans cesse publiés et utilisés par la communauté mathématique pour se former sur de nouveaux thèmes et éventuellement les enseigner. Cependant, cette littérature est massive et il est parfois difficile de s'y retrouver dans ces dédales de pages et de livres. Fréquemment les livres en question sont davantage écrits à l'intention d'experts d'un domaine précis, permettant ainsi de faire un bilan sur les dernières découvertes obtenues, que pour le néophyte souhaitant se former. Cet objectif louable et incontournable présente malgré tout les défauts suivants : la lecture est difficile et l'exhaustivité rebutante. Le lecteur peut alors se noyer sous une montagne d'informations, d'innombrables résultats techniques et pointus ou se figer face à des démonstrations trop élaborées.

Face à cela, une alternative semble apparaître avec les « écoles d'été » durant lesquelles de prestigieux mathématiciens viennent dispenser leur savoir dans un domaine particulier ; cela donne régulièrement lieu à des notes de cours, plus ou moins élaborées et détaillées, dont la qualité pédagogique varie d'un auteur à l'autre. Les raisons de ces défauts me semblent principalement être dues à un manque de temps. En effet, le mathématicien croule sous de nombreuses tâches : la recherche et la publication de résultats, la diffusion de son travail en effectuant des exposés dans d'autres universités, la partie administrative permettant de faire fonctionner son laboratoire de recherche, sa charge d'enseignement,... Il semble alors compréhensible qu'une partie des mathématiciens ne consacre qu'un temps limité à la rédaction d'ouvrages introductifs à visée pédagogique car cela est un travail de longue haleine.

En 2018, j'ai pris la décision de quitter le monde de la recherche pour me consacrer à l'enseignement des mathématiques. J'ai, depuis, la nette conviction que je suis plus compétent pour transmettre un savoir plutôt que de le dénicher. Néanmoins, je ne souhaitais pas oublier toutes les merveilleuses choses que j'ai pu découvrir dans le monde de la recherche. Ceci m'a alors poussé à dresser un bilan de ce que j'ai pu apprendre jusque-là. Je me suis alors employé à rassembler les sujets qui m'avaient particulièrement intéressés à la fin de mes études et durant ma thèse. J'ai donc commencé à rédiger cet ouvrage à partir des choses que l'on m'avait enseigné tout en ajoutant celles que j'ai dû acquérir par moi-même. En plus d'être un exercice extrêmement satisfaisant d'un point de vue personnel, ce fut également l'occasion de créer un livre introductif sur des domaines qui m'intéressent et qui me semblent essentiels à n'importe quel jeune probabiliste.

J'ai donc essayé de rédiger ce livre sous la forme d'un cours introductif prêt à l'emploi. J'espère, et cela est sûrement prétentieux de ma part, avoir pu écrire un ouvrage qui oscille entre des passages techniques, des survols (référéncés) de certains aspects élaborés et de suffisamment

d'exemples pour faciliter l'apprentissage du lecteur. En cela, l'ouvrage se veut pédagogique. Il permet d'entrouvrir de nombreuses portes au lecteur ; lequel pourra ensuite s'y engouffrer, avec suffisamment de bagage mathématique, pour approfondir les thématiques de son choix.

Dans le cours qui va suivre, il est supposé que lecteur soit au moins en possession d'une licence de mathématiques. Il est également supposé qu'il ait suivi un premier cours de probabilités et d'intégration (au sens de Lebesgue) ; dans une moindre mesure, il est également attendu qu'il ait des notions basiques d'analyse fonctionnelle.

Le livre est découpé en trois parties. La première d'entre-elles est assez classique et peut servir de matériel de cours durant un semestre de master 2 recherche. Dans celle-ci, nous débutons par des rappels, en dimension finie, sur les différentes notions de convergence de variables aléatoires. Un effort pédagogique particulier a été apporté pour motiver l'étude de la convergence en loi et des nombreux théorèmes de dualité associés. Le chapitre 2 traite de la convergence en loi de variables aléatoires dans des espaces de dimension infinie et expose les outils essentiels liés à ce nouveau cadre d'étude. Le résultat le plus important de cette partie est sans conteste le théorème de Prokhorov 2.4.3 qui permet d'établir un résultat de convergence en loi (pour une sous-suite) à partir d'un résultat de compacité de l'espace ambiant. Pour illustrer ces notions abstraites, nous proposons le théorème de Donsker 2.5.1 qui peut s'apparenter à un théorème de la limite centrale dans l'espace des fonctions continues $C^0[0, 1]$. De plus, ce théorème permet de mettre en évidence l'existence de la mesure de Wiener associée au mouvement Brownien.

Le chapitre 3 introduit la notion du principe de grandes déviations : il s'agit de résultats asymptotiques permettant de quantifier certaines vitesses de convergence. Les exemples les plus emblématiques sont certainement les théorèmes de Cramér 3.2.1 et 3.3.1 qui quantifient la vitesse de convergence de la loi forte des grands nombres de Kolmogorov. Ce chapitre sera également l'occasion d'obtenir des résultats similaires dans différents cadres : lorsque les variables aléatoires ne sont pas indépendantes (cf. le théorème de Gärtner-Ellis 3.4.1), lorsque les variables aléatoires sont à valeurs dans un espace métrique discret (cf. théorème de Sanov 3.3.2). Ensuite, nous présentons quelques principes généraux permettant d'établir des principes de grandes déviations dans un cadre relativement large.

Le chapitre 4 est alors l'occasion de présenter un certain nombre de résultats satisfaits par la mesure gaussienne standard sur $\mathbb{R}^d$ pour ensuite les étendre à des mesures sur des espaces de dimension infinie (notamment la mesure de Wiener). En particulier, nous introduisons la notion de concentration de la mesure et de problèmes isopérimétriques. Le théorème le plus important de cette partie est celui de Schilder 4.3.2 qui établit un principe de grandes déviations pour la mesure de Wiener.

La deuxième partie du livre se focalise sur le phénomène de la concentration de la mesure et de ses différents avatars. En sélectionnant les thèmes de cette partie, il est possible d'élaborer facilement un cours d'un semestre pour des étudiants de master 2 recherche. Le chapitre 5 aborde de manière historique le phénomène de la concentration de la mesure et présente deux applications : l'une permettant d'effectuer de la réduction de dimension (cf. théorème de Johnson-Lindenstrauss 5.4.1), l'autre, plus géométrique, concerne les sections de domaines convexes (cf. théorème de Dvoretzky 5.5.1). Le chapitre 6 est plus classique car il concerne des inégalités de déviations

pour des sommes de variables aléatoires. Plusieurs méthodologies sont alors décrites (inégalité de Markov, méthode de Chernoff,...) et des exemples nombreux et variés sont donnés pour les illustrer. Le chapitre 7 expose une première étape d'abstraction. Celle-ci permet d'obtenir des résultats de concentration pour des fonctions plus complexes que des sommes de variables aléatoires. En particulier, ceci s'effectue en contrôlant de différentes manières la variance d'une fonction aléatoire grâce à l'inégalité d'Efron-Stein 7.1.1 (et ses variantes) ou par l'inégalité de Poincaré 7.3.1; à nouveau de nombreux exemples sont donnés pour illustrer notre propos. Le chapitre 7 laisse penser que certaines inégalités fonctionnelles sont étroitement liées au phénomène de concentration de la mesure, le chapitre 8 montre que c'est effectivement le cas et se focalise sur les inégalités de Poincaré ou de Sobolev logarithmique ainsi que leurs conséquences vis-à-vis de la concentration de la mesure. La mesure gaussienne standard sur $\mathbb{R}^n$ apparaît alors comme un exemple parfait permettant d'illustrer ces notions. Ce chapitre est également l'occasion de mettre en évidence la polyvalence des outils liés à la concentration de la mesure : en particulier, nous aborderons des résultats issus de la Statistique comme le choix de modèle et la méthode du LASSO. Le chapitre 8 permet également d'introduire la théorie de Bakry et Émery. Celle-ci permet d'établir de manière élémentaire, pour une large classe de mesures de probabilités, des inégalités fonctionnelles à partir de méthode d'interpolation par semi-groupes (cette section fait alors écho à une démonstration du chapitre 4). Le chapitre 9 montre qu'un contrôle avisé de l'entropie d'une fonction f permet d'obtenir de nombreux résultats de concentration pour cette dernière; en guise d'exemple, nous proposons tout au long de ce chapitre une étude de processus empiriques. Le chapitre 10 présente une nouvelle approche du phénomène de concentration de la mesure et conclut cette deuxième partie. Dans le chapitre 10, nous introduisons la théorie du transport optimal et montrons de quelle manière les objets qui lui sont liés permettent d'établir des résultats de concentration. Ce chapitre est également l'occasion de dresser un bilan en hiérarchisant les différents résultats de concentration exposés dans la deuxième partie du livre.

La dernière partie de ce livre peut aisément être utilisée pour un travail d'étude et de recherche. En effet, les chapitres 11, 12 et 13 sont un peu à part puisqu'ils ont pour objectif d'introduire des questions qui sont apparues récemment dans le monde de la recherche. Soyons plus précis : la deuxième partie de l'ouvrage s'évertuait à démontrer la polyvalence du phénomène de concentration de la mesure. Les outils qui lui sont liés sont indéniablement incontournables dans la recherche actuelle. Néanmoins, cette polyvalence entraîne forcément un manque de précision sur certains cas particuliers. Autrement dit, pour certains exemples particuliers, il est possible de renforcer considérablement les résultats fournis par la théorie « classique » de la concentration de la mesure. Ce phénomène porte le nom de *superconcentration* et a été mis en évidence par le mathématicien Chatterjee aux alentours des années 2008. Depuis, ce phénomène apparaît ponctuellement dans de nombreux (et variés) domaines des mathématiques et ces améliorations (de la concentration) sont souvent obtenus au prix d'efforts techniques considérables. L'étude de ce domaine fut précisément le sujet de ma thèse et je me propose, modestement, dans ces derniers chapitres, d'introduire au lecteur certains de ces aspects les plus abordables. Il va sans dire qu'une partie de moi-même espère convaincre suffisamment de lecteurs de la pertinence de ce sujet. Ainsi, une partie d'entre-eux pourraient s'investir dans ce domaine et éventuellement résoudre certains problèmes ouverts dont la solution m'intéresse fortement et pour lesquels je ne trouverais certainement jamais le fin mot de l'histoire seul.

Le livre ne contient pas d'exercices car nous n'avons pas jugé bon de reproduire ceux déjà

présents dans la littérature. Le lecteur pourra donc se référer aux ouvrages [82, 89, 44] à ce sujet. Le lecteur souhaitant enseigner certains passages contenus dans ce livre pourra, à sa guise, facilement composer des exercices à partir du matériel exposé.

Pour conclure cet avant-propos, il me semble essentiel de rendre à César ce qui lui appartient. L'ensemble des connaissances que j'ai rassemblées dans ce cours sont issues des différents enseignements que l'on m'a prodigué durant ma scolarité à l'Université de Toulouse ainsi que de certains ouvrages ou articles mathématiques. Ainsi, je souhaite remercier chaleureusement M. Ledoux qui m'a permis de m'appuyer sur la structure de ses cours de master 2 dont les mots qui vont suivre semblent être un prolongement.

Il convient également de remercier chaleureusement C. Tanguy et M. Deléglise qui m'ont apporté une aide inestimable en relisant attentivement ce manuscrit à la recherche de fautes d'orthographe, tout en proposant différents conseils permettant de l'améliorer. Je souhaite également remercier J. Bernard pour ses précieuses figures ; sans sa dextérité précoce (elle n'est encore qu'au lycée !) ce livre n'aurait pu profiter de graphiques.

Je tiens aussi à remercier P. Petit, P. Monmarché, M. Talagrand et I. Nourdin pour leurs remarques et leurs conseils m'ayant permis d'améliorer la qualité du manuscrit.

Première partie

Convergence de mesures et grandes déviations

Chapitre 1

Introduction

Dans ce chapitre nous allons procéder à de nombreux rappels. Nous commencerons par les différentes notions de convergence qu'il est possible d'étudier dans le domaine des probabilités. Naturellement, cela nous mène rapidement vers la notion de convergence en loi. Pour mieux cerner celle-ci, nous procédons à différents rappels concernant les notions de convergences faibles sur un espace vectoriel normé. Nous ferons ensuite un bref résumé concernant certains résultats de dualité pour des espaces de fonctions avec pour objectif l'étude de convergence de mesures (i.e. la convergence en loi de variables aléatoires). Enfin, nous présenterons une liste de résultats standards (critères, caractérisations, applications) portant sur la convergence en loi de variables aléatoires dans $\mathbb{R}^d$. Certaines démonstrations seront omises et nous renvoyons le lecteur vers les ouvrages [22, 89] à ce sujet.

1.1 Convergence de variables aléatoires

Par la suite, nous allons considérer des variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace polonais E muni de sa tribu borélienne $\mathcal{B}(E)$.

Profitons-en pour rappeler (cf. [89]) les faits suivants :

- un espace topologique $(E, \mathcal{T})$ est un espace polonais si et seulement si $\mathcal{T}$ est métrisable pour une distance d et que (E, d) soit un espace métrique complet et séparable ;
- un espace (E, d) est séparable lorsqu'il contient une suite $(x_n)_{n \geq 0}$ dense.

Exemple 1.1.1. Voici quelques exemples d'espaces polonais que nous allons rencontrer tout au long de cet ouvrage :

- $(\mathbb{R}^n, |\cdot|)$ avec $|\cdot|$ la distance euclidienne est un espace polonais et les vecteurs à coordonnées rationnelles forment un ensemble dénombrable et dense dans $\mathbb{R}^n$.
- Tout espace de Hilbert $(\mathbb{H}, \|\cdot\|)$ est un espace polonais.
- L'espace des fonctions continues sur $[0, 1]$, $(C^0[0, 1], \|\cdot\|_\infty)$ est un espace polonais.

Il convient de garder en tête ces exemples tout au long de cet ouvrage ; comme cela est flagrant pour les exemples ci-dessus, il est fréquent qu'une structure d'espace vectoriel soit également

présente.

Sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ nous allons nous concentrer sur différents modes de convergence d'une suite de variables aléatoires $(X_n)_{n \geq 0}$ (à valeurs dans E). Il est traditionnel d'étudier les modes de convergences suivants.

Définition 1.1.1. Soient $(X_n)_{n \geq 0}$ et X des variables aléatoires. Nous dirons que

- X_n converge presque sûrement (p.s. en abrégé) vers X si

$$\mathbb{P} \left(\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \right) = 1.$$

Nous noterons ceci par $X_n \xrightarrow{p.s.} X$.

- X_n converge en probabilité vers X si, pour tout $\epsilon > 0$, nous avons

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

Nous noterons ceci par $X_n \xrightarrow{\mathbb{P}} X$.

Rappelons quelques résultats concernant ces deux modes de convergences (cf.[22]).

Proposition 1.1.1. La convergence presque sûre de X_n vers X lorsque $n \rightarrow +\infty$ entraîne la convergence en probabilité de X_n vers X .

Remarque. Ceci peut s'observer facilement en remarquant que la convergence presque sûre de X_n vers X peut se formuler comme suit, pour tout $\epsilon > 0$,

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\sup_{n \geq m} d(X_n, X) \geq \epsilon \right) = 0.$$

La présence du supremum rend cette convergence plus contraignante que la convergence en probabilité.

Bien que la convergence en probabilité n'implique pas, en général, une convergence presque sûre, il existe tout de même un résultat allant dans ce sens.

Proposition 1.1.2. Si X_n converge en probabilité vers X alors il existe une sous-suite $(n_k)_{k \geq 0}$ telle que X_{n_k} converge presque sûrement vers X .

Pour établir ce résultat, il faut employer le lemme de Borel-Cantelli qui est un outil essentiel permettant d'établir un résultat de convergence presque sûre ; nous rappelons son énoncé ci-dessous.

Lemme 1.1.1 (Borel-Cantelli). Si pour tout $\epsilon > 0$,

$$\sum_{n \geq 0} \mathbb{P}(d(X_n, X) \geq \epsilon) < \infty$$

alors $X_n \rightarrow X$ p.s..

Remarque. Lorsque $(X_n)_{n \geq 0}$ est une suite de variables aléatoires mutuellement indépendantes et X une variable constante presque sûrement, le résultat précédent est alors une équivalence.

Nous allons enfin aborder un dernier mode de convergence, celui de la convergence en loi. Cette notion est plus faible que la convergence en probabilité ou la convergence presque sûre mais elle est très importante dans le domaine des probabilités. Tentons de motiver l'utilité de cette notion.

1.2 A propos de la convergence faible

Pour plus de détails concernant les quelques mots qui vont suivre, nous renvoyons le lecteur vers les ouvrages suivants [89, 204, 54, 171, 170].

De manière générale lorsque nous avons à disposition une suite $(z_n)_{n \in \mathbb{N}}$ dans un espace topologique, sa convergence peut-être obtenue en deux étapes. Dans un premier temps, il est possible de montrer que l'ensemble $\{z_n : n \in \mathbb{N}\}$ est relativement compact, puis, dans un second temps, il suffit de montrer que l'ensemble des valeurs d'adhérence est réduit à un unique point. Afin d'établir la première étape, il est donc important de connaître les ensembles compacts de l'espace étudié.

Dans des espaces de Banach de dimension finie, la compacité d'un ensemble est simple à établir car ces ensembles sont complètement caractérisés par le théorème de Riesz (cf. [13]) : il s'agit d'ensembles fermés et bornés. En dimension infinie, cette caractérisation n'est plus valable et il devient plus délicat d'identifier les ensembles compacts. Les théorèmes établissant ce genre de résultat sont alors plus compliqués. A titre d'exemple, nous pouvons mentionner le théorème d'Arzela-Ascoli qui fournit un critère de compacité relative dans $(C^0[0, 1]; \|\cdot\|_\infty)$ ou encore celui de Riesz-Fréchet-Kolmogorov pour l'espace $(L^p(\Omega), \|\cdot\|_p)$ avec $\Omega \subset \mathbb{R}^n$ et $1 \leq p < \infty$. (cf. [54])

Une des raisons expliquant la difficulté précédente est la suivante : la convergence classique dans un espace vectoriel normé (induite par la topologie des boules ouvertes définies via la norme de l'espace vectoriel) est très contraignante. De plus la topologie induite par cette norme (aussi appelée topologie forte) ne capture pas tous les modes de convergence (convergence ponctuelle ou convergence sur les ensembles compacts par exemple). Afin de contourner cette difficulté, il est possible d'introduire des notions plus faibles de convergence (qui seront induites par des topologies moins fines). Il sera alors possible d'obtenir plus facilement un résultat de convergence faible qui pourra éventuellement être renforcé en une convergence forte à l'aide d'arguments supplémentaires. Il est également possible de laisser de côté la convergence forte pour se focaliser uniquement sur la notion de convergence faible.

A présent, procédons à des rappels concernant les notions évoquées ci-dessus.

De manière traditionnelle, les topologies faibles sont introduites via l'espace dual E' d'un espace vectoriel normé E . Rappelons que E' est l'espace vectoriel composé des formes linéaires continues $f : E \rightarrow \mathbb{R}$.

Définition 1.2.1. Soient E un espace vectoriel normé et E' son espace dual.

1. Une suite $(x_n)_{n \in \mathbb{N}}$ de E converge faiblement vers $x \in E$ si pour tout $f \in E'$

$$\lim_{n \rightarrow +\infty} f(x_n) = f(x).$$

Lorsque $n \rightarrow +\infty$, ceci sera noté par $x_n \rightharpoonup x$.

2. Une suite $(f_n)_{n \in \mathbb{N}}$ de E' converge faiblement-* vers $f \in E'$ si pour tout point $x \in E$

$$\lim_{n \rightarrow +\infty} f_n(x) = f(x).$$

Lorsque $n \rightarrow +\infty$, ceci sera noté par $f_n \xrightarrow{*} f$.

- Remarque.* 1. La notation « faiblement- $*$ » (lire *faiblement étoile*) provient de la littérature anglaise pour rappeler que la convergence a lieu dans l'espace dual E^* (que nous avons noté E').
2. Bien entendu, en dimension finie les convergences faibles et fortes coïncident ; cette affirmation est loin d'être évidente et découle d'un théorème que le lecteur pourra trouver dans [54] par exemple. En revanche, en dimension infinie, ces différents modes de convergences induisent de nouvelles topologies, moins fines, sur les espaces E et E' . Nous n'entrerons pas dans la description de celles-ci et renvoyons le lecteur vers l'ouvrage [54] pour plus de détails.

L'intérêt principal derrière cet affaiblissement de la topologie traditionnelle est le suivant : puisque la topologie est moins fine, il y a moins d'ensembles ouverts et donc plus d'ensembles compacts. Par exemple, le résultat suivant nous sera extrêmement utile au cours de ce chapitre. Ceci sera mis en valeur ultérieurement via le théorème de Banach-Alaoglu 1.2.1 dont nous rappelons l'énoncé.

Théorème 1.2.1 (Banach-Alaoglu). *Soit $(E, \|\cdot\|)$ un espace de Banach ; les boules fermées de l'espace dual E' sont faiblement- $*$ compactes. En outre, elles sont métrisables si E est un espace séparable.*

Remarque. Autrement dit, lorsque E est séparable, si $(f_n)_{n \in \mathbb{N}} \subset E'$ est bornée,

$$\text{i.e.} \quad \sup_{n \in \mathbb{N}} \|f_n\|_{E'} < +\infty,$$

alors il existe $f \in E'$ ainsi qu'une sous-suite $(f_{n_k})_{k \in \mathbb{N}}$ qui converge faiblement vers f . Autrement dit, pour tout $x \in E$, nous avons

$$f_{n_k}(x) \rightarrow f(x) \quad \text{lorsque} \quad k \rightarrow +\infty.$$

A présent, nous allons développer ces notions de convergences faibles dans le cadre de la convergence de variables aléatoires.

A ce propos, nous indiquons au lecteur que la convergence de variables aléatoires peut s'exprimer de manière équivalente en termes de mesures de probabilités (i.e. les lois associées aux variables aléatoires). Nous passerons donc fréquemment du point de vue probabiliste (avec des variables aléatoires $(X_n)_{n \geq 0}$) au point de vue de l'analyse fonctionnelle (avec des mesures de probabilités $(\mu_n)_{n \geq 0}$).

Sachant que nous allons étudier la convergence de mesures et au vu de ce qui précède, il est important de rappeler des théorèmes de structures concernant ce genre d'espace ; en particulier, nous allons nous focaliser sur des résultats de dualité. La plupart des démonstrations seront admises, le lecteur pourra trouver les preuves dans [204, 171, 170].

1.3 Dualité et structure des espaces de mesures

Tout d'abord, précisons quels types de mesures nous allons étudier dans ce livre.

1.3.1 Mesures signées et bornées, mesures de Radon

Un bref moment de réflexion nous montre qu'un ensemble de mesures positives ne forme pas un espace vectoriel. Il est donc nécessaire de considérer un ensemble légèrement plus grand. Nous débutons donc par la définition d'une mesure signée et bornée.

Définition 1.3.1. *Etant donné un espace mesurable $(E, \mathcal{A})$, une mesure signée μ est une fonction $\mu : \mathcal{A} \rightarrow [-\infty, \infty]$ telle que*

1. $\mu(\emptyset) = 0$.
2. Pour toute famille dénombrable d'ensembles disjoints $(A_n)_{n \geq 0}$ de $\mathcal{A}$, nous avons

$$\mu\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n \geq 0} \mu(A_n).$$

Remarque. 1. Il faut prendre garde au fait que les sommes intervenant dans la définition (ci-dessus) doivent toutes être définies. Ainsi, une mesure signée ne peut prendre au plus qu'une des deux valeurs $+\infty$ ou $-\infty$ (cf. [89]).

2. Une mesure signée est dite bornée s'il existe $M \in \mathbb{R}$ telle que $|\mu(A)| \leq M$ pour tout $A \in \mathcal{A}$.
3. L'ensemble des mesures signées et bornées sera désigné par $\mathcal{M}(E)$.

Le théorème suivant nous assure que les mesures signées et bornées s'obtiennent toujours comme la différence de deux mesures positives.

Théorème 1.3.1 (Décomposition de Jordan-Hahn). 1. Soit μ une mesure signée sur $(E, \mathcal{A})$ alors il existe un unique couple (μ^+, μ^-) de mesures positives, mutuellement singulières, dont l'une au moins est bornée, tel que $\mu = \mu^+ - \mu^-$.

2. Le couple (μ^+, μ^-) correspond la décomposition de Jordan-Hahn de μ .
3. μ est bornée si μ^+ et μ^- le sont.
4. La quantité $|\mu| = \mu^+ + \mu^-$ est la mesure de variation totale de μ .

Remarque. Rappelons que deux mesures μ et ν sont dites singulières si il existe $A \in \mathcal{A}$ tel que pour tout $B \in \mathcal{A}$

$$\mu(B) = \mu(B \cap A) \quad \text{et} \quad \nu(B) = \nu(B \cap A^c).$$

Définissons également $\|\mu\|$ par $|\mu(E)| = \mu^+(E) + \mu^-(E)$; il s'agit d'une norme sur l'espace vectoriel $\mathcal{M}(E)$ des mesures signées bornées sur E . Pour plus de détails, nous renvoyons le lecteur vers [204].

Il est assez simple de construire des mesures signées à partir de la théorie de la mesure classique.

Exemple 1.3.1. Soient ν une mesure positive sur $(E, \mathcal{A})$ et $f \in L^1(\nu)$. En utilisant les parties positives et négatives de f , il suffit alors de poser

$$d\mu^+ = f^+ d\nu \quad \text{et} \quad d\mu^- = f^- d\nu$$

pour obtenir une mesure signée $\mu = (\mu^+, \mu^-)$. Notons au passage que la théorie de Lebesgue peut aisément s'étendre au cas des mesures signées.

Certaines mesures ont des rôles assez particuliers, c'est notamment le cas des mesures dites de Radon.

Définition 1.3.2. Soit E un espace séparé et localement compact. Une mesure positive μ sur $(E, \mathcal{B}(E))$ est dite de Radon si $\mu(K) < +\infty$ pour tout compact $K \subset E$.

Remarque. Rappelons qu'un espace topologique E est dit séparé si la propriété suivante est vérifiée : pour tout $x, y \in E$ distincts il existe un voisinage (ouvert) U de x et un voisinage (ouvert) V de y tels que $U \cap V = \emptyset$; E est dit localement compact si tout point $x \in E$ admet un voisinage dont l'adhérence est compacte.

Exemple 1.3.2. La mesure de Lebesgue est une mesure de Radon. Plus généralement, une mesure de Stieljes sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est également de Radon.

Les mesures de Radon vérifient une propriété de régularité.

Proposition 1.3.1 (Régularité). Soit E un espace séparé, localement compact et σ -compact. Si μ est une mesure de Radon sur $(E, \mathcal{B}(E))$ alors μ est régulière. C'est-à-dire, pour tout borélien $B \in \mathcal{B}(E)$,

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

Remarque. Dire qu'un espace topologique E est σ -compact signifie que E peut s'écrire comme une réunion dénombrable d'ensembles compacts.

1.3.2 Espace dual

Par la suite nous allons être amenés à considérer des espaces de fonctions continues sur un espace E . Dans ce qui va suivre nous supposons que E est un espace séparé, localement compact et σ -compact. Nous adopterons donc les notations suivantes :

- $C^0(E)$ pour désigner l'ensemble des fonctions continues sur E ;
- $C_b(E)$ pour désigner l'ensemble des fonctions continues et bornées sur E ;
- $C_0(E)$ pour désigner l'ensemble des fonctions continues sur E qui tendent vers 0 à l'infini ;
- $C_K(E)$ pour désigner l'ensemble des fonctions continues sur E à support compact.

Remarque. Les inclusions suivantes sont satisfaites $C_K(E) \subset C_0(E) \subset C_b(E) \subset C^0(E)$ avec égalité lorsque E est un ensemble compact.

Voyons ce qu'il est déjà possible d'établir (en termes de dualité) concernant l'ensemble des mesures positives de Radon. Tout d'abord, observons le fait suivant. Si μ est une mesure positive de Radon sur $(E, \mathcal{B}(E))$, il est possible de lui associer la forme linéaire φ_μ sur $C_K(E)$ définie par

$$\varphi_\mu(f) = \int_E f d\mu \text{ avec } f \in C_K(E).$$

De plus, cette forme linéaire est dite positive : si $f \geq 0$ alors $\varphi_\mu(f) \geq 0$. Le théorème suivant nous assure qu'il s'agit de la seule manière de définir une forme linéaire positive sur $C_K(E)$.

Théorème 1.3.2 (dualité de Riesz pour les mesures de Radon). *Soient E est un espace séparé, localement compact et σ -compact. Si φ une forme linéaire positive sur $C_K(E)$ alors il existe une unique mesure positive de Radon μ sur $(E, \mathcal{B}(E))$ telle que*

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

Il est également possible d'obtenir un résultat similaire concernant les formes linéaires sur $C_0(E)$, il s'agit du théorème suivant.

Théorème 1.3.3 (dualité de Radon-Riesz pour les mesures signées). *Soit E est un espace séparé, localement compact et σ -compact. Si $\varphi : C_0(E) \rightarrow \mathbb{R}$ est une forme linéaire continue alors il existe $\mu \in \mathcal{M}(E)$ telle que*

$$\varphi(f) = \int_E f d\mu \quad \text{avec} \quad f \in C_0(E)$$

et $\|\varphi\| = \|\mu\|$. Autrement dit, $C_0(E)' = \mathcal{M}(E)$.

Remarque. Si E est un ensemble compact, le résultat précédent se résume en $C^0(E)' = \mathcal{M}(E)$.

Ces résultats de dualité, entre espaces de fonctions continues et espaces de mesures, combinés au théorème de Banach-Alaoglu sont particulièrement utiles pour établir des résultats de convergences de mesures. Notons que, dans ce qui suit, $\mathcal{M}(E)$ sera équipé de la topologie faible-*. Voyons ce que fournit la combinaison des résultats précédents lorsque $E = \mathbb{R}^d$.

Si $(\mu_n)_{n \in \mathbb{N}}$ est une suite bornée dans $\mathcal{M}(\mathbb{R}^d)$ (i.e. $\sup_{n \in \mathbb{N}} \|\mu_n\| < +\infty$) alors, d'après le théorème de Banach-Alaoglu 1.2.1 il existe $\mu \in \mathcal{M}(\mathbb{R}^d)$ et $(\mu_{n_k})_{k \in \mathbb{N}}$ telles que, pour toute fonction $f \in C_0(\mathbb{R}^d)$, nous avons la convergence suivante

$$\int_{\mathbb{R}^d} f d\mu_{n_k} \rightarrow \int_{\mathbb{R}^d} f d\mu \quad \text{lorsque} \quad k \rightarrow +\infty.$$

En particulier, ceci s'applique à des suites de mesures de probabilités $(\mu_n)_{n \in \mathbb{N}}$ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Il faut toutefois être prudent avec cette limite faible $\mu \in \mathcal{M}(\mathbb{R}^d)$. Nécessairement, il s'agit une mesure positive de norme inférieure ou égale à 1, néanmoins ce n'est pas forcément une mesure de probabilité.

Exemple 1.3.3. Dans $\mathbb{R}$, considérons la suite de mesures $\mu_n = \delta_n$ avec $n \in \mathbb{N}$. Alors, d'après ce qui précède, il existe une sous-suite $(\mu_{n_k})_{k \geq 0}$ telle que pour toute fonction $f \in C_0(\mathbb{R})$ nous avons

$$\int_{\mathbb{R}} f d\mu_{n_k} = f(n_k) \rightarrow 0 = \int_{\mathbb{R}} f d\mu \quad \text{lorsque} \quad k \rightarrow +\infty.$$

Autrement dit, $\mu = 0$; la limite n'est pas une mesure de probabilité car la masse s'échappe à l'infini.

Malheureusement, comme le montre l'exemple précédent, la convergence faible ne préserve pas les mesures de probabilités. Nous allons alors devoir considérer une notion de convergence un peu plus restrictive. A cet effet, par dualité, il suffit de considérer un espace de fonction plus grand pour obtenir un espace de mesure plus petit (ne contenant que des mesures de probabilités).

1.4 Convergence faible et convergence étroite dans $\mathbb{R}^d$

Le lecteur souhaitant avoir plus de détails sur les notions étudiées ci-dessous est prié de se référer à [22]. Considérons maintenant l'espace $E = \mathbb{R}^d$ muni de sa tribu borélienne $\mathcal{B}(\mathbb{R}^d)$ ainsi que $(\mu_n)_{n \geq 1}$ et μ des mesures positives. Par la suite, nous poursuivons l'étude de la convergence faible de μ_n vers μ entamée plus tôt ; ceci nécessite de formaliser quelques définitions dans ce cadre.

Définition 1.4.1. Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures positives bornées sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

1. Nous dirons que $(\mu_n)_{n \in \mathbb{N}}$ converge faiblement vers μ si pour toute fonction $f \in C_0(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu \quad \text{lorsque } n \rightarrow +\infty.$$

Ceci sera noté par $\mu_n \rightharpoonup \mu$.

2. Nous dirons que $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ si pour toute fonction $f \in C_b(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu \quad \text{lorsque } n \rightarrow +\infty.$$

Ceci sera noté par $\mu_n \Rightarrow \mu$.

Remarque. Puisque $C_0(X) \subset C_b(X)$, la convergence étroite implique la convergence faible. La réciproque est fautive comme nous l'avons vu dans l'exemple 1.3.3. Dans la littérature, il est également possible de rencontrer la notion de convergence vague : $(\mu_n)_{n \in \mathbb{N}}$ converge vaguement vers μ si pour toute fonction $f \in C_K(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu \quad \text{lorsque } n \rightarrow +\infty.$$

Cependant, nous n'étudierons pas cette notion dans ce cours.

Faisons un bref aparté concernant le vocabulaire associé à l'étude de la convergence de mesure ; laquelle a pour vocation de traiter la convergence de variable aléatoires.

Définition 1.4.2. Soient $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Désignons la loi de X_n par μ_n , et celle de X par μ . Les assertions suivantes sont alors équivalentes :

- $X_n \rightarrow X$ en loi lorsque $n \rightarrow +\infty$ (nous pourrions noter ceci par $X_n \xrightarrow{\mathcal{L}} X$) ;
- $\mu_n \Rightarrow \mu$ étroitement lorsque $n \rightarrow +\infty$;
- pour toute fonction $f \in C_b(\mathbb{R}^d)$, $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ lorsque $n \rightarrow +\infty$.

1.4.1 Hiérarchie des modes de convergence

Voyons où se place la convergence en loi (ou convergence étroite) par rapport à la convergence presque sûre ou la convergence en probabilité. Rappelons au passage que la convergence presque sûre entraîne la convergence en probabilité.

Proposition 1.4.1. Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et $p \in \mathbb{R}^d$. Si $X_n \xrightarrow{\mathcal{L}} p$ lorsque $n \rightarrow +\infty$ alors $X_n \xrightarrow{\mathbb{P}} p$ lorsque $n \rightarrow +\infty$.

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

Proposition 1.4.2. Avec les notations introduites précédemment, si $X_n \xrightarrow{\mathbb{P}} X$ alors $X_n \xrightarrow{\mathcal{L}} X$ lorsque $n \rightarrow +\infty$.

Démonstration. Soit $(n_k)_{k \in \mathbb{N}}$ une sous-suite. Par hypothèse, $X_n \xrightarrow{\mathbb{P}} X$ alors, d'après la proposition 1.1.2, il existe alors une sous-suite $(n_{k_l})_{l \in \mathbb{N}}$ telle que

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

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

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

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

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

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

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

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

ce qui est absurde. □

La convergence en loi est donc le mode de convergence le plus faible.

1.4.2 Critère de convergence étroite dans $\mathbb{R}^d$

Nous allons voir qu'il est possible d'obtenir de la convergence étroite à partir de la convergence faible, à condition qu'il y ait aussi convergence des masses (ceci assure que le cas de figure décrit dans l'exemple 1.3.3 ne se produise pas).

Proposition 1.4.3. Si $(\mu_n)_{n \in \mathbb{N}}$ converge faiblement vers μ et si $\lim_{n \rightarrow +\infty} \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$ alors $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ .

Démonstration. La démonstration est laissée à titre d'exercice □

Voici un critère suffisant, permettant d'assurer qu'une suite $(\mu_n)_{n \in \mathbb{N}}$ converge faiblement.

Proposition 1.4.4. *Pour qu'une suite $(\mu_n)_{n \in \mathbb{N}}$ converge faiblement, il suffit que cette suite soit bornée*

$$\text{i.e.} \quad \sup_{n \in \mathbb{N}} \mu_n(\mathbb{R}^d) < +\infty$$

et que pour toute fonction f appartenant à un ensemble total de $C_0(\mathbb{R}^d)$ la suite $\left(\int_{\mathbb{R}^d} f d\mu_n \right)_{n \in \mathbb{N}}$ converge.

Remarque. Rappelons qu'un ensemble total d'un espace de Banach correspond à un sous-espace vectoriel dense. Par exemple,

1. $C_K(\mathbb{R}^d)$ est un ensemble total de $C_0(\mathbb{R}^d)$.
2. La famille de fonctions $x \mapsto \exp(-a|x|^2 - \langle b, x \rangle)$ avec $a \in \mathbb{R}_+^*$ et $b \in \mathbb{R}^d$ engendre également un espace total. Il est même possible de supposer que $a \in \mathbb{Q}_+$ et $b \in \mathbb{Q}^d$.

Revenons sur l'utilisation du théorème de Banach-Alaoglu 1.2.1 pour obtenir l'existence d'une sous-suite convergente dans un tel contexte.

Théorème 1.4.1. *De toute suite $(\mu_n)_{n \in \mathbb{N}}$ de mesures positives sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que $\sup_{n \in \mathbb{N}} \mu_n(\mathbb{R}^d) < +\infty$, il est possible d'extraire une sous-suite qui converge faiblement.*

Démonstration. Nous avons déjà vu qu'il suffisait d'appliquer le théorème de Banach-Alaoglu 1.2.1 pour obtenir le résultat. En effet, celui-ci nous assure qu'il existe une sous-suite $(\mu_{n_k})_{k \geq 0}$ qui converge faiblement vers une mesure positive μ .

Présentons également une démonstration plus intuitive du théorème précédent. Pour cela, considérons $(f_p)_{p \in \mathbb{N}}$ une suite totale dans $C_0(\mathbb{R}^d)$. Nous devons alors trouver une sous-suite $(\mu_{n_k})_{k \in \mathbb{N}}$ telle que

$$\int_{\mathbb{R}^d} f_p d\mu_{n_k} \quad \text{converge pour tout} \quad p \in \mathbb{N}. \quad (1.4.1)$$

Pour $p = 1$ considérons la suite de nombres réels $(\int_{\mathbb{R}^d} f_1 d\mu_n)_{n \in \mathbb{N}}$. Puisque μ_n est bornée (par hypothèse) et f_1 est bornée en tant qu'élément de $C_0(\mathbb{R}^d)$, il est possible d'appliquer le théorème de Bolzano-Weierstrass (cf. [13]). Celui-ci nous assure qu'il est possible d'extraire une sous-suite $(\mu_n^{(1)})_{n \in \mathbb{N}}$ telle que $(\int_{\mathbb{R}^d} f_1 d\mu_n^{(1)})_{n \in \mathbb{N}}$ converge.

Poursuivons le même raisonnement, lorsque $p = 2$, avec $(\int_{\mathbb{R}^d} f_2 d\mu_n^{(1)})_{n \in \mathbb{N}}$. Par les mêmes arguments, nous pouvons extraire une sous-suite, notée $(\mu_n^{(2)})_{n \in \mathbb{N}}$, de la suite précédente $(\mu_n^{(1)})_{n \in \mathbb{N}}$ telle que $(\int_{\mathbb{R}^d} f_1 d\mu_n^{(2)})_{n \in \mathbb{N}}$ converge. Il suffit de reproduire cet argument pour toutes les valeurs de $p \geq 3$. L'argument diagonal de Cantor nous assure que ce procédé nous permet d'obtenir la sous-suite initialement désirée (i.e. une sous-suite vérifiant (1.4.1)). $\square$

La notion qui va suivre est essentielle. Cette propriété est en quelque sorte l'analogue de la notion d'équicontinuité (apparaissant dans le théorème d'Arzela-Ascoli et permettant d'obtenir de la compacité relative) pour les espaces de mesures bornées. En particulier, lorsque cette propriété est satisfaite, nous pourrions justifier que la limite faible (obtenue via le théorème de Banach-Alaoglu) est une mesure de probabilité.

Définition 1.4.3. Une partie $\mathcal{P} \subset \mathcal{M}(\mathbb{R}^d)$ de mesures positives sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est dite tendue si

1. $\sup_{\mu \in \mathcal{P}} \|\mu\| = \sup_{\mu \in \mathcal{P}} \mu(\mathbb{R}^d) < +\infty$,
2. Pour tout $\epsilon > 0$, il existe K un compact de $\mathbb{R}^d$ tel que, pour tout $\mu \in \mathcal{P}$,

$$\mu(K^c) \leq \epsilon.$$

Remarque. Les deux faits suivants sont vérifiés et permettent de visualiser plus facilement quelles formes peuvent prendre des parties tendues de $\mathcal{M}(\mathbb{R}^d)$.

1. Si $\mathcal{P}$ est une partie finie de $\mathcal{M}(\mathbb{R}^d)$ alors $\mathcal{P}$ est tendue.
2. Si $(\mu_n)_{n \geq 1}$ est une suite de mesures convergeant étroitement alors $\mathcal{P} = (\mu_n)_{n \geq 1}$ est tendue.

Le résultat suivant est essentiel et sera étendu à un cadre plus général par la suite. Celui-ci induit que l'étude de la convergence étroite se ramène à la recherche d'ensembles compacts de l'espace sous-jacent via la notion de tension.

Proposition 1.4.5. De toute suite tendue $(\mu_n)_{n \in \mathbb{N}}$ de mesures de probabilité sur $\mathbb{R}^d$, il est possible d'en extraire une sous-suite qui converge étroitement.

Démonstration. D'après ce qui précède, nous savons qu'il existe une sous-suite $(n_k)_{k \geq 0}$ et ν une mesure positive telles que

$$\mu_{n_k} \xrightarrow[k \rightarrow +\infty]{} \nu.$$

Il suffit à présent d'utiliser la propriété de tension pour démontrer que ν est bien une mesure de probabilité vers laquelle μ_{n_k} converge étroitement. A cet effet, il suffit de démontrer que pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ continue telle que $0 \leq f \leq 1$,

$$\lim_{k \rightarrow +\infty} \int f d\mu_{n_k} = \int f d\nu.$$

Pour tout $M > 0$, introduisons l'ensemble compact K_M défini par

$$K_M = \{x \in \mathbb{R}^d \ ; \ |x| \leq M\}.$$

La propriété de tension, nous assure que pour tout $\epsilon > 0$, il existe $M > 0$ tel que $\mu_n(K_M) \geq 1 - \epsilon$ pour tout $n \in \mathbb{N}$. Soit $\psi_M(x) = \min([M + 1 - |x|]_+; 1)$ pour $x \in \mathbb{R}^d$ et notons que

$$1_{K_M}(x) \leq \psi_M(x) \leq 1.$$

Alors, pour tout $n \in \mathbb{N}$, nous avons

$$\int f d\mu_n \leq \int_{K_M} f d\mu_n + \epsilon \leq \int \psi_M f d\mu_n + \epsilon.$$

En particulier, puisque $\psi_M f \in C_0(\mathbb{R}^d)$, nous avons

$$\limsup_{k \rightarrow +\infty} \int f d\mu_{n_k} \leq \int \psi_M f d\nu + \epsilon \leq \int f d\nu + \epsilon.$$

En outre, si f est la fonction constante égale à 1, nous avons $\nu(\mathbb{R}^d) \geq 1 - \epsilon$. Puisque ν est un élément de la boule unité de $\mathcal{M}(\mathbb{R}^d)$, ceci implique que $\nu(\mathbb{R}^d) = 1$. De plus, ce fait permet de choisir M tel que $\nu(K_M) \geq 1 - \epsilon$. Ainsi,

$$\int f d\nu \leq \int_{K_M} f d\nu + \epsilon \leq \int \psi_M f d\nu + \epsilon$$

ce qui implique ensuite que

$$\int f d\nu - \epsilon \leq \liminf_{k \rightarrow +\infty} \int \psi_M f d\mu_{n_k} \leq \liminf_{k \rightarrow +\infty} \int f d\mu_{n_k}.$$

La conclusion s'ensuit. $\square$

Remarque. En dimension finie, la démonstration est facilitée par la caractérisation des ensembles compacts K_M intervenant dans la propriété de tension de la suite $(\mu_n)_{n \geq 0}$.

1.5 Convergence étroite dans $\mathbb{R}^d$ et transformation de Fourier

Il est bien connu que deux variables aléatoires X et Y définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ ont même loi si et seulement leurs fonctions caractéristiques coïncident. Voyons quelle influence a cette propriété vis-à-vis de la convergence en loi. Pour plus de détails, notamment les démonstrations, le livre [22] est tout indiqué. Concernant la terminologie, la transformée de Fourier fait référence à la fonction caractéristique d'une variable aléatoire. A ce propos, rappelons la définition suivante.

Définition 1.5.1. *Si μ est une mesure positive bornée sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, sa fonction caractéristique $\hat{\mu}$ est définie par*

$$\hat{\mu}(t) = \phi_\mu(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} d\mu(x) \quad \text{pour tout } t \in \mathbb{R}^d.$$

En dimension finie, il est possible d'obtenir une caractérisation de la convergence étroite via la convergence des fonctions caractéristiques.

Théorème 1.5.1 (Lévy). *Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures positives, bornées, sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que $(\phi_{\mu_n})_{n \in \mathbb{N}}$ converge simplement vers ϕ une fonction continue en 0. Alors, il existe μ une mesure positive bornée telle que*

$$\mu_n \Rightarrow \mu \quad \text{lorsque } n \rightarrow +\infty \quad \text{et} \quad \phi_\mu = \phi.$$

Remarque. Au vu de ce que nous avons observé dans la section précédente, pour démontrer ce résultat il nous faudra comprendre de quelle manière l'hypothèse portant sur la convergence des fonctions caractéristiques permet d'obtenir de la tension. Comme nous allons l'observer, le fait que $\mathbb{R}^d$ soit un espace de dimension finie joue un rôle essentiel. Le cas de la dimension infinie, plus délicat, sera traité dans le chapitre prochain.

L'inégalité suivante, due à Lévy, joue un rôle crucial pour établir le théorème éponyme. Pour alléger les notations, nous présentons son énoncé seulement en dimension 1. Soulignons le fait que cette inégalité n'est valable qu'en dimension finie, le choix de $E = \mathbb{R}^d$ est donc crucial ici.

Lemme 1.5.1 (Lévy). *Soit μ une mesure de probabilité. Alors, pour tout $u > 0$, l'inégalité suivante est satisfaite*

$$\mu\left(x \in \mathbb{R}; |x| \geq \frac{1}{u}\right) \leq \frac{7}{u} \int_0^u [1 - \operatorname{Re}(\phi_\mu(t))] dt.$$

Remarque. Si $z = a + ib \in \mathbb{C}$ alors $a = \operatorname{Re}(z)$ désigne la partie réelle de z .

Démonstration. Tout d'abord, observons que pour tout $|x| \geq 1$, nous avons

$$\frac{\sin(x)}{x} \leq \sin(1) \leq \frac{6}{7}.$$

Ensuite, grâce au théorème de Fubini-Tonelli, nous avons

$$\begin{aligned} \frac{1}{u} \int_0^u (1 - \operatorname{Re}(\phi_\mu(t))) dt &= \frac{1}{u} \int_0^u \int_{\mathbb{R}} (1 - \cos(tx)) d\mu(x) dt \\ &= \int_{\mathbb{R}} \left(1 - \frac{\sin(ux)}{ux}\right) d\mu(x) \\ &\geq (1 - \sin(1)) \mu\left(|x| \geq \frac{1}{u}\right). \end{aligned}$$

La conclusion s'ensuit grâce à l'observation faite au début de la démonstration. $\square$

C'est grâce à ce résultat qu'il est possible d'obtenir de la tension sous les hypothèses du théorème 1.5.1 de Lévy. En effet, si nous appliquons le lemme 1.5.1 à μ_n pour tout n , nous obtenons

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \mu_n(|x| \geq M) &\leq \limsup_{n \rightarrow +\infty} 7M \int_0^{1/M} [1 - \operatorname{Re}(\phi_{\mu_n}(t))] dt \\ &= 7M \int_0^{1/M} [1 - \operatorname{Re}(\phi(t))] dt. \end{aligned}$$

Le majorant obtenu peut être rendu arbitrairement petit en choisissant M arbitrairement grand puisque, par hypothèse, $\lim_{t \rightarrow 0} \operatorname{Re}(\phi(t)) = 1$ et $t \mapsto \phi(t)$ est une fonction continue. En conséquence, ceci permet d'obtenir un compact de $\mathbb{R}$ qui contient presque toute la masse et assure la tension de la suite $(\mu_n)_{n \geq 0}$.

Une application célèbre du théorème de Lévy est celle du théorème de la limite centrale (cf. [22]).

Rappelons au passage que l'abréviation *i.i.d.* associés à une famille de variables aléatoires signifie qu'elles sont *indépendantes et identiquement distribuées*; le symbole $\sim$ apparaissant dans

$$X \sim \mathcal{N}(0, 1)$$

signifie que la variable aléatoire X suit une loi normale centrée et réduite. Autrement dit, X admet la densité (par rapport à la mesure de Lebesgue) suivante

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{pour tout } x \in \mathbb{R}.$$

Enfin, si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire centré de $\mathbb{R}^d$ sa structure de covariance $\text{Cov}(X)$ correspond à une matrice symétrique $(\Gamma_{ij})_{1 \leq i, j \leq d}$ avec

$$\Gamma_{ij} = \mathbb{E}[X_i X_j] \quad \text{pour tout } i, j \in \{1, \dots, d\}.$$

Théorème 1.5.2 (Limite centrale). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Supposons que les variables aléatoires $X_1, \dots, X_n$ soient toutes i.i.d. (de même loi que X_1), que $\mathbb{E}[X_1] = 0$ et $\mathbb{E}[|X_1|^2] < +\infty$ alors*

$$\frac{S_n}{\sqrt{n}} \rightarrow \mathcal{N}_d(0, \text{Cov}(X_1)) \quad \text{en loi lorsque } n \rightarrow +\infty,$$

où, pour $n \geq 1$, $S_n = X_1 + \dots + X_n$ avec $\text{Cov}(X_1)$ la matrice de covariance associée au vecteur X_1 .

Démonstration. Établissons ce résultat lorsque $d = 1$. Sans perdre en généralité, il est possible de supposer que $\mathbb{E}[X_1^2] = 1$. Observons que si $G \sim \mathcal{N}(0, 1)$ alors $\phi_G(t) = \mathbb{E}[e^{itG}] = e^{-\frac{t^2}{2}}$. Il n'est pas difficile de montrer ensuite, sous l'hypothèse i.i.d. des variables aléatoires, que l'égalité suivante (en loi) est satisfaite

$$\phi_{\frac{S_n}{\sqrt{n}}}(t) = \phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n.$$

Ainsi,

$$\lim_{n \rightarrow +\infty} \phi_n(t) = e^{-\frac{t^2}{2}} \quad \text{pour tout } t \in \mathbb{R}.$$

Il suffit ensuite de conclure grâce au théorème de Lévy 1.5.1. □

Remarque. La démonstration précédente n'est pas complètement rigoureuse car la fonction caractéristique est à valeurs complexes. Il est alors nécessaire de démontrer le fait suivant : si $(z_n)_{n \geq 0}$ est une suite de nombre complexes tendant vers 0 alors

$$\lim_{n \rightarrow +\infty} \left(1 + \frac{z_n}{n}\right)^n = 1.$$

Le lecteur trouvera les détails de ceci dans [22].

1.6 Autres caractérisations de la convergence étroite

Dans cette section, nous présentons d'autres moyens permettant d'établir un résultat de convergence étroite.

Les lois de probabilités étant aussi caractérisées par leurs fonctions de répartition, il est également possible de les employer pour obtenir de la convergence en loi. Pour simplifier, nous ne traitons que le cas réel (les notations dans $\mathbb{R}^d$ étant plus lourdes).

Théorème 1.6.1 (Helly-Bray). *Soient $(X_n)_{n \geq 0}$ et X des variables aléatoires de fonctions de répartition respectives $t \mapsto F_{X_n}(t) = \mathbb{P}(X_n \leq t)$ et $t \mapsto F_X(t) = \mathbb{P}(X \leq t)$. Alors nous avons le résultat suivant :*

$$X_n \xrightarrow{\mathcal{L}} X \iff \lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t) \text{ pour tout } t \in D_{F_X}$$

où D_{F_X} désigne l'ensemble des points de continuité de F_X .

De manière surprenante, il est également possible de relier la convergence en loi à une convergence presque sûre (à condition de travailler sur un autre espace de probabilité).

Théorème 1.6.2 (Skorokhod). *Dans $\mathbb{R}^d$, avec les notations précédentes, $\lim_{n \rightarrow +\infty} X_n \stackrel{\mathcal{L}}{=} X$ si et seulement si il existe des variables aléatoires $(X'_n)_{n \geq 0}$ et X' définies sur un autre espace de probabilité $(\Omega', \mathcal{A}', \mathbb{P}')$ telles que*

1. $X'_n = X_n$ et $X' = X$ en loi;

2. $X_n \xrightarrow{p.s.} X$.

Ceci conclut ce chapitre. Dans le suivant, nous allons voir de quelle manière étendre l'étude de convergence de mesure à des espaces de Banach de dimension infinie.

1.7 Références historiques

Les quelques idées et évènements qui suivent ont pour but de dresser un panorama historique de l'évolution de la notion de dimension infinie et d'espaces fonctionnels ainsi que les questions mathématiques naturellement liées à de tels concepts. Le lecteur pourra compléter ce court descriptif à l'aide des ouvrages [83, 45, 171, 95, 96] dont nous nous sommes inspirés.

Jusqu'en 1830, les mathématiciens de l'époque travaillaient principalement sur des systèmes linéaires d'équations dans $\mathbb{R}^n$. En 1850 la notion de matrice émergea via les travaux de Cayley et de Sylvester, ceci conduisit à l'étude de réduction d'endomorphismes ou de forme quadratiques. En fait, dès la fin du 19^{ième} siècle, les résultats essentiels d'algèbre linéaire et bilinéaire étaient connus mais reposaient sur des notions encore floues et mal comprises, ce qui ne facilitait pas le passage de la dimension finie à la dimension infinie.

Par exemple, la notion de fonction ne commença à se développer qu'au début du 19^{ième} siècle. Il fallut attendre les travaux de Riemann, fin 19^{ième}, pour considérer des espaces de fonctions muni de différentes topologies. A ce propos, au tout début, la différence entre convergence faible et convergence uniforme n'était pas bien identifiée par Cauchy. Il fallu patienter jusqu'aux années 1850 pour qu'une véritable définition soit adoptée; nous la devons à Stokes et Seidel (en 1847 – 1848) et, de manière indépendante, à Cauchy en 1853.

L'étude de la convergence uniforme et des questions connexes fut au centre des préoccupations du moment. L'expansion des travaux sur ces thématiques fut portée par les écoles allemandes (citons, par exemple, Hankel ou encore du Bois-Reymond) et italiennes (les mathématiciens Dini et Arzela notamment). De manière contemporaine, Ascoli introduisit la notion d'équicontinuité qui lui permet de caractériser les sous-ensembles relativement compacts des espaces de fonctions

continues (sur un ensemble compact).

La première apparition de la dimension infinie, provint de l'étude de l'équation de la chaleur par Fourier. Dans son étude, il s'attacha à résoudre des systèmes infinis d'équations impliquant des polynômes trigonométriques.

A la fin du 19^{ième} siècle, la notion d'espace de fonctions ainsi que les différentes topologies qu'il était possible de leur associer commençaient à se répandre dans le monde mathématique. Entre 1900 et 1910 l'accumulation de ces idées commença à cristalliser et à donner naissance à de nouveaux concepts. Les travaux suivants furent source de nombreux développements et jouèrent un rôle prépondérant dans l'histoire de l'analyse fonctionnelle.

- les travaux de Fredholm (en 1900) sur des équations intégrales,
- la thèse de Lebesgue (en 1902) permettant d'étendre la notion d'intégrale de Riemann à un cadre beaucoup plus général (via la notion d'ensemble négligeable et de mesurabilité),
- les résultats d'Hilbert (en 1906) en théorie spectrale,
- la thèse de Fréchet (en 1906) portant sur l'étude des espaces métriques.

Les travaux d'Hilbert donnèrent naissance aux espaces qui portent désormais son nom et mirent en évidence la notion de structure. En effet, jusqu'à la moitié du 19^{ième} siècle les objets mathématiques étaient essentiellement concrets : nombres, points, courbes, surfaces, fonctions, opérateurs, . . . Le monde mathématique prit ensuite conscience que la nature des objets importait moins que les règles permettant de les manipuler.

Ceci permit à la notion d'espace vectoriel normé et de dualité de naître. Fréchet se mit à étudier des espaces de fonctions muni de la norme uniforme et étudia la notion de forme linéaire continue sur ces espaces. Dans le cadre des espaces de Hilbert, la caractérisation des formes linéaires continues fut obtenue, de manière indépendante, par Fréchet et Riesz. Ensuite, Riesz continua l'étude des formes linéaires continues sur les espaces L^p et l^p pour $p > 1$. Il établit par ailleurs la dualité, désignée sous une autre terminologie à l'époque, entre les espaces L^q et L^p (pour des exposants p et q conjugués).

En 1911, Riesz voulu étudier les formes linéaires de $C[a, b]$ avec $[a, b]$ un intervalle compact. Il souhaitait résoudre des équations de la forme

$$\int_a^b g_\alpha(x) d\xi(x) = c_\alpha \quad \text{avec} \quad \alpha \in I$$

où $(g_\alpha)_{\alpha \in I}$ est une famille de fonctions continues et $(c_\alpha)_{\alpha \in I}$ des scalaires donnés. L'inconnue de son problème était alors ξ une fonction à variation bornées. Dans la démonstration de la résolution de ce problème, Riesz laisse apparaître l'idée de construire une forme linéaire sur un sous-espace, puis de la prolonger cette même forme à l'espace tout entier. Dans la terminologie actuelle, ce procédé s'apparente au contenu du théorème de Hahn-Banach.

A la suite des travaux de Riesz, Helly s'intéressa à la convergence de suites dans des espaces vectoriels. A la différence des exemples concrets (l^p ou L^p pour $p > 1$) étudié par Riesz, Helly se plaça dans un cadre abstrait plus général. Il fut, parmi d'autres mathématiciens, à l'origine de la notion de dualité et l'idée de prolongement d'une forme linéaire était également présente dans ses travaux.

La synthèse des travaux de Riesz et d'Helly permit d'extirper la notion générale d'espaces vectoriels normés. Cette nouvelle notion permit à Hahn et Banach de travailler sur des espaces vectoriels normés complets afin d'étudier les points suivants :

- opérateurs linéaires,
- formes linéaires,
- notion de contraction.

A la suite de cela, Hahn introduisit, pour la première fois de façon formelle, la définition de l'espace dual d'un espace vectoriel. Tout ce qui précède donna alors lieu de nombreux développements :

- la notion d'espace métrique complet, de fonctions semi-continues inférieurement,
- le lemme de Baire,
- le théorème de Banach-Steinhaus,
- ...

En 1932, Banach écrivit un ouvrage qui permit de faire la synthèse des travaux des dernières décennies. Dans son livre, la notion la convergence faible y est grandement mis en avant ; Banach y présente également son fameux théorème portant sur des conditions nécessaires et suffisantes pour assurer la continuité d'une application linéaire entre deux espaces de Banach.

Chapitre 2

Convergence de mesures dans des espaces polonais

Nous allons maintenant étudier la convergence en loi de suites de variables aléatoires $(X_n)_{n \geq 1}$ à valeurs dans un espace polonais E (i.e. un espace métrique complet et séparable); de manière alternative, nous allons étudier sur la convergence étroite des lois $(\mu_n)_{n \geq 1}$ associées aux variables aléatoires $(X_n)_{n \geq 1}$. Contrairement à ce que nous avons présenté dans le chapitre précédent, dans ce chapitre, E sera un espace de Banach ou de Hilbert de dimension infinie.

Nous débuterons ce chapitre en étudiant les propriétés de la convergence étroite dans ce nouveau cadre et en observant les différences avec celui de la dimension finie. Dans la section 2.3, nous établirons de nouveaux critères (en particulier, le théorème du portemanteau 2.3.1) permettant d'établir de la convergence étroite dans le cadre des espaces métriques polonais. La section 2.4 sera l'occasion d'étudier cette convergence d'un point de vue métrique en introduisant différentes distances (celles de Lévy-Prokhorov et de Kantorovich-Rubinstein). Nous nous attarderons sur la topologie (induites par ces distances) de l'espace des mesures de probabilités $\mathcal{P}(E)$; le point culminant de ce chapitre est le théorème de Prokhorov 2.4.3. La section 2.5 consistera à illustrer cette approche abstraite lorsque $E = C^0[0, 1]$, nous en profiterons pour établir le théorème de Donsker 2.5.1 et pour détailler la méthodologie à suivre permettant d'établir un résultat de convergence en loi en dimension infinie. La section 2.6 explore brièvement les propriétés vérifiées par la mesure de Wiener et le mouvement Brownien; une construction alternative, à l'aide d'ondelettes, de celui-ci sera également donnée.

Pour plus de détails, nous renvoyons le lecteur vers [89, 126].

2.1 Différences avec la dimension finie

Tout d'abord, il est important de préciser que les définitions des différents modes de convergence introduits dans le chapitre précédent restent valables dans ce nouveau contexte; de manière similaire, la hiérarchie entre ces modes est conservée.

Passons maintenant en revue certaines caractérisations de la convergence en loi que nous avons évoquée au chapitre précédent. En dimension infinie, le théorème d'Helly-Bray 1.6.1 impliquant les fonctions de répartition n'est, bien entendu, plus satisfait. Dans un cadre aussi général, la transformée de Fourier d'une mesure de probabilité μ peut se définir par

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

où $\langle \cdot, \cdot \rangle$ désigne le crochet de dualité entre E et son dual E' . Malheureusement, il n'existe pas, en dimension infinie, un analogue du théorème de Lévy 1.5.1 permettant de caractériser simplement la convergence étroite. Ceci provient de l'absence, en dimension infinie, d'une inégalité similaire à celle de Lévy (cf. lemme 1.5.1). Nous allons donc nous cantonner à la définition classique de la convergence en loi via les fonctions continues et bornées :

$$\mu_n \Rightarrow \mu \iff \lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \quad \text{pour toute fonction } f \in C_b(E).$$

2.2 Convergence étroite de mesures de probabilités

Dans tout ce qui va suivre (E, d) désignera un espace polonais. Nous munirons E de sa tribu borélienne $\mathcal{B}(E)$ et si $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{B}(E))$ est une variable aléatoire, nous noterons sa loi par μ . L'ensemble des mesures de probabilités sur $(E, \mathcal{B}(E))$ sera noté $\mathcal{P}(E)$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

La propriété de tension présentée en dimension finie dans le premier chapitre (cf. définition 1.4.3) se généralise à des variables aléatoires à valeurs dans un espace polonais.

Définition 2.2.1. Une partie $\mathcal{P} \subset \mathcal{M}(E)$ composée de mesures de probabilités sur $(E, \mathcal{B}(E))$ est dite uniformément tendue si pour tout $\epsilon > 0$ il existe K_ϵ un compact de E tel que, pour tout $\mu \in \mathcal{P}$,

$$\mu(K_\epsilon^c) \leq \epsilon.$$

Remarque. Par simplicité, la plupart du temps nous parlerons de partie tendue pour désigner une partie uniformément tendue.

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

Proposition 2.2.2 (Ulam). Soit (E, d) un espace polonais et $\mu \in \mathcal{P}(E)$; alors μ est tendue.

Remarque. Ainsi toute partie finie de $\mathcal{P}(E)$ définit une partie tendue et comme dans le chapitre précédent, il est aussi possible de montrer le fait suivant : si $\mu_n \Rightarrow \mu$ alors $(\mu_n)_{n \in \mathbb{N}}$ est tendue.

Dans tous les cas, la difficulté sera de construire le compact K_ϵ . En effet, en dimension infinie, ce genre d'ensemble n'est plus caractérisé par les ensembles fermés et bornés. Nous allons donc profiter de la structure d'espace métrique pour laquelle les ensembles compacts sont caractérisés comme étant des ensembles complets et précompacts (cf. [13]).

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

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

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

$$\forall k \geq 1, \exists N_k \in \mathbb{N} \text{ tel que } \mu\left(\bigcup_{n \leq N_k} \overline{B(x_n, \delta)}\right) \geq 1 - \frac{\epsilon}{2^k}. \quad (2.2.1)$$

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

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

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

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

la dernière inégalité est obtenue grâce à (2.2.1). $\square$

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

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

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

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

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

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

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

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

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

d'autre part,

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

En résumé, nous avons montré que

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

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

2.3 Quelques critères pour établir de la convergence étroite

Voyons à présent s'il est possible de dégager des critères pratiques permettant d'obtenir un résultat de convergence étroite dans ce nouveau cadre (celui des espaces polonais).

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

Théorème 2.3.1 (Portemanteau). *Soient (E, d) un espace polonais, $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(E)$ et $\mu \in \mathcal{P}(E)$. Les assertions suivantes sont alors équivalentes :*

1. $\mu_n \Rightarrow \mu$ lorsque $n \rightarrow +\infty$.
2. Pour tout ouvert $O \subset E$, $\liminf_{n \rightarrow +\infty} \mu_n(O) \geq \mu(O)$.
3. Pour tout fermé $F \subset E$, $\limsup_{n \rightarrow +\infty} \mu_n(F) \leq \mu(F)$.
4. Pour tout $A \in \mathcal{B}(E)$ tel que $\mu(\partial A) = \mu(\bar{A} \setminus \mathring{A}) = 0$ alors

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

Remarque. Les boréliens de E vérifiant la condition $\mu(\partial A) = 0$ sont appelés ensembles de continuité pour la mesure μ . Le point 4 est donc l'analogue du théorème 1.6.1 de Helly-Bray en dimension infinie. D'ailleurs, lorsque $E = \mathbb{R}$, nous allons voir que le théorème de Helly-Bray peut se démontrer à partir du théorème du portemanteau.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Nous désignerons cette réunion dénombrable d'ensembles par I .

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

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

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

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

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

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

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

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

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

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

Comme annoncé plus tôt, il est possible de démontrer le théorème 1.6.1 d'Helly-Bray à partir du point (4) du théorème précédent. Rappelons d'ailleurs que la fonction de répartition d'une mesure de probabilité $\mu \in \mathcal{P}(\mathbb{R})$ est définie par $F_\mu(t) = \mu(]-\infty, t])$ pour tout $t \in \mathbb{R}$.

Corollaire 2.3.1 (Helly-Bray). *Si $E = \mathbb{R}$ les assertions suivantes sont équivalentes :*

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

Démonstration. Démontrons le sens direct. Si t est un point de continuité de F_μ alors $A =]-\infty, t]$ est un ensemble de continuité pour μ (i.e. $\mu(\partial A) = 0$). Si $\mu_n \Rightarrow \mu$ lorsque $n \rightarrow +\infty$ alors, d'après le théorème du portemanteau 2.3.1, $F_{\mu_n}(t) \rightarrow F_\mu(t)$ lorsque $n \rightarrow +\infty$.

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

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

afin de conclure grâce au théorème du portemanteau 2.3.1. Notons qu'un ouvert O de $\mathbb{R}$ peut s'écrire comme réunion dénombrable d'intervalles ouverts $O = \cup_{j \geq 1}]a_j, b_j[$ avec $a_j, b_j \in \mathbb{R}$ pour tout $j \geq 1$. Soit $\epsilon > 0$, par convergence monotone, il est possible de choisir $m \in \mathbb{N}$ suffisamment grand de sorte que

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

De plus, comme a_j et b_j sont des éléments de O pour tout $j \geq 1$, il est également possible de trouver $a'_j < a_j$ et $b'_j > b_j$ tels que $]a'_j, b'_j[\subset O$ tout en imposant que a'_j et b'_j ne soient pas des points de discontinuité de F_μ ou de F_{μ_n} et ceci pour tout $n \in \mathbb{N}$ (ceci est possible car, étant croissantes,

les applications $t \mapsto F_\mu(t)$ et $t \mapsto F_{\mu_n}(t)$ n'admettent qu'un nombre au plus dénombrable de points de discontinuité).

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

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

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

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

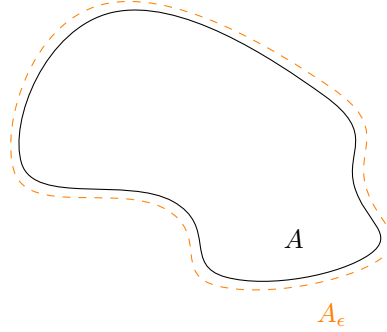
Enfin, ϵ étant arbitraire, la conclusion s'ensuit. $\square$

2.4 Distances et convergence étroite

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

Définition 2.4.1. Soit $A \subset E$, pour tout $\epsilon > 0$ l'ouvert A_ϵ désigne l'épaissi de A . Celui-ci est défini par

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



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

2.4.1 Distance de Lévy-Prokhorov

Définition 2.4.2. Soient $\mu, \nu \in \mathcal{P}(E)$, la distance de Lévy-Prokhorov entre μ et ν est définie par

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

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

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

A titre d'illustration, nous avons le résultat suivant.

Exemple 2.4.1. Soient $x, y \in E$, si $\mu = \delta_x$ et $\nu = \delta_y$ alors $\rho(\delta_x, \delta_y) = \min(1, d(x, y))$.

Proposition 2.4.1. ρ est une distance sur $\mathcal{P}(E)$.

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

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

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

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

En conséquence, nous venons de montrer que $\rho(\nu, \mu) \geq \epsilon$ donc, en faisant tendre ϵ vers $\rho(\mu, \nu)$ nous en déduisons que $\rho(\nu, \mu) \geq \rho(\mu, \nu)$. Il suffit ensuite d'échanger les rôles de μ et ν pour conclure que $\rho(\mu, \nu) = \rho(\nu, \mu)$.

Montrons à présent que ρ sépare les points (i.e. $\rho(\mu, \nu) = 0$ implique alors $\mu = \nu$). Si $\rho(\nu, \mu) = 0$ alors, pour tout fermé $F \subset E$ et pour $\epsilon = \frac{1}{k}$ avec $k \geq 1$, nous avons par définition de ρ

$$\mu(F) \leq \nu(F_{\frac{1}{k}}) + \frac{1}{k} \quad \text{pour tout } k \geq 1.$$

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

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

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

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

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

donc $\rho(\gamma, \mu) \leq \epsilon + \delta$. En conséquence, en faisant tendre δ vers $\rho(\mu, \nu)$ et ϵ vers $\rho(\gamma, \mu)$, nous obtenons $\rho(\gamma, \mu) \leq \rho(\gamma, \nu) + \rho(\mu, \nu)$. $\square$

2.4.2 Distance de Kantorovich-Rubinstein

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

Définition 2.4.3. Soit $f : E \rightarrow \mathbb{R}$ une fonction définie sur l'espace métrique (E, d) . Cette fonction est dite *lipschizienne* si le rapport suivant est fini :

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

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

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

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

Proposition 2.4.2 (Kantorovich-Rubinstein). Soient $\mu, \nu \in \mathcal{P}(E)$, la quantité suivante est une distance sur $\mathcal{P}(E)$.

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

β est appelée *distance de Kantorovich-Rubinstein*.

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

Nous allons constater que les distances de Kantorovich-Rubinstein et de Lévy-Prokhorov, introduites précédemment, permettent de quantifier la convergence étroite entre des mesures de probabilités.

Théorème 2.4.1. Soient (E, d) un espace polonais ainsi que $\mu_n, \mu \in \mathcal{P}(E)$. Les assertions suivantes sont alors équivalentes.

1. $\mu_n \xrightarrow{n \rightarrow +\infty} \mu$.
2. Pour toute fonction $f \in \text{BL}(E)$, nous avons $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$.
3. $\lim_{n \rightarrow +\infty} \beta(\mu_n, \mu) = 0$.
4. $\lim_{n \rightarrow +\infty} \rho(\mu_n, \mu) = 0$.

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

Théorème 2.4.2 (Arzéla-Ascoli). Soit (K, d) un espace métrique compact et considérons $C(K)$ muni de la distance uniforme $\|\cdot\|_\infty$. Les assertions suivantes sont alors équivalentes :

1. Une partie $A \subset C(K)$ est relativement compacte ;
2. A est bornée et équicontinue : pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que

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

Mentionnons enfin que d'autres distances qui peuvent être utilisées sur $\mathcal{P}(E)$: la distance en variation totale, la distance de Fortet-Mourier, la distance de Kantorovich-Wasserstein (que nous retrouverons en théorie du transport optimal),... mais nous n'en parlerons pas dans cet ouvrage.

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

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

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

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

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

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

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

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

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

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

2ième étape : Posons $\mathcal{F} = \{f \in \text{BL}(E); \|f\|_{\text{BL}(E)} \leq 1\}$ et observons que la famille des restrictions à l'ensemble K des fonctions $f \in \mathcal{F}$ forme, d'après le théorème d'Arzéla-Ascoli, une partie relativement compact de $(C(K), \|\cdot\|_\infty)$.

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

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

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

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

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

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

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

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

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

Ainsi,

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

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

Montrons à présent que (3) $\Rightarrow$ (4). Pour cela, nous allons montrer qu'il est possible de comparer les distances ρ et β . Soient $B \in \mathcal{B}(E)$ et $\mu, \nu \in \mathcal{P}(E)$, sans perdre en généralité, il est possible de supposer que $\mu(B) \leq \nu(B)$. Soit $\epsilon > 0$ et considérons à nouveau la fonction g définie par

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

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

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

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

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

Deux cas de figures se présentent alors à nous :

- Si $\beta(\mu, \nu) < 1$ et ϵ est choisi tel que $\beta(\mu, \nu) < \epsilon^2 < 1$, nous en déduisons que

$$(1 + \frac{1}{\epsilon})\beta(\mu, \nu) \leq \epsilon + \epsilon^2 < 2\epsilon.$$

Par suite, $\rho(\mu, \nu) \leq 2\epsilon$. En faisant tendre ϵ vers $\beta(\mu, \nu)$ nous obtenons alors l'inégalité suivante

$$\rho(\mu, \nu) \leq 2\sqrt{\beta(\mu, \nu)}.$$

- Maintenant, si $\beta(\mu, \nu) \geq 1$ nous en déduisons que $\rho(\mu, \nu) \leq 2\sqrt{\beta(\mu, \nu)}$ puisque $\rho(\mu, \nu) \leq 1$.

Dans tout les cas, nous pouvons conclure que (3) $\Rightarrow$ (4).

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

$$\mu(A_\delta \setminus A) < \epsilon \quad \text{et} \quad \mu((A^c)_\delta \setminus A^c) < \epsilon. \quad (2.4.1)$$

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

Pour n assez grand, puisque $\lim_{n \rightarrow +\infty} \rho(\mu_n, \mu) = 0$ et d'après les relations (2.4.1), nous avons

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

et, par passage au complémentaire,

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

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

Le théorème qui va suivre est essentiel car il permet de caractériser les parties relativement compactes de $\mathcal{P}(E)$. De manière assez remarquable, ces ensembles sont obtenus à partir des compacts de l'espace (E, d) et de la notion de tension (cf. définition 2.2.1).

Théorème 2.4.3 (Prokhorov). *Soit (E, d) un espace métrique polonais et $A \subset \mathcal{P}(E)$. Les assertions suivantes sont équivalentes :*

1. *A est une partie tendue.*
2. *De toute suite de A , il existe une sous-suite qui converge étroitement.*

3. l'ensemble A est relativement compact pour la distance β (ou ρ).
4. l'ensemble A est pré-compact pour la distance β (ou ρ).

Enonçons un corollaire immédiat avant de démontrer ce résultat.

Corollaire 2.4.1. *L'espace métrique $(\mathcal{P}(E), \beta)$ est un espace complet.*

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

Démonstration. En effet, si $(\mu_n)_{n \in \mathbb{N}}$ est une suite de Cauchy de (E, β) elle définit un ensemble pré-compact. D'après le théorème de Prokhorov 2.4.3, il est alors possible d'en extraire une sous-suite convergente. Puisqu'une suite de Cauchy n'admet qu'au plus une valeur d'adhérence, nous en déduisons qu'elle converge et l'espace est donc complet (pour la métrique induite par ρ ou β). $\square$

Passons à la démonstration du théorème de Prokhorov.

Démonstration. (du théorème de Prokhorov 2.4.3) Notons que (2) $\iff$ (3) car $\mathcal{P}(E)$ est un espace métrique. (3) $\Rightarrow$ (4) est trivial et (4) $\Rightarrow$ (1) est laissé à titre d'exercice (*indication* : il faut utiliser le fait qu'une partie finie est tendue ainsi que la relation $\rho \leq 2\sqrt{\beta}$ démontrée plus tôt).

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

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

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

En effet, pour cela il suffit de choisir une famille d'ensembles compacts $(L_j)_{j \in \mathbb{N}}$ tels que $\mu_n(L_j) \geq 1 - \frac{1}{j}$ grâce à la propriété de tension pour ensuite définir $K_j = \cup_{i \leq j} L_i$; la famille d'ensembles compacts $(K_j)_{j \in \mathbb{N}}$ vérifie alors la propriété désirée.

Ensuite, pour tout $j \geq 1$ et $n \in \mathbb{N}$, désignons par $\nu_n^j = \mu_n|_{K_j}$ la restriction de μ_n à l'ensemble K_j . Notons ensuite que l'espace $C(K_j, \|\cdot\|_\infty)$ un espace de Banach séparable (puisque K_j est un ensemble compact) et $C(K_j)' = \mathcal{M}(K_j)$ l'espace des mesures signées et bornées sur K_j (d'après le théorème 1.3.2 de dualité de Riesz).

En outre, puisque nous avons à disposition une suite bornée $(\nu_n^j)_{n \geq 1}$ dans $\mathcal{M}(K_j)$, le théorème 1.2.1 de Banach-Alaoglu nous assure qu'il est possible d'extraire une sous-suite convergente de $(\nu_n^j)_{n \in \mathbb{N}}$ (rappelons que j est fixé). Plus précisément, il existe ν^j telle que la sous-suite $(\nu_n^j)_{n \in \mathbb{N}}$ (pour alléger les notations, la sous-suite extraite sera notée comme la suite initiale) converge faiblement (et donc étroitement puisque K_j est un ensemble compact) vers ν^j .

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

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

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

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

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

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

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

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

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \left| \int_E f d\mu_n - \int_E f d\mu \right| &\leq \frac{1}{j} + \left| \int_{K_j} f d\nu^j - \int_E f d\mu \right| \\ &\leq \frac{1}{j} + |\nu^j(K_j) - \mu(E)|. \end{aligned}$$

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

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

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

2.5 Applications

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

Nous cherchons donc à démontrer un résultat de convergence en loi $(X_n \xrightarrow{\mathcal{L}} X$ lorsque $n \rightarrow +\infty)$ dans cet espace de dimension infinie. Il est tentant d'essayer d'utiliser à nouveau la transformée de Fourier. Celle-ci est définie par dualité

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

Malheureusement, comme mentionné plus tôt, l'outil de la fonction caractéristique ne suffit plus à décrire la convergence en loi (en dimension infinie, l'inégalité de Lévy (cf. lemme 1.5.1) fournissant de la tension à partir de la convergence des fonctions caractéristiques n'est plus disponible). Voici alors le schéma de démonstration qu'il faut suivre :

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

Remarque. 1. En pratique, l'étape délicate revient à établir cette propriété de tension.

2. Concernant l'identification de la limite, il est possible d'utiliser les lois finies dimensionnelles (via la transformée de Fourier). Plus précisément, nous étudierons pour tout d -uplet $(t_1, t_2, \dots, t_d)$ la convergence en loi dans $\mathbb{R}^d$ du vecteur aléatoire

$$(X_n(t_1), \dots, X_n(t_d))$$

afin d'identifier la limite X obtenue grâce à l'hypothèse de tension combinée au théorème de Prokhorov. Soulignons le fait, qu'en général, la convergence des lois marginales ne suffit pas à assurer la convergence en loi du processus $(X_n(t))_{t \in [0,1]}$, il est donc important d'avoir établi la propriété de tension au préalable.

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

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

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

Proposition 2.5.1. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variable aléatoire à valeurs dans $E = C^0[0, 1]$. La suite $(X_n)_{n \in \mathbb{N}}$ est alors tendue si et seulement si*

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

Remarque. Pour démontrer ce résultat nous allons devoir utiliser le théorème d'Arzélà-Ascoli 2.4.2. A ce propos, étant donnée une suite de fonctions continues $(x_n)_{n \in \mathbb{N}}$, observons que l'hypothèse : il existe $M > 0$ telle que

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

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

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

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

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

$$\sup_{n \in \mathbb{N}} \mathbb{P}(|X_n(t_0)| > M) \leq \eta.$$

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

$$\sup_{n \in \mathbb{N}} \mathbb{P} \left(\sup_{|s-t| \leq \delta_k} |X_n(s) - X_n(t)| > \frac{1}{k} \right) \leq \frac{\eta}{2^k}.$$

Définissons ensuite l'ensemble suivant :

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

et observons qu'il s'agit d'un ensemble fermé dans $(E, \|\cdot\|_\infty)$. De plus, d'après le théorème 2.4.2 d'Arzela-Ascoli, l'ensemble K est donc compact dans $(E, \|\cdot\|_\infty)$. Vérifions à présent que K vérifie la propriété de tension voulue. Pour tout $n \in \mathbb{N}$, nous avons, par construction,

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

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

$$\mathbb{P}(X_n \in K) = \mu_n(K) \geq 1 - 2\eta \quad \text{pour tout } n \in \mathbb{N}.$$

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

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

$n \rightarrow +\infty$

2.5.2 Principe d'invariance de Donsker

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

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

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

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

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

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

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

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

Identification de la limite

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

Proposition 2.5.2. *Soient $0 \leq t_1 < t_2 < \dots < t_d \leq 1$ alors*

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

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

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

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

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

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

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

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

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

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

Lemme 2.5.1. *Soient Y_n, X_n et X des variables aléatoires réelles. Supposons que $Y_n \xrightarrow{\mathcal{L}} X$ et $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$ lorsque $n \rightarrow +\infty$. Alors $X_n \xrightarrow{\mathcal{L}} X$ lorsque $n \rightarrow +\infty$.*

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

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

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

puisque $S_{[nt_1]}$ et $(S_{[nt_2]} - S_{[nt_1]})$ sont des variables aléatoires indépendantes (car $t_1 < t_2$). Nous pouvons alors utiliser le théorème 1.5.2 de la limite centrale dans $\mathbb{R}^2$ qui nous assure que

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

En d'autres termes,

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

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

Remarque. L'étape précédente nous montre la forme de la structure de covariance d'un éventuel processus limite X :

$$\mathbb{E}[X_s X_t] = \min(s, t) \quad \text{pour tout } s, t \in [0, 1].$$

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

Nous devons établir une propriété de tension pour la suite $\mu_n = \mathcal{L}(X^n)$ dans $E = C([0, 1])$. Pour cela nous allons nous reposer sur la proposition 2.5.1.

Nous devons donc montrer que les assertions suivantes sont satisfaites :

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

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

Soit $\delta > 0$, pour démontrer que la deuxième assertion est vérifiée, nous allons devoir étudier des quantités de la forme

$$\sup_{(s,t) \in [0,1]^2; |s-t| \leq \delta} |X_s^n - X_t^n| = \frac{1}{\sqrt{n}} \sup_{(u,v) \in [0,n]^2; |u-v| \leq n\delta} |Y_u - Y_v| \quad (2.5.1)$$

où, rappelons-le,

$$Y_u = S_{[u]} + (u - [u])V_{[u]+1} \quad X_t^n = \frac{1}{\sqrt{n}} Y_{nt} \quad \text{avec } t \in [0, 1].$$

Observons à présent que les majorations suivantes sont satisfaites :

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

$k=1, \dots, [n\delta]+1$

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

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

puisque $V_1 \in L^2$; la dernière inégalité est obtenue grâce à l'inégalité de Markov. En effet, si Z est une variable aléatoire positive alors l'inégalité de Markov nous assure que

$$\mathbb{P}(Z \geq t) \leq \frac{1}{t} \int_{\{Z \geq t\}} Z d\mathbb{P} \quad \text{pour tout } t > 0.$$

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

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

Soient ϵ et $\eta > 0$ et posons

$$T_k = \sum_{j=1}^k V_j 1_{\{|V_j| \leq \eta\sqrt{n}\}} \quad \text{pour } k = 1, \dots, \lfloor n\delta \rfloor + 1.$$

Notons que ces variables aléatoires ne sont pas centrées (contrairement aux variables aléatoires $(V_j)_{j=1, \dots, n}$). Mettons en place un argument de troncature.

$$\mathbb{P}\left(\max_{k=1, \dots, \lfloor n\delta \rfloor + 1} |S_k| \geq \epsilon\sqrt{n}\right) \leq \mathbb{P}\left(\max_{j=1, \dots, \lfloor n\delta \rfloor + 1} |V_j| \geq \eta\sqrt{n}\right) + \mathbb{P}\left(\max_{k=1, \dots, \lfloor n\delta \rfloor + 1} |T_k| \geq \epsilon\sqrt{n}\right) \quad (2.5.2)$$

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

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

puisque les variables aléatoires $(V_j)_{j=1, \dots, n}$ sont centrées. D'où, pour tout $k \in \{1, \dots, \lfloor n\delta \rfloor + 1\}$, nous avons

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

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

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

Ceci entraîne, par l'inégalité triangulaire, que

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

Posons ensuite $Z_k = T_k - \mathbb{E}[T_k]$ pour $k = 1, \dots, \lfloor n\delta \rfloor + 1$ et observons qu'il s'agit d'une somme de variables aléatoires centrées, de même loi. En particulier, il s'agit d'une martingale (pour la filtration induite par les variables aléatoires sous-jacentes). Il est donc possible d'utiliser l'inégalité maximale de Doob avec $p = 4$ (cf. [22]) : il existe donc une constante $C > 0$ telle que

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

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

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

En utilisant le fait que les variables $(\tilde{V}_j)_{j=1, \dots, \lfloor n\delta \rfloor + 1}$ sont indépendantes et centrées, nous en déduisons l'existence de constantes $C_1, C_2 > 0$ telles que

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

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

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

C'est pourquoi, pour tout $j \in \{1, \dots, \lfloor n\delta \rfloor + 1\}$, nous avons

$$\mathbb{E}[|\tilde{V}_j|^p] \leq C_p \mathbb{E}[|V_j|^p 1_{\{|V_j| \leq \eta \sqrt{n}\}}].$$

En particulier, pour $p = 4$ et $p = 2$, nous en déduisons que

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

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

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

Ceci étant vrai pour tout $\eta > 0$, nous avons donc

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

En résumé, tous les calculs qui précèdent permettent de justifier le résultat suivant.

Proposition 2.5.3. *Pour tout $\epsilon > 0$, $\lim_{\delta \rightarrow 0} \frac{1}{\delta} \left[\limsup_{n \rightarrow +\infty} \mathbb{P}\left(\max_{k=1,\dots,[n\delta]+1} |S_k| \geq \epsilon\sqrt{n}\right) \right] = 0$.*

Il est possible d'en déduire le résultat suivant (cf. [126]) dont la démonstration, reposant sur le même genre de technique, sera admise.

Corollaire 2.5.1. *Avec les notations précédentes, nous avons*

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

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

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

Théorème 2.5.1 (Donsker). *Soit $(X^n)_{n \geq 1}$ la suite d'interpolations linéaires introduite plus tôt. Cette suite converge en loi dans $E = C([0, 1])$ vers une variable aléatoire $B = (B_t)_{t \in [0, 1]}$ de E . De plus, l'objet limite est gaussien au sens suivant : pour tout $0 \leq t_1 < t_2 < \dots < t_d \leq 1$, le vecteur aléatoire $(B_{t_1}, \dots, B_{t_d})$ est un vecteur gaussien centré de $\mathbb{R}^d$ de matrice de covariance*

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

et $B_0 = 0$ p.s.

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

Démonstration. Si $\mu_n = \mathcal{L}(X^n)$ pour tout $n \in \mathbb{N}$. La combinaison du corollaire 2.5.1 et de la proposition 2.5.1 nous assure que la suite $(\mu_n)_{n \in \mathbb{N}}$ est tendue. Alors, d'après le théorème de Prokhorov, il est possible d'extraire une nouvelle sous-suite $(\mu_{n_k})_{k \in \mathbb{N}}$ qui converge étroitement vers une mesure de probabilité μ (correspondant à la loi d'un processus limite $(X_t)_{t \in [0, 1]}$).

A présent, les conditions induites par les lois finies dimensionnelles (cf. proposition 2.5.2) permettent d'identifier la loi limite (il s'agit là d'une propriété remarquable des processus gaussien cf. [89, 126]). Comme annoncé plus tôt, il s'agit de la loi d'un processus gaussien dont la structure de covariance est donnée par

$$\text{Cov}(X_{t_i}, \dots, B_{t_j}) = \min(t_i, t_j), \quad \text{pour tout } (i, j) \in \{1, \dots, d\} \quad \text{et tout } 0 \leq t_1 < t_2 < \dots < t_d \leq 1.$$

Par convention, ce processus sera noté $(B_t)_{t \in [0, 1]}$. □

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

L'étude complète du mouvement Brownien, notamment la construction de l'intégrale d'Ito fait l'objet d'un cours complet de calcul stochastique (cf. [180, 168, 126, 89]). Nous allons rapidement passer en revue quelques propriétés de ce nouvel objet, les démonstrations sont laissées en exercices.

Dans ce qui suit, $B = (B_t)_{t \in [0,1]} \in E = C^0[0,1]$ désignera un mouvement brownien et nous noterons par $\mu = \mathcal{L}(B)$ la mesure de Wiener.

Proposition 2.6.1. 1. Les trajectoires $t \mapsto B_t$ du mouvement brownien sont continues et $B(0) = 0$ p.s..
 2. Ce processus est à accroissements indépendants. Autrement dit, pour tout $0 \leq t_1 < \dots < t_d \leq 1$ les variables aléatoires

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

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

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

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

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

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

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

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

Bien sûr, la liste ne s'arrête pas là (cf. [180, 168, 126, 89]) et il est possible d'exposer de nombreuses autres propriétés.

Présentons rapidement une autre construction du mouvement brownien. Pour cela, considérons l'espace de Hilbert $L^2([0,1], d\lambda)$ où λ désigne la mesure de Lebesgue et considérons $(e_i)_{i \in \mathbb{N}}$ une base orthonormée de H . Il est possible de choisir, par exemple, la base définie à l'aide des ondelettes de Haar (cf. [180, 89]). Celles-ci s'obtiennent par translation et dilatation de la fonction suivante

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

2.6. QUELQUES PROPRIÉTÉS DU MOUVEMENT BROWNIEN ET DE LA MESURE DE WIENER 57

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

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

Pour de tels entiers, nous posons alors, pour tout $t \in [0, 1]$,

$$e_n(t) = 2^{j/2} H(2^j t - k) \quad \text{pour} \quad n = 2^j + k \quad \text{avec} \quad j \geq 0 \quad \text{et} \quad 0 \leq k < 2^j$$

et imposons $e_0(t) = 1$. La famille ainsi définie $(e_n)_{n \geq 0}$ est une base orthonormée de $L^2[0, 1]$.

A partir d'une telle famille, il est possible de construire une base de Schauder $(v_n)_n$ de $C[0, 1]$ (i.e. il s'agit d'une famille dense dans un espace de Banach). Pour cela, posons

$$v_n(t) = \int_0^t e_n(u) du \quad \text{pour tout} \quad t \in [0, 1]. \quad (2.6.1)$$

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

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

et $d_n(t) = d(2^j t - k)$ pour $n = 2^j + k$ avec $0 \leq k < 2^j$. Il est alors possible de montrer que

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

avec $\lambda_0 = 1$ et, pour $n \geq 1$, $\lambda_n = \frac{1}{2} 2^{-j/2}$ où $n = 2^j + k$ avec $0 \leq k < 2^j$ et $j \geq 0$.

A partir de cette famille, il est possible de construire le mouvement brownien. En effet, soit $(g_i)_{i \in \mathbb{N}}$ une suite de variables gaussiennes standards, indépendantes et identiquement distribuées (i.e. $g_1 \sim \mathcal{N}(0, 1)$) sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et posons

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

Cette suite de fonctions aléatoires vérifie la convergence suivante :

Théorème 2.6.1. *Presque sûrement, $(X^n)_{n \in \mathbb{N}}$ converge uniformément sur $[0, 1]$ vers une limite X qui est un mouvement brownien.*

Démonstration. Tout d'abord (en supposant que la limite existe) voyons ce qui se produit au niveau des fonctions de covariances et vérifions rapidement que

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_s^n X_t^n] = \mathbb{E}[X_s X_t] = \min(s, t).$$

Pour cela, nous utiliserons le fait que les variables $(g_i)_{i \in \mathbb{N}}$ sont des variables aléatoires gaussiennes standard i.i.d. .

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

Pour conclure, il suffit d'utiliser le théorème de Parseval (cf. [204]) qui nous assure que

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

Cette construction sera de nouveau employée lorsque nous démontrerons le théorème de Schilder dans le chapitre 4.

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

Lemme 2.6.1. *Dans le cadre décrit dans la section 2.6, il existe une variable aléatoire C telle que, presque sûrement, pour tout $n \geq 2$, nous avons*

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

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

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

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

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

De plus, puisque $\alpha > 1$, cette dernière quantité est sommable et le lemme de Borel-Cantelli 1.1.1 entraîne alors que

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

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

Fort de ce lemme nous pouvons établir la convergence uniforme de $X_n(t)$ vers X_t . Pour cela, nous allons montrer que, presque sûrement, le reste de la série (en valeur absolue) converge uniformément vers 0. Avant toutes choses, observons que pour tout $n \in [2^j, 2^{j+1}[$ nous avons :

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

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

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

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

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

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

est un vecteur gaussien. Ces détails sont laissés à l'attention du lecteur et peuvent être trouvés, par exemple, dans [180]. $\square$

2.7 Remarques historiques

Il est difficile d'obtenir un résumé de l'évolution de certains travaux mathématiques lorsque ceux-ci sont relativement récents. Nous empruntons donc ce qui va suivre aux ouvrages [89, 126, 168, 83].

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

« *Abréviation que, pour rire, j'ai suggérée à N. Bourbaki en 1949 après avoir appris le sujet dans la Topologie de Casimir Kuratowski [...] et constaté la contribution des Polonais ; j'aurais du*

reste aussi bien pu les appeler « polono-russes ». La plaisanterie, qui n'en était pas entièrement une, fut prise au sérieux et tous les experts ont depuis adopté cette étrange terminologie, généralement sans mentionner N. Bourbaki à l'exception de Kuratowski lui-même qui, dans un petit livre paru en 1974 sur l'histoire des mathématiques polonaises, y voit à juste titre un hommage à celles-ci. »

D'après [165], en 1905 Borel essaya de traduire le théorème des segments emboîtés en terme d'ensembles; cela donna naissance à la notion de fermés. Dans des ouvrages de philosophie mathématiques (traitant de la notion d'infini et de continuité), Russel introduit la notion de « compacité » pour désigner des suites de nombres rationnels vérifiant une certaine propriété. Les idées de Russel sont ensuite développées dans « Principles of mathematics » en 1903. L'introduction de ce nouveau terme se fit presque de manière simultanée avec la définition des ensembles compacts proposée par Fréchet. Cette proposition apparaît en 1906 dans la thèse de Fréchet et fait suite au mouvement (initié par Hilbert) cherchant trouver des espaces généraux permettant d'englober plusieurs théories mathématiques à la fois (afin d'en saisir plus profondément les points essentiels).

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

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

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

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

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

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

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

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

Deuxième partie

Concentration de la mesure

Chapitre 5

Introduction à la concentration de la mesure

Dans les chapitres précédents nous avons démontré différents résultats asymptotiques : le théorème d'invariance de Donsker 2.5.1, des principes de grandes déviations (les théorèmes de Cramér 3.2.1 et 3.3.1, celui de Gärtner-Ellis 3.4.1 ou encore le théorème de Schilder 4.3.2). Afin de mieux comprendre ceux-ci, nous allons voir qu'il est essentiel d'obtenir des inégalités, non-asymptotiques, qui quantifient ces résultats. Ces dernières sont souvent le reflet du phénomène de la concentration de la mesure. Dans ce chapitre nous allons décrire de manière formelle ce phénomène, nous présenterons ensuite des résultats historiques d'isopérimétrie (ceux associés aux problèmes euclidiens et gaussiens) et expliquerons de quelle manière ces derniers sont liés au phénomène de concentration de la mesure ; de manière équivalente, nous présenterons ensuite le phénomène de concentration par le prisme des fonctions lipschitziennes. Pour conclure ce chapitre, nous produirons deux exemples d'applications : le premier sera lié à la réduction de la dimension (cf. théorème de Johnson-Lindenstrauss 5.4.1) et le deuxième sera de nature géométrique (cf. théorème de Dvoretzky 5.5.1).

5.1 Cadre abstrait

Considérons (E, d, μ) un espace métrique mesuré. Supposons de plus que μ soit une mesure de probabilité sur les boréliens de E et rappelons que, pour tout $A \subset E$ et $r > 0$,

$$A_r = \{x \in E, d(x, A) < r\}$$

désigne le r -élargissement de A . Il est évident que $\mu(A_r) \rightarrow 1$ lorsque $r \rightarrow \infty$. Comme nous allons le voir, il est surprenant qu'il soit possible d'exprimer, sur un grand nombre d'exemples non-triviaux, la vitesse à laquelle ceci s'accomplit. Dans ce but, introduisons la définition suivante

Définition 5.1.1. *Dans le cadre précédent, nous définissons la fonction de concentration associée à (E, d, μ) par*

$$\alpha_{(E,d,\mu)}(r) = \sup \left\{ 1 - \mu(A_r) ; A \subset E, \mu(A) \geq \frac{1}{2} \right\} \quad \text{avec } r > 0.$$

Remarque. La valeur $\frac{1}{2}$ de la définition est arbitraire et peut-être remplacée par $\epsilon \in]0, 1[$ sans que cela n'altère véritablement les lignes qui vont suivre. Toutefois, comme nous le verrons par la suite, la valeur $\frac{1}{2}$ est pratique car elle permet de relier la fonction de concentration à des inégalités de déviations autour d'une médiane de fonctions lipschitziennes.

Il se trouve qu'il est souvent possible de contrôler cette fonction par un terme qui tend rapidement vers 0 lorsque $r \rightarrow \infty$. Par exemple, si jamais il existe des constantes $C, c > 0$ telles que

$$\alpha_{(E,d,\mu)}(r) \leq C e^{-cr^2} \text{ pour tout } r > 0, \quad (5.1.1)$$

nous parlerons de concentration gaussienne. Tandis que s'il existe des constantes $C, c > 0$ telles que

$$\alpha_{(E,d,\mu)}(r) \leq C e^{-cr} \text{ pour tout } r > 0$$

nous parlerons de concentration exponentielle. Il est à noter que d'autres comportements asymptotiques peuvent survenir (cf. [139, 7, 19]). Notons également que lorsque E est un espace vectoriel, la constante c peut dépendre de la dimension de ce dernier.

La terminologie de concentration de la mesure provient de l'observation suivante : si nous considérons le cas (5.1.1) avec un ensemble A de mesure $1/2$, alors quasiment tous les points (au sens de la mesure) de E seront à une distance $r = \frac{10}{\sqrt{c}}$ (par exemple) de l'ensemble A .

A titre d'illustration, considérons l'exemple suivant : soient $X_1, X_2, \dots$ une suite de variables aléatoires indépendantes, prenant les valeurs ± 1 avec la même probabilité et posons, pour $n \geq 1$, $S_n = X_1 + \dots + X_n$. Il est utile de considérer S_n comme une fonction des variables aléatoires X_i . L'inégalité classique suivante, exprimant une version non-asymptotique du théorème de la loi forte des grands nombres, satisfait par les variables X_i ,

$$\mathbb{P} \left(\frac{|S_n|}{n} \geq t \right) \leq 2e^{-nt^2/2} \text{ pour tout } t \geq 0 \quad (5.1.2)$$

nous assure que S_n/n est essentiellement nul (cf. [190]). Cette affirmation peut sembler choquante à première vue puisque les fluctuations de S_n sont d'ordre $\sqrt{n}$ qui est loin d'être nul. Il convient alors de prêter attention aux renormalisations employées (i.e. n ou $\sqrt{n}$) pour étudier S_n . Ce point de vue, proposé par Talagrand (cf. [190]), fut incroyablement fructueux et a permis de développer la théorie de la concentration de la mesure ; ses travaux novateurs mirent en évidence les nombreux champs d'application de ce nouveau domaine.

Dans ce qui suit, nous allons présenter l'approche historique de ce phénomène. Pour cela, nous allons poursuivre l'introduction (débutée dans la section 4.2.2) de la notion de problèmes isopérimétriques et montrerons de quelle manière les solutions de ces derniers permettent de contrôler la fonction de concentration $\alpha_{(E,d,\mu)}$.

5.2 Isopérimétrie

Nous rappelons au lecteur que la notion de problèmes isopérimétriques a déjà été introduite dans la section 4.2.2. Nous allons compléter cette introduction et, pour plus de détails, nous renvoyons le lecteur vers [32, 139, 37].

Nous présentons ci-dessous trois résultats isopérimétriques, celui de l'espace euclidien, celui de la sphère et enfin celui portant sur $\mathbb{R}^n$ muni de la mesure gaussienne standard. Nous ne présenterons pas le cas du cube discret $\{-1, 1\}^n$ équipé de la distance de Hamming et de la mesure uniforme (μ^n avec $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$) ; à ce sujet nous renvoyons le lecteur vers [139].

5.2.1 Problème isopérimétrique dans l'espace euclidien

Dans ce qui suit, Vol_n désignera la mesure de Lebesgue sur $(\mathbb{R}^n, \|\cdot\|_2)$. La solution du problème isopérimétrique euclidien s'exprime comme suit.

Théorème 5.2.1 (Isopérimétrie euclidienne). *Soient $A \subset \mathbb{R}^n$ et D une boule euclidienne telle que $\text{Vol}_n(A) = \text{Vol}_n(B)$, alors*

$$\text{Vol}_n(A_r) \geq \text{Vol}_n(B_r) \quad \text{pour tout } r \geq 0.$$

La démonstration de ce résultat peut s'effectuer via l'inégalité de Brunn-Minkowski dont voici l'énoncé.

Théorème 5.2.2 (Brunn-Minkowski). *Soient A et B deux ensembles mesurables bornés de $\mathbb{R}^n$. L'inégalité suivante est alors satisfaite*

$$\text{Vol}_n(A + B)^{1/n} \geq \text{Vol}_n(A)^{1/n} + \text{Vol}_n(B)^{1/n} \quad (5.2.1)$$

où $A + B = \{x + y; x \in A \text{ et } y \in B\}$ désigne la somme de Minkowski des ensembles A et B .

Remarque. Une preuve de ce résultat, à l'aide de la théorie du transport optimal, sera donnée dans le chapitre 10.

Armé de ce résultat, il n'est pas difficile de démontrer le théorème 5.2.1.

Démonstration. (Isopérimétrie euclidienne 5.2.1) Sous les hypothèses du théorème, considérons B la boule euclidienne de centre 0 et de rayon $r > 0$. D'après l'inégalité de Brunn-Minkowski (5.2.1), nous avons

$$\text{Vol}_n(A_r)^{1/n} = \text{Vol}_n(A + B)^{1/n} \geq \text{Vol}_n(A)^{1/n} + v(r)^{1/n}$$

où $v(r)$ désigne le volume d'une boule euclidienne de rayon $r > 0$. Par hypothèse, $\text{Vol}_n(A) = \text{Vol}_n(D)$ avec D une boule de rayon $s > 0$. Alors, puisque $r \mapsto v^{1/n}(r)$ est linéaire, nous avons également

$$\text{Vol}_n(A)^{1/n} + v(r)^{1/n} = \text{Vol}_n(D)^{1/n} + v(r)^{1/n} = v(r + s)^{1/n} = \text{Vol}_n(D_r)^{1/n}.$$

D'où le résultat escompté,

$$\text{Vol}_n(A_r) \geq \text{Vol}_n(D_r)^{1/n}.$$

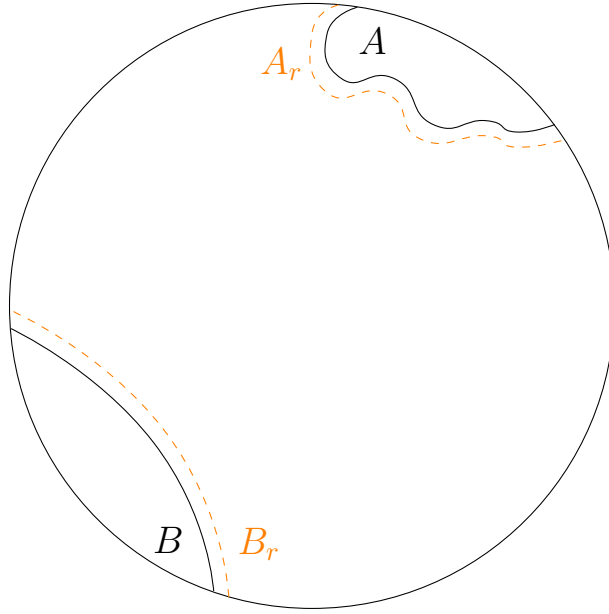
□

5.2.2 Isopérimétrie sur la sphère et l'espace gaussien

Maintenant que la terminologie associée aux problèmes isopérimétriques est introduite, voyons deux autres résultats essentiels. Considérons tout d'abord, pour $n \geq 1$, celui sur la sphère $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ équipée de sa mesure de volume (renormalisée) riemannienne σ_n .

Théorème 5.2.3 (Schmidt). *Soient $A \subset \mathbb{S}^n$ et C une calotte sphérique satisfaisant $\sigma_n(C) = \sigma_n(A)$, alors*

$$\sigma_n(A_r) \geq \sigma_n(C_r) \quad \text{pour tout } r \geq 0.$$



Ce résultat propose un premier exemple non trivial du phénomène de concentration de la mesure en grande dimension. En effet, ce théorème implique que l'élargissement d'ordre $1/\sqrt{n}$ d'un ensemble de demi-mesure recouvre presque toute la sphère. En effet (cf. [139]), si $\sigma_n(A) = 1/2$,

$$\sigma_n(A_r) \geq \sigma_n(C_r) \geq 1 - 2e^{-nr^2/2}.$$

Ce résultat permet notamment de résoudre le problème isopérimétrique dans l'espace gaussien $(\mathbb{R}^n, \|\cdot\|_2, \gamma_n)$. En effet, une observation attribuée, par erreur semble-t-il, à Poincaré stipule que l'espace gaussien peut-être « vu » comme limite de sphère N dimensionnelle de rayon $\sqrt{N}$ (cf. [139]). Dans le problème isopérimétrique gaussien, les ensembles extrémaux ne sont plus des calottes sphériques ou des boules euclidiennes mais des demi-espaces ; nous rappelons que de tels ensembles sont définis par

$$H = \{x \in \mathbb{R}^n ; \langle x, u \rangle \leq a\}$$

avec $u \in \mathbb{R}^n$ un vecteur unitaire et $a \in \mathbb{R}$. Pour le confort du lecteur, nous reproduisons ci-dessous l'énoncé de la solution du problème isopérimétrique gaussien.

Théorème 5.2.4 (Borell, Sudakov, Tsirel'son, Ibragimov). *Soit $A \subset \mathbb{R}^n$ et H un demi-espace tel que $\gamma_n(A) = \gamma_n(H)$. Alors*

$$\gamma_n(A_r) \geq \gamma_n(H_r) \quad \text{pour tout } r \geq 0.$$

Remarque. Une propriété remarquable et inhérente aux mesures gaussiennes est que le résultat précédent ne dépend pas de la dimension, contrairement à celui des sphères euclidiennes. Comme nous l'avons déjà observé dans le chapitre 4, le champ d'application de telles inégalités est plutôt large.

Voyons de quelle manière ces inégalités sur les ensembles peuvent se traduire en terme de fonctions.

5.3 Inégalités de déviations

Ci-dessous, nous expliquons la pertinence du choix de la valeur $\frac{1}{2}$ dans la définition de la fonction de concentration $\alpha_{(E,d,\mu)}$. Pour alléger les notations, nous remplacerons $\alpha_{(E,d,\mu)}$ par α_μ , le contexte ne laissant aucune ambiguïté sur l'espace (E, d) sous-jacent ; étant donné une fonction f , m_f désignera une de ses médianes (sous μ) : ‘

$$\text{i.e. } \mu(f \geq m_f) \geq \frac{1}{2} \quad \text{et} \quad \mu(f \leq m_f) \geq \frac{1}{2}.$$

Proposition 5.3.1. *Soit μ une mesure de probabilité sur l'espace métrique (E, d) . Si $f : E \rightarrow \mathbb{R}$ est une fonction lipschitzienne alors*

$$\mu(f \geq m_f + r) \leq \alpha_\mu(r/\|f\|_{\text{Lip}}) \quad \text{pour tout } r > 0. \quad (5.3.1)$$

et

$$\mu(|f - m_f| \geq r) \leq 2\alpha_\mu(r/\|f\|_{\text{Lip}}) \quad \text{pour tout } r > 0. \quad (5.3.2)$$

Réciproquement, s'il existe une fonction positive $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ telle que, pour toute fonction f 1-lipschitzienne de médiane m_f nous ayons

$$\mu(f \geq m_f + r) \leq \alpha(r) \quad \text{pour tout } r > 0,$$

alors $\alpha_\mu \leq \alpha$.

Remarque. L'inégalité (5.3.1) est désignée sous l'appellation d'inégalité de déviation (par rapport à la médiane), tandis que l'inégalité (5.3.2) est appelée inégalité de concentration. Notons également, qu'en général, les paramètres m_f et $\|f\|_{\text{Lip}}$ peuvent-être d'ordre très différents et les inégalités de concentration ne donnent aucune information quant aux ordres de grandeur de ces derniers.

Démonstration. Soit f une fonction lipschitzienne et considérons l'ensemble $A = \{f \leq m\}$ avec $m \in \mathbb{R}$, alors

$$A_r \subset \{f < m + r\|f\|_{\text{Lip}}\} \quad \text{pour tout } r > 0.$$

En particulier, si $m = m_f$, nous avons (par définition de la fonction α_μ)

$$\mu(f \geq m_f + r) \leq \alpha_\mu(r/\|f\|_{\text{Lip}}) \quad \text{pour tout } r > 0.$$

Réciproquement, si A est un borélien de (E, d) tel que $\mu(A) \geq \frac{1}{2}$, posons $f(x) = d(x, A)$ pour $x \in E$. Clairement, $\|f\|_{\text{Lip}} \leq 1$ tandis que

$$\mu(f = 0) \geq \mu(A) \geq \frac{1}{2}.$$

Ainsi, puisque $f \leq 0$, cela signifie que 0 est une médiane de f . Alors, par hypothèse, nous avons

$$1 - \mu(A_r) = \mu(f \geq r) \leq \alpha_\mu(r)$$

d'où le résultat. $\square$

Au vu de la proposition précédente, il est alors possible d'exprimer de manière fonctionnelle les résultats isopérimétriques présentés plus tôt.

Théorème 5.3.1 (Lévy). *Soit $f : \mathbb{S}^n \rightarrow \mathbb{R}$ une fonction 1-lipschitzienne et $m_f = \text{Med}(f)$ sa médiane. Alors,*

$$\sigma_n(|f - m_f| \geq t) \leq 2e^{-\frac{nt^2}{2}} \quad \text{pour tout } t \geq 0.$$

Dans le cas gaussien, voici ce que nous obtenons.

Théorème 5.3.2. *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction 1-lipschitzienne et $m_f = \text{Med}(f)$ sa médiane, alors*

$$\gamma_n(|f - m_f| \geq t) \leq 2e^{-\frac{t^2}{2}} \quad \text{pour tout } t \geq 0. \quad (5.3.3)$$

Remarque. Faisons deux remarques quant à ce nouveau résultat.

1. L'inégalité (5.3.3) est à comparer avec celle obtenue dans le corollaire 4.2.1. Dans le chapitre 4 le résultat avait alors été obtenu par des méthodes d'interpolation et de monotonie le long du flot de la chaleur (cf. [135]) et l'inégalité obtenue impliquait $\mathbb{E}_{\gamma_n}(f)$ au lieu de m_f . Comme cela est expliqué dans [139] (page 9-10), il est possible de passer d'une formulation à l'autre au prix de constantes numériques. Soulignons toutefois, qu'en général, il est plus simple de mettre en place une méthode d'interpolation que de résoudre un problème isopérimétrique.
2. Il est à noter que l'inégalité de concentration (5.3.3) (ou celle du corollaire 4.2.1) permet également de contrôler la taille de la variance de la fonction f . Pour cela, il suffit d'observer que

$$\text{Var}_{\gamma_n}(f) = 2 \int_0^\infty r \mathbb{P}(|f - m_f| \geq r) dr.$$

En particulier, si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction lipschitzienne nous avons

$$\text{Var}_{\gamma_n}(f) \leq 2 \int_0^\infty r e^{-\frac{r^2}{2\|f\|_{\text{Lip}}^2}} dr = 2\|f\|_{\text{Lip}}^2.$$

Présentons à présent deux applications du phénomène de concentration de la mesure.

5.4 Réduction de la dimension

Le résultat suivant montre qu'il est possible de plonger une famille de N vecteurs d'un espace vectoriel dans un espace vectoriel de dimension plus petite, au prix d'une perte d'information quantifiable.

Proposition 5.4.1 (Johnson-Lindenstrauss). *Soient $p_1, \dots, p_N \in \mathbb{R}^n$, $\epsilon > 0$ et $k \in \mathbb{N}$ tels que $k > \frac{4}{\epsilon^2} \log N$ alors il est possible de trouver une application linéaire $\pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ telle que*

$$(1 - \epsilon) \|p_i - p_j\|_{\mathbb{R}^n} \leq \|\pi(p_i) - \pi(p_j)\|_{\mathbb{R}^k} \leq (1 + \epsilon) \|p_i - p_j\|_{\mathbb{R}^n}.$$

Remarque. L'application linéaire π est une quasi-isométrie que nous allons construire de manière aléatoire; elle permet de réduire la dimension.

Pour qu'il n'y ait aucune confusion dans la démonstration, nous noterons par $\|\cdot\|_{\mathbb{R}^d}$ la norme euclidienne de $\mathbb{R}^d$.

Démonstration. Soient $X = (X_{ij})_{(i,j) \in \{1, \dots, k\} \times \{1, \dots, n\}}$ une famille i.i.d. de variables aléatoires gaussiennes standards et $u \in \mathbb{R}^n$ un vecteur fixé. Il est possible d'assimiler la famille X à une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^k$; i. e. cela revient à identifier X avec une matrice de taille $k \times n$. Considérons alors l'application suivante

$$f_u(X) = \|Xu\|_{\mathbb{R}^k}.$$

Autrement dit, à $u \in \mathbb{R}^n$ fixé, $f_u : \mathbb{R}^{k \times n} \rightarrow \mathbb{R}_+$. Il n'est pas difficile de montrer qu'il s'agit d'une application L -lipschitzienne avec $L \leq \|u\|_{\mathbb{R}^n}$, c'est pourquoi, d'après (4.2.5), nous avons

$$\mathbb{P}\left(|f_u - \mathbb{E}[f_u]| \geq r\|u\|_{\mathbb{R}^n}\right) \leq 2e^{-\frac{r^2}{2}} \quad \text{pour tout } r > 0.$$

De plus, après quelques efforts techniques (cf. [139, 44]), il est possible de montrer que $\mathbb{E}[f_u] \simeq \sqrt{k}\|u\|_{\mathbb{R}^n}$.¹ Choisissons alors $r = \sqrt{k}\epsilon$, ceci fournit

$$\mathbb{P}\left(\frac{\|Xu\|_{\mathbb{R}^k}}{\sqrt{k}} \notin [(1 - \epsilon)\|u\|_{\mathbb{R}^n}; (1 + \epsilon)\|u\|_{\mathbb{R}^n}]\right) \leq 2e^{-\frac{k\epsilon^2}{2}}$$

pour tout vecteur $u \in \mathbb{R}^n$. En particulier, si $l = \binom{N}{2}$ et $u_1, \dots, u_l$ est une énumération des différences $p_i - p_j$ avec $(i, j) \in \{1, \dots, k\} \times \{1, \dots, n\}$, $i \neq j$ nous pouvons appliquer ce qui précède afin d'obtenir

$$\mathbb{P}\left(\exists i \in \{1, \dots, l\}; \frac{\|Xu_i\|_{\mathbb{R}^k}}{\sqrt{k}} \notin [(1 - \epsilon)\|u_i\|_{\mathbb{R}^n}; (1 + \epsilon)\|u_i\|_{\mathbb{R}^n}]\right) \leq 2le^{-\frac{k\epsilon^2}{2}}.$$

Or, par hypothèse $k > \frac{4}{\epsilon^2} \log N > c4\epsilon \log((2l)^{1/2}) = 2\epsilon \log 2l$, c'est pourquoi

$$\mathbb{P}\left(\exists i \in \{1, \dots, l\}; \frac{\|Xu_i\|_{\mathbb{R}^k}}{\sqrt{k}} \notin [(1 - \epsilon)\|u_i\|_{\mathbb{R}^n}; (1 + \epsilon)\|u_i\|_{\mathbb{R}^n}]\right) < 1.$$

Autrement dit, en passant au complémentaire, nous avons

1. Cette égalité est vraie modulo des constantes numériques multiplicatives indépendantes de la dimension. Pour alléger les notations dans la démonstration, nous ferons comme s'il s'agissait d'une véritable égalité.

$$\mathbb{P}\left(\forall i \in \{1, \dots, l\}; \frac{\|Xu_i\|_{\mathbb{R}^k}}{\sqrt{k}} \in [(1-\epsilon)\|u_i\|_{\mathbb{R}^n}; (1+\epsilon)\|u_i\|_{\mathbb{R}^n}]\right) > 0,$$

il existe donc $\omega \in \Omega$ tel que l'application π définie par $\pi(u) = \frac{X(\omega)u}{\sqrt{k}}$ vérifie la propriété voulue. $\square$

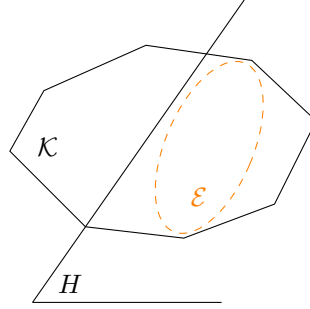
Passons à présent à une application géométrique de la concentration de la mesure.

5.5 Théorème de Dvoretzky

Voici l'énoncé du problème auquel le théorème de Dvoretzky 5.5.1 apporte une solution.

Dans ce qui suit $\mathcal{K}$ est un ensemble compact, convexe, d'intérieur non vide et symétrique par rapport à l'origine. Nous dirons que $\mathcal{K}$ contient des sections euclidiennes de dimension k lorsque pour tout $\epsilon > 0$, il existe un sous-espace H de dimension k ainsi qu'un ellipsoïde $\mathcal{E} \subset H$ tels que

$$(1-\epsilon)\mathcal{E} \subset \mathcal{K} \cap H \subset (1+\epsilon)\mathcal{E}.$$



Une manière équivalente de formuler ceci est la suivante. Considérons $(E, \|\cdot\|)$ un espace de Banach, dire que $\mathcal{K} \subset E$ contient des sections euclidiennes de dimension k est équivalent à dire que E est $(1+\epsilon)$ -isomorphe à l'espace euclidien $\mathbb{R}^k$. Ceci signifie qu'il existe une famille $v_1, \dots, v_k$ de k vecteurs de E tels que pour tout $t = (t_1, \dots, t_k) \in \mathbb{R}^k$ nous ayons

$$(1-\epsilon)\|t\|_2 \leq \left\| \sum_{i=1}^k t_i v_i \right\| \leq (1+\epsilon)\|t\|_2$$

où $\|\cdot\|_2$ désigne la norme euclidienne sur $\mathbb{R}^k$. Autrement dit, tout espace de Banach de dimension finie n admet un sous-espace vectoriel « presque » euclidien à condition d'imposer une restriction sur la dimension de celui-ci. Le théorème de Dvoretzky 5.5.1 quantifie la dimension admissible de $\mathbb{R}^k$.

Théorème 5.5.1 (Dvoretzky). *Pour tout $\epsilon > 0$, il existe $\eta(\epsilon)$ tel que tout espace de Banach E de dimension n admet un sous-espace isomorphe à $\mathbb{R}^k$ avec $k = \lfloor \eta(\epsilon) \log n \rfloor$.*

La démonstration de ceci s'effectue en trois étapes. La première consiste à construire une application lipschitzienne vérifiant un certain nombre de propriétés. La deuxième, plus technique, permet de se ramener à un nombre fini de vecteurs en utilisant des δ -réseaux et la dernière, la plus

importante, utilise les propriétés de concentration de la mesure gaussienne ainsi que sa propriété d'invariance par rotation.

Débutons par la construction de l'application lipschitzienne adéquate.

Lemme 5.5.1 (Dvoretzky-Rodgers). *Soient $(E, \|\cdot\|)$ un espace de Banach de dimension n et $m = \lfloor n/2 \rfloor$. Il existe des vecteurs $w_1, \dots, w_m$ de E tels que $\|w_j\| \leq \frac{1}{2}$ pour tout $j = 1, \dots, m$ et*

$$\left\| \sum_{j=1}^m t_j w_j \right\| \leq \|t\|_2$$

pour tout $t \in \mathbb{R}^m$.

Démonstration. • (1ère étape) :

Construisons tout d'abord une application $T : \mathbb{R}^n \rightarrow E$ telle que

$$\|T\| \leq 1 \quad \text{et} \quad \|T|_V\| \geq \dim(V)/n$$

pour tout sous-espace vectoriel V de $\mathbb{R}^n$. En particulier, $\|T\| = 1$. A cet effet, soit T tel que

$$\det(T) = \max\{\det(S) ; S : \mathbb{R}^n \rightarrow E, \|S\| \leq 1\}.$$

Alors, pour tout $\epsilon > 0$ et tout $S : \mathbb{R}^n \rightarrow E$, le caractère maximal de T implique que

$$\det\left(\frac{T + \epsilon S}{\|T + \epsilon S\|}\right) \leq \det(T) \iff \det(T + \epsilon S) \leq \det(T) \|T + \epsilon S\|^n. \quad (5.5.1)$$

En outre, en utilisant un développement limité (du déterminant) lorsque ϵ tend vers 0, nous avons

$$\begin{aligned} \det(T + \epsilon S) &= \det(T) \det(I_d + \epsilon T^{-1} S) \\ &= \det(T) (1 + \epsilon \operatorname{Tr}(T^{-1} S) + o(\epsilon)) \end{aligned}$$

D'où, en substituant cette identité dans (5.5.1),

$$\begin{aligned} 1 + \epsilon \operatorname{Tr}(T^{-1} S) &\leq \|T + \epsilon S\|^n + o(\epsilon) \\ &\leq (1 + \epsilon \|S\|)^n + o(\epsilon) \\ &\leq 1 + \epsilon n \|S\| + o(\epsilon). \end{aligned}$$

Ainsi, $\operatorname{Tr}(T^{-1} S) \leq n \|S\|$. Soient V un sous-espace vectoriel de $\mathbb{R}^n$ et $P : \mathbb{R}^n \rightarrow V$ la projection orthogonale sur celui-ci et posons $S = TP$. D'après ce qui précède,

$$\|T|_V\| = \|TP\| \quad \text{et} \quad \dim(V) = \operatorname{Tr}(P) = \operatorname{Tr}(T^{-1} S) \leq n \|S\|$$

et donc $\|TP\| \geq \dim(V)/n$. Il n'est plus difficile d'achever la preuve de ce lemme.

• (2ième étape) :

Considérons T l'application que nous venons de construire. Soit $y_1 \in \mathbb{R}^n$ tel que $\|y_1\|_2 = 1$ et $\|Ty_1\| = \|T\| = 1$. Ensuite, choisissons $y_2 \in \text{vect}(y_1)^\perp$ tel que $\|Ty_2\| = \|TQ\| \geq 1 - \frac{1}{n}$ où Q désigne la projection orthogonale sur $\text{vect}(y_1)^\perp$. En poursuivant ce procédé, nous obtenons une famille de vecteurs tels que $y_j \in \{y_1, \dots, y_{j-1}\}^\perp$ et $\|y_j\|_2 = 1$ et $\|Ty_j\| \geq \frac{n-j+1}{n}$.

• (Conclusion) :

Il suffit ensuite de poser $w_j = Ty_j$, pour $j = 1, \dots, m$ pour déterminer la démonstration. $\square$

Remarque. Grâce au lemme 5.5.1 de Dvoretzky-Rodgers, nous pouvons construire l'application lipschitzienne F qui nous permettra d'utiliser l'inégalité de concentration (5.3.3). Pour cela, définissons $F : \mathbb{R}^m \rightarrow \mathbb{R}$ par

$$F(x) = \left\| \sum_{j=1}^m x_j w_j \right\| \quad \text{pour tout } x = (x_1, \dots, x_m) \in \mathbb{R}^m. \quad (5.5.2)$$

D'après le lemme 5.5.1, $\|F\|_{\text{Lip}} \leq 1$. En effet, pour tout $x, y \in \mathbb{R}^m$,

$$\|F(x) - F(y)\|_2 \leq \left\| \sum_{j=1}^m (x_j - y_j) w_j \right\| \leq \|x - y\|_2.$$

Poursuivons notre présentation des résultats préliminaires qui nous permettront de démontrer le théorème de Dvoretzky 5.5.1. Durant la démonstration de ce théorème, nous aurons besoin d'utiliser un réseau de la sphère unité pour se ramener à un nombre fini de vecteurs. Rappelons à cet effet la définition suivante

Définition 5.5.1. Un δ -réseau d'un ensemble $A \subset \mathbb{R}^k$ désigne un sous ensemble $\mathcal{S} \subset A$ tel que pour tout $t \in A$, il existe $s \in \mathcal{S}$ tel que $\|t - s\|_2 \leq \delta$.

Remarque. Comme cela nous sera utile ultérieurement, notons qu'il existe (cf. [139]) un δ -réseau de la sphère unité de $\mathbb{R}^k$ de cardinalité inférieure à $(1 + \frac{2}{\delta})^k \leq e^{2k/\delta}$.

Voici enfin le dernier résultat technique nécessaire pour démontrer le théorème de Dvoretzky 5.5.1.

Lemme 5.5.2. Pour tout $\epsilon > 0$, il existe $\delta = \delta(\epsilon)$ satisfaisant la propriété suivante. Soit $\mathcal{S}$ un δ -réseau de la sphère unité de $\mathbb{R}^k$; s'il existe $v_1, \dots, v_k$ des vecteurs d'un espace de Banach E tels que $1 - \delta \leq \|\sum_{i=1}^k t_i v_i\| \leq 1 + \delta$ pour tout $t \in \mathcal{S}$ alors

$$(1 - \epsilon)\|t\|_2 \leq \left\| \sum_{i=1}^k t_i v_i \right\| \leq (1 + \epsilon)\|t\|_2$$

pour tout $t \in \mathbb{R}^k$.

Démonstration. Par homogénéité, il suffit de démontrer le résultat pour tout $t \in \mathbb{R}^k$ tel que $\|t\|_2 = 1$. Par définition de $\mathcal{S}$, il existe $t^0 \in \mathcal{S}$ tel que $\|t - t^0\|_2 \leq \delta$. Par conséquent, $t = t^0 + \lambda_1 s$ avec s un vecteur unitaire de $\mathbb{R}^k$ et $|\lambda_1| \leq \delta$. En répétant cette procédure k fois, nous obtenons le fait suivant :

il est possible de décomposer t en $t = \sum_{l \geq 0} \lambda_l t^l$ avec $t^l \in \mathcal{S}$, $|\lambda_l| \leq \delta^l$ pour tout $l \geq 1$ et $\lambda_0 = 1$. Ceci entraîne alors

$$\left\| \sum_{i=1}^k t_i v_i \right\| = \left\| \sum_{l=0}^{\infty} \lambda_l \sum_{i=1}^k t_i^l v_i \right\| \leq \sum_{l=0}^{\infty} \delta^l \left\| \sum_{i=1}^k t_i^l v_i \right\| \leq \frac{1+\delta}{1-\delta} \quad \text{avec } 0 < \delta < 1.$$

De la même façon, en utilisant l'inégalité triangulaire inverse, nous avons

$$\begin{aligned} \left\| \sum_{i=1}^k t_i v_i \right\| &= \left\| \sum_{l=1}^n t_i^0 v_i - \left(- \sum_{l=1}^n \lambda_l \sum_{i=1}^n t_i^l v_i \right) \right\| \\ &\geq \left\| \sum_{l=1}^n t_i^0 v_i \right\| - \left\| \sum_{l=1}^{\infty} \lambda_l \sum_{i=1}^n t_i^l v_i \right\| \\ &\geq 1 - \delta - \frac{\delta(1+\delta)}{1-\delta} = \frac{1-3\delta}{1-\delta}. \end{aligned}$$

Il suffit ensuite de choisir $\delta = \delta(\epsilon)$ suffisamment petit pour conclure. $\square$

Nous pouvons à présent établir le théorème de Dvoretzky 5.5.1.

Démonstration. L'idée principale est de combiner l'inégalité de concentration gaussienne (5.3.3) pour les fonctions lipschitziennes avec la propriété d'invariance par rotation de la mesure gaussienne standard : autrement dit, pour tout vecteur unitaire $t \in \mathbb{R}^k$ nous avons l'égalité en loi

$$\langle t, X \rangle = X$$

avec $X = (X_1, \dots, X_k) \sim \mathcal{N}_k(0, I_d)$.

Considérons l'application F définie par (5.5.2) et γ_m (pour $m = \lfloor n/2 \rfloor$) la mesure gaussienne standard sur $\mathbb{R}^m$. Soit $k \geq 1$ que nous spécifierons plus tard et considérons $t = (t_1, \dots, t_k)$ un vecteur de la sphère unité de $\mathbb{R}^k$. Grâce à l'invariance par rotation de γ_m , la loi de l'application

$$(X_1, \dots, X_k) \in (\mathbb{R}^m)^k \mapsto \sum_{i=1}^k t_i X_i \in \mathbb{R}^m$$

sous la mesure produit γ_m^k est la même que γ_m . C'est pourquoi, d'après (5.3.3), pour tout $r \geq 0$,

$$\gamma_m^k \left((x_1, \dots, x_k) \in (\mathbb{R}^m)^k ; \left| F \left(\sum_{i=1}^k t_i x_i \right) - M_F \right| \geq r \right) \leq 2e^{-r^2/2}$$

où $M_F = \int F d\gamma_m$. A présent, soit $\epsilon > 0$ fixé et considérons $\delta = \delta(\epsilon)$ fourni par le lemme 5.5.2. De plus, considérons $\mathcal{S}$ un δ -réseau de la sphère unité de $\mathbb{R}^k$; d'après la remarque 5.5, celui-ci vérifie $\text{Card}(\mathcal{S}) \leq e^{2k/\delta}$. Enfin, choisissons $r = \delta M_F$ dans l'inégalité précédente. Tout ceci entraîne alors que

$$\begin{aligned} \gamma_m^k \left(\exists t \in \mathcal{S}; \left\| \sum_{i=1}^k t_i \frac{Z_i}{M_F} \right\| - 1 \geq \delta \right) &\leq 2\text{Card}(\mathcal{S})e^{-\delta^2 M_F^2/2} \\ &\leq 2e^{(2k/\delta) - (\delta^2 M_F^2/2)} \end{aligned}$$

où $Z_i = \sum_{j=1}^m X_{ij} w_j \in E$ et $X_i = (X_{ij})_{1 \leq j \leq m}$ pour tout $i = 1, \dots, k$.

Il ne reste plus qu'à minorer M_F . A cet effet, des arguments standards permettent de montrer qu'il existe $\rho > 0$ telle que

$$M_F \geq \int_{\mathbb{R}^m} \max_{j=1, \dots, m} x_j d\gamma_m(x) \geq \rho \sqrt{\log m}.$$

D'où, pour tout $n \geq 3$, nous avons $M_F \geq \rho \sqrt{\log n}$. Il est alors possible de choisir $k = \lfloor \eta \log n \rfloor$ avec $\eta = \eta(\delta) = \eta(\epsilon)$ suffisamment petit de sorte que $2e^{(2k/\delta) - (\delta^2 M_F^2/2)} < 1$. C'est pourquoi il existe au moins une réalisation de $(X_1, \dots, X_k)$ tel que $(v_1, \dots, v_k) = (M_F)^{-1}(Z_1, \dots, Z_k)$ satisfait

$$1 - \delta \leq \left\| \sum_{i=1}^k t_i v_i \right\| \leq 1 + \delta$$

pour tout $t \in \mathcal{S}$. Le lemme 5.5.1 de Dvoretzky-Rodgers permet de conclure alors la démonstration. $\square$

Remarque. De nombreux travaux ont poursuivi l'étude de ce résultat, notamment vis à vis de la dépendance en ϵ de la fonction $\eta(\epsilon)$. Nous renvoyons vers les travaux de Schechtmann [176], Paouris et al. [162], etc ...

5.6 Références historiques

Les notes qui suivent sont principalement tirées de [139, 44]; nous renvoyons le lecteur vers ces ouvrages pour plus de détails.

La théorie de la concentration de la mesure s'est développée durant le siècle dernier et s'est imposée comme un outil incontournable dans différents domaines des mathématiques : théorie des probabilités, statistiques, géométrie, analyse fonctionnelle ... [139]. Historiquement, cette théorie est née au début des années soixante-dix avec une nouvelle preuve, reposant sur l'inégalité de Lévy (de nature isopérimétrique), de Milman, du Théorème de Dvoretzky [155] sur les sections sphériques de domaines convexes en grande dimension. Durant les années 1980 et 1990, cette théorie fut révolutionnée par Talagrand et ses remarquables travaux sur le phénomène de la concentration de la mesure dans les espaces produits [192, 184, 185]. Il proposa un regard novateur sur la notion d'indépendance de variables aléatoires. Il suggéra de considérer une variable aléatoire dépendant (de manière régulière) d'une grande famille de variables aléatoires comme étant quasiment constante. Cette approche lui permet d'obtenir des résultats majeurs dans cette théorie pouvant facilement s'appliquer à la résolution de problèmes complexes. Par exemple, son

outil de l'inégalité de distance convexe lui a permis de développer des applications en probabilité combinatoire, de mécaniques statistiques et de processus empiriques ainsi que l'étude du supremum de processus [187, 188, 197]. Parallèlement, une remarque de Herbst sur les inégalités de Sobolev logarithmiques permit une approche supplémentaire, via les inégalités fonctionnelles, pour étudier les propriétés de la concentration de la mesure. Il est à noter que cette approche est plus souple que l'approche isopérimétrique historique. Enfin, en 1997 une remarque de Marton [152, 151], permit une troisième approche, complémentaire à celle des inégalités fonctionnelles, par le biais de la théorie du transport optimal. Cet aspect de la concentration sera développé dans un chapitre ultérieur.

Concernant le problème isopérimétrique euclidien en dimension 2, il semblerait que sa solution, les disques euclidiens, avait déjà été démontrée de manière pragmatique par la reine Didon de Carthage. Comme énoncé dans ce chapitre, ce résultat se généralise à l'espace euclidien, $\mathbb{R}^n$, pour lequel les ensembles minimisant la mesure de bord à volume fixé sont les boules euclidiennes. Bien que cette solution semble naturelle, la résolution rigoureuse de ce problème isopérimétrique n'a été trouvée qu'à la fin du 19^{ième} siècle par Brunn, Minkowski puis Lusternik [149].

Son analogue sphérique a été obtenu par Lévy au début des années 20 [146] puis, par Schmidt [177], de manière rigoureuse dans les années 40, par des techniques délicates de symétrisation. Au milieu des années 70, en combinant le lemme de Poincaré (justifiant que la mesure gaussienne pouvait être obtenue comme limite de la mesure uniforme sur la sphère) avec le résultat de Lévy, les mathématiciens Sudakov et Tsirelson [182], ainsi que Borell, [40] de manière indépendante, résolvent le problème isopérimétrique dans l'espace gaussien $(\mathbb{R}^n, \|\cdot\|_2, \gamma_n)$ et démontrent que les ensembles extrémaux sont des demi-espaces H . En 1983, Ehrhard propose une démonstration plus intrinsèque de ce résultat ; celle-ci repose sur une version gaussienne de l'inégalité de Brunn-Minkowski (cf. [92]). Bien que nous ne l'ayons pas présenté dans ce chapitre, il est important de mentionner certains résultats de Gromov [111, 113, 112] dans ce domaine. Il a notamment obtenu un théorème (de comparaison) qui permet de montrer que la mesure (renormalisée) d'une variété riemannienne dont la courbure de Ricci est bornée inférieurement par une constante strictement positive vérifie des propriétés de concentration au moins aussi bonnes que la mesure uniforme sur la sphère.

Dans ses travaux novateurs, Milman utilisait les propriétés de concentration de la mesure uniforme sur la sphère (plutôt que celle de la mesure gaussienne) pour proposer une nouvelle démonstration du théorème de Dvoretzky (cf. [155]). De nombreux théorèmes similaires ont ensuite été obtenus durant les années 1970 – 1990 en théorie locale des espaces de Banach ; citons en particulier, la contribution importante de Figiel, Lindenstrauss and Milman dans laquelle ils établirent un lien avec la notion de *type/cotype* associée à un espace de Banach (cf. [148, 143]). De nombreux domaines mathématiques se sont inspirés de cette construction probabiliste : en statistique, pour des problèmes d'estimations sous certaines contraintes (méthode du LASSO par exemple, cf [19]) ou encore en théorie des équations aux dérivées partielles (les travaux de Burq et Lebeaux [55] ou encore le survol [206]).

Chapitre 6

Principes de base et inégalités exponentielles classiques

Après la brève introduction théorique et historique du chapitre précédent, nous allons nous attarder sur des exemples classiques. A ce titre, nous débuterons par les inégalités de concentration les plus simples : celles portant sur la somme de variables aléatoires indépendantes. Dans ce qui suit, Z désignera une variable aléatoire intégrable :

$$Z : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

et nous souhaitons obtenir une majoration pertinente des queues de déviations. Autrement dit, nous cherchons à établir des inégalités de la forme

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq t) \leq \theta_1(t) \quad \text{et} \quad \mathbb{P}(Z - \mathbb{E}[Z] \leq -t) \leq \theta_2(t) \quad \text{pour tout } t \geq 0$$

avec $t \mapsto \theta_i(t) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, pour $i = 1, 2$, des fonctions positives qui tendent vers 0 à l'infini.

Comme nous l'avons mentionné dans l'introduction du chapitre précédent, les fonctions $(\theta_i)_{i=1,2}$ sont typiquement des fonctions de la forme $t \mapsto Ce^{-ct^p}$ où $p > 0$ avec $c, C > 0$ des constantes numériques. Les inégalités précédentes sont désignées sous l'appellation *déviations (par rapport à la moyenne) à droite (resp. à gauche) de Z* et, lorsque de telles inégalités sont établies, elles permettent d'en déduire une inégalité de concentration qui prend la forme suivante

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 2 \max(\theta_1(t), \theta_2(t)) \quad \text{pour tout } t \geq 0.$$

Dans ce chapitre, nous allons donc présenter différentes inégalités de concentration classique. Nous débuterons par l'inégalité de Markov [6.1.1](#), puis nous présenterons la méthode de Cramér-Chernoff [6.2.1](#) (déjà employée dans le chapitre [3](#)). Nous nous attarderons ensuite sur la notion de variables aléatoires sous-gaussiennes ainsi que l'inégalité d'Hoeffding [6.3.1](#) pour enfin conclure avec l'inégalité de Bennett [6.4.2](#) et de Bernstein [6.4.1](#). De nombreuses applications ponctueront le chapitre.

6.1 Inégalité de Markov et ses conséquences

L'inégalité de Markov fait partie des inégalités de déviation les plus célèbres et ses conséquences sont des points essentiels de la théorie.

Proposition 6.1.1. *Soit Y une variable aléatoire réelle et positive. Alors, pour tout $t > 0$, nous avons*

$$\mathbb{P}(Y \geq t) \leq \frac{1}{t} \int_{\{Y \geq t\}} Y d\mathbb{P} \leq \frac{1}{t} \mathbb{E}[Y].$$

Démonstration. Soit $t > 0$, il suffit d'observer que l'inégalité suivante est satisfaite de manière évidente :

$$1_{\{Y \geq t\}} \leq \frac{Y}{t} 1_{\{Y \geq t\}}$$

puis d'intégrer celle-ci par rapport à $d\mathbb{P}$. □

Remarque. 1. Le procédé décrit plus haut se généralise aisément. En effet, considérons

$$\phi : I \rightarrow \mathbb{R}$$

une fonction positive, croissante, définie sur un sous-ensemble $I \subset \mathbb{R}$; si $Y \in I$ p.s. alors nous avons

$$\mathbb{P}(Y \geq t) \leq \frac{1}{\phi(t)} \mathbb{E}[\phi(Y)] \quad \text{pour tout } t \in I.$$

Bien entendu, cette majoration est pertinente uniquement si le membre de droite (de l'inégalité précédente) est inférieur ou égal à 1. En général, il est coutume de choisir $\phi(t) = t^q$ avec $q > 1$ ou $\phi(t) = e^{\lambda t}$ avec $\lambda > 0$.

2. Bien entendu, les inégalités de concentration sont étroitement liées aux moments d'une variable aléatoire. En effet, les résultats suivants sont satisfaits :

- S'il existe $q > 0$ et $C > 0$ tel que

$$\mathbb{P}(|Z| \geq t) \leq \frac{C}{t^q} \quad \text{pour tout } t > 0$$

alors $\mathbb{E}[|Z|^p]$ pour tout $0 < p < q$.

- S'il existe $q > 0$ tel que $\mathbb{E}[|Z|^q] < +\infty$ alors

$$\mathbb{P}(|Z| \geq t) \leq \frac{\mathbb{E}[|Z|^q]}{t^q} \quad \text{pour tout } t > 0.$$

- De manière similaire, les assertions ci-dessous sont équivalentes.

(a) Il existe des constantes $(C, c, p) \in (\mathbb{R}_+^*)^3$ telles que $\mathbb{P}(Z \geq t) \leq Ce^{-ct^p}$ avec $t \geq 0$.

(b) Il existe $c' > 0$ telle que $\mathbb{E}[e^{c'Z^p}] < +\infty$.

Voici quelques exemples classiques d'inégalités de concentrations obtenues à partir des conséquences de l'inégalité de Markov [6.1.1](#).

Exemple 6.1.1 (Bienaymé-Tchebychev). Si $\phi(t) = t^2$ et $Y = |Z - \mathbb{E}[Z]|$, nous obtenons

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq t\right) \leq \frac{1}{t^2} \text{Var}(Z) \quad \text{pour tout } t > 0. \quad (6.1.1)$$

Cette inégalité de déviation permet d'obtenir une quantification, non-asymptotique, du théorème de la loi forte des grands nombres. En effet, supposons que $Z = \sum_{i=1}^n X_i$ avec $(X_i)_{i=1, \dots, n}$ une suite de variables aléatoires i.i.d. de carré intégrable, alors nous avons

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right]\right| \geq t\right) \leq \frac{n}{t^2} \text{Var}(X_1) \quad \text{pour tout } t > 0.$$

Le choix $t = u\sqrt{n}$ indique alors que les fluctuations de Z sont d'ordre $\sqrt{n}$.

Exemple 6.1.2 (Chernoff). Si $\phi(t) = e^{\lambda t}$ (avec $\lambda > 0$) et $Z = \sum_{i=1}^n X_i$ avec $(X_i)_{i=1, \dots, n}$ une suite de variables aléatoires i.i.d. (centrées) admettant des moments de tout ordre, nous obtenons, pour tout $t \geq 0$,

$$\begin{aligned} \mathbb{P}(Z \geq t) &\leq e^{-\lambda t} \prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}] \\ &= e^{-\lambda t} \mathbb{E}[e^{\lambda X_1}]^n \\ &= e^{-\lambda t + n\Lambda_{X_1}(\lambda)} \end{aligned}$$

où, rappelons-le, $\Lambda_{X_1}(\lambda) = \log \mathbb{E}[e^{\lambda X_1}]$ avec $\lambda \in \mathbb{R}$ désigne la log-Laplace de la variable X_1 . Le lecteur attentif aura remarqué que la majoration précédente est proche de celle employée dans la démonstration du théorème de Cramér 3.2.1. Nous reviendrons sur ce point dans la section suivante.

6.2 Méthode de Cramér-Chernoff

L'emploi de l'inégalité de Chernoff permet d'obtenir la proposition suivante. Implicitement nous supposons que la variable aléatoire Z admet des moments de tout ordre (de sorte que la transformée de Laplace est finie sur $\mathbb{R}$).

Proposition 6.2.1 (Cramér-Chernoff). *Soit Z une variable aléatoire réelle alors, pour tout $t \geq 0$, nous avons*

$$\mathbb{P}(Z \geq t) \leq e^{-\sup_{\lambda \geq 0} (\lambda t - \Lambda_Z(\lambda))} = e^{-\Lambda_Z^*(t)}$$

où $\Lambda_Z^* = \sup_{\lambda \geq 0} \lambda t - \Lambda_Z(\lambda)$ désigne la transformée de Fenchel-Legendre de Λ_Z .

Remarque. 1. Il est intéressant d'observer que le membre de gauche de l'inégalité précédente est un objet purement probabiliste tandis que le membre de droite est un objet analytique.
2. Cette inégalité est précisément celle que nous avons utilisée pour démontrer la majoration dans le théorème de Cramér.

Voici quelques exemples d'applications de la méthode de Cramér-Chernoff 6.2.1 pour des lois usuelles.

Exemple 6.2.1 (Loi normale). Si $Z \sim \mathcal{N}(0, 1)$ alors $\mathbb{P}(Z \geq t) \leq e^{-\frac{t^2}{2}}$ pour tout $t \geq 0$.

Remarque. La majoration ci-dessus est presque du même ordre que les bornes de Mills (cf. [44, 14]) :

$$\frac{t}{(1+t^2)\sqrt{2\pi}} e^{-\frac{t^2}{2}} \leq \mathbb{P}(Z \geq t) \leq \frac{1}{t\sqrt{2\pi}} e^{-\frac{t^2}{2}} \quad \text{pour tout } t > 0.$$

Exemple 6.2.2 (Loi de Poisson). Si $Y \sim \mathcal{P}_o(\theta)$ avec $\theta > 0$ et $Z = Y - \theta$, nous avons

$$\Lambda_Z(\lambda) = \theta(e^\lambda - \lambda - 1) \quad \text{pour tout } \lambda \in \mathbb{R}$$

et, par suite,

$$\Lambda_Z^*(t) = \theta h\left(\frac{t}{\theta}\right) \quad \text{pour tout } t \geq 0$$

avec $h(x) = (1+x)\log(1+x) - x$ pour $x > -1$.

Remarque. Observons que h admet les comportements asymptotiques suivants :

- lorsque $x \rightarrow 0$, $h(x) \sim x^2$ ce qui correspond au comportement gaussien issu du théorème de la limite centrale ;
- lorsque $x \rightarrow +\infty$, $h(x) \sim x \log x$ ce qui correspond au comportement issu du théorème de la limite centrale poissonnienne. En effet, rappelons le résultat suivant (cf. [22]) :

Théorème 6.2.1 (TCL poissonnien). Soit $(X_i)_{1 \leq i \leq n}$ une famille de variables aléatoires indépendantes suivant une loi de Bernoulli $\mathcal{B}_e(p_i)$ avec $p_i \in]0, 1[$ pour tout $i \in \{1, \dots, n\}$. Si $S_n = \sum_{i=1}^n X_i$ et, lorsque $n \rightarrow \infty$, les hypothèses suivantes sont vérifiées

$$\max_{i=1, \dots, n} p_i \rightarrow 0 \quad \text{et} \quad \mathbb{E}[S_n] = \sum_{i=1}^n p_i \rightarrow \lambda < +\infty.$$

Alors, lorsque $n \rightarrow +\infty$,

$$S_n \rightarrow \mathcal{P}_o(\lambda) \quad \text{en loi.}$$

Exemple 6.2.3 (Loi de Bernoulli). Si $Y \sim \mathcal{B}_e(p)$ avec $p \in [0, 1]$ et $Z = Y - p$ alors

$$\Lambda_Z(\lambda) = \log(pe^\lambda + 1 - p) - \lambda p \quad \text{pour tout } \lambda \in \mathbb{R}$$

et

$$\Lambda_Z^*(t) = (1-p-t) \log\left(\frac{1-p-t}{1-p}\right) + (p+t) \log\left(\frac{p+t}{p}\right) \quad \text{pour tout } t \in [0, 1-p].$$

Exemple 6.2.4 (Majoration principe grande déviation). Si $Z = X_1 + \dots + X_n$ avec $(X_i)_{i=1, \dots, n}$ une suite de variables i.i.d. nous avons

$$\Lambda_Z(\lambda) = n\Lambda_{X_1}(\lambda) \quad \text{pour tout } \lambda \in \mathbb{R} \quad \text{et} \quad \Lambda_Z^*(t) = n\Lambda_{X_1}^*\left(\frac{t}{n}\right) \quad \text{pour tout } t \geq 0.$$

En particulier, $\mathbb{P}(Z \geq t) \leq e^{-n\Lambda_{X_1}^*\left(\frac{t}{n}\right)}$ pour tout $t \geq 0$. Il s'agit précisément de la majoration employée lors de la démonstration du théorème de Cramer 3.2.1.

6.3 Variables aléatoires sous-gaussiennes

Rappelons le fait suivant : si $Z \sim \mathcal{N}(0, \sigma^2)$ alors $\Lambda_Z(\lambda) = \frac{\lambda^2 \sigma^2}{2}$ pour tout $\lambda \in \mathbb{R}$. Nous allons voir qu'il est possible d'introduire une classe de variables aléatoires qui vérifie une propriété similaire et donne lieu à des conséquences intéressantes de stabilité.

Définition 6.3.1 (Loi sous-gaussienne). *Une variable aléatoire Z est dite sous-gaussienne de paramètre $\sigma^2 > 0$ si*

$$\Lambda_Z(\lambda) \leq \frac{\lambda^2 \sigma^2}{2} \quad \text{pour tout } \lambda \in \mathbb{R}.$$

Remarque. 1. Le contrôle précédent sur la Λ_Z entraîne que $\sigma^2 \leq \text{Var}(Z)$. En revanche, il ne s'agit pas en général d'une égalité.

2. Par l'argument de Chernoff, une variable aléatoire sous-gaussienne de paramètre σ^2 vérifie une propriété de concentration au moins gaussienne. C'est-à-dire,

$$\mathbb{P}(Z \geq t) \leq e^{-\frac{t^2}{2\sigma^2}} \quad \text{pour tout } t > 0.$$

3. Si Z est une variable aléatoire sous-gaussienne de paramètres σ^2 , $-Z$ l'est également (avec le même paramètre σ^2). Par suite,

$$\mathbb{P}(|Z| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}} \quad \text{pour tout } t > 0.$$

4. Il est également possible d'exhiber la classe des variables aléatoires dites « sous-gamma » de paramètres $(a, b) \in]0; +\infty[^2$. Celles-ci vérifient alors les mêmes propriétés de concentration qu'une loi gamma standard $\Gamma(a, b)$. Ces variables aléatoires vérifient alors des résultats de stabilité similaire aux lois sous-gaussiennes. Néanmoins, nous n'aborderons pas cet aspect dans ce cours et renvoyons le lecteur vers [44] pour plus de détails.

Voici quelques propriétés vérifiées par des variables aléatoires sous-gaussiennes.

Proposition 6.3.1. 1. *L'ensemble des variables aléatoires sous-gaussiennes est stable par addition. Plus précisément, si $X_1, \dots, X_n$ sont mutuellement indépendantes et sous-gaussiennes de paramètres respectifs $\sigma_i^2 > 0$ alors $Z = \sum_{i=1}^n X_i$ est une variable aléatoire sous-gaussienne de paramètre $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.*

2. *Il est possible de caractériser les variables aléatoires sous-gaussiennes d'après la croissance de leurs moments. Plus précisément, Z est une variable aléatoire sous-gaussienne si et seulement si il existe $C > 0$ telle que*

$$\|Z\|_q \leq C\sqrt{q} \quad \text{pour tout } q \geq 1$$

$$\text{où } \|Z\|_q = \left(\int_{\Omega} |Z|^q d\mathbb{P} \right)^{1/q}.$$

Remarque. Notons que la première propriété se comporte médiocrement vis-à-vis de la dimension (i.e. du nombre de variables impliquées).

Comme nous allons le voir, le résultat suivant nous assure que les variables bornées (p.s.) sont sous-gaussiennes.

Proposition 6.3.2 (Hoeffding). *Si Z est une variable aléatoire réelle, centrée, telle que*

$$Z \in [a, b] \quad \text{p.s.} \quad \text{avec} \quad a, b \in \mathbb{R}$$

alors Z est une variable aléatoire sous-gaussienne de paramètre $\sigma^2 = \frac{(b-a)^2}{4}$.

Démonstration. La démonstration repose sur deux arguments : un changement de mesure combiné à un développement de Taylor de la transformée de Laplace. Avant de mettre en place ceci, il va être utile d'observer le fait élémentaire suivant.

La variance de Z est bornée par σ^2 . En effet, puisque $Z \in [a, b]$ p.s. nous avons

$$\left| Z - \frac{b+a}{2} \right| \leq \frac{b-a}{2}.$$

En particulier, $\text{Var}(Z) = \text{Var}(Z - \frac{b+a}{2}) \leq \sigma^2$.

Nous pouvons à présent procéder au changement de mesure. Si $\mathbb{P} = \mathcal{L}(Z)$ nous désignerons par $\mathbb{P}_\lambda$ (avec $\lambda \in \mathbb{R}$) la mesure de probabilité définie par

$$d\mathbb{P}_\lambda = \frac{e^{\lambda Z}}{\mathbb{E}[e^{\lambda Z}]} d\mathbb{P}.$$

Dans ce qui suit, pour tout $\lambda \in \mathbb{R}$, Z_λ correspondra à une variable aléatoire de loi $\mathbb{P}_\lambda$. Observons au passage que $Z_\lambda \in [a, b]$ p.s. pour tout $\lambda \in \mathbb{R}$.

Comme annoncé, nous allons utiliser un développement de Taylor (à l'ordre 2) de Λ_Z . Celui-ci nous assure que, pour tout $\lambda > 0$ il existe $\theta \in]0, \lambda[$ tel que

$$\Lambda_Z(\lambda) = \Lambda_Z(0) + \lambda \Lambda'_Z(0) + \frac{\lambda^2}{2} \Lambda''(\theta).$$

Or, des calculs élémentaires montrent que $\Lambda'_Z(\lambda) = \mathbb{E}[Z_\lambda]$ et $\Lambda''_Z(\lambda) = \text{Var}(Z_\lambda)$. En conséquence, puisque Z est centré et $Z_\theta \in [a, b]$ p.s., nous avons

$$\Lambda_Z(\lambda) = \frac{\lambda^2}{2} \Lambda''(\theta) \leq \frac{\lambda^2 (b-a)^2}{8}.$$

D'où, $\mathbb{E}[e^{\lambda Z}] \leq e^{\frac{\lambda^2 (b-a)^2}{8}}$ pour tout $\lambda > 0$. Il suffit ensuite d'utiliser l'inégalité de Chernoff 6.1.2 pour enfin optimiser en λ afin d'obtenir l'inégalité annoncée. Le cas $\lambda < 0$ se traite de manière similaire en remplaçant Z par $-Z$ qui est une variable aléatoire à valeurs dans $[-b, -a]$. $\square$

Ce dernier résultat permet d'obtenir un corollaire évident.

Corollaire 6.3.1 (Inégalité d'Hoeffding). *Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes telles que, pour tout $i = 1, \dots, n$, il existe des constantes $a_i, b_i \in \mathbb{R}$ telles que*

$$X_i \in [a_i, b_i] \quad \text{p.s.}$$

alors

$$\mathbb{P}(S_n \geq t) \leq e^{-\frac{2t^2}{\sigma_n^2}} \quad \text{pour tout} \quad t > 0, \quad (6.3.1)$$

où $S_n = \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$ et $\sigma_n^2 = \sum_{i=1}^n (b_i - a_i)^2$.

Remarque. Le lecteur a sûrement déjà rencontré ce genre d'inégalités dans un cours de statistiques. En effet, ce type d'inégalité de concentration est fort utile pour établir des régions de confiance. Nous n'aborderons pas cet aspect par la suite; pour plus de détails à ce sujet, nous renvoyons le lecteur vers l'excellent cours [42].

Ceci permet d'obtenir, dans le cadre des variables bornées, une inégalité de concentration reflétant de manière non asymptotique le théorème de la limite centrale.

Exemple 6.3.1. Si $X_1, \dots, X_n$ sont des variables aléatoires i.i.d. telles que

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2} \quad (\text{i.e des variables aléatoires de Rademacher})$$

alors, d'après l'inégalité de Hoeffding (6.3.1), nous avons

$$\mathbb{P}(S_n \geq t) \leq e^{-\frac{t^2}{2n}} \quad \text{pour tout } t > 0.$$

D'où, en posant $r = \frac{t}{\sqrt{n}}$, nous avons

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \geq r\right) \leq e^{-\frac{r^2}{2}} \quad \text{pour tout } r > 0.$$

Voici deux autres applications de l'inégalité de Hoeffding (6.3.1).

Exemple 6.3.2 (Loi forte des grands nombres). Le lecteur est bien entendu familier avec le théorème de Kolmogorov (cf. [22]) qui s'énonce comme suit. Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d. telles que $\mathbb{E}[|X_1|] < +\infty$. Si $S_n = \sum_{i=1}^n X_i$ alors

$$\lim_{n \rightarrow +\infty} \frac{S_n}{n} = \mathbb{E}[X_1] \quad p.s.$$

Sans perdre en généralité, nous pouvons supposer que les variables sont centrées. En particulier, $\mathbb{E}[X_1] = 0$. Nous allons maintenant démontrer la loi forte des grands nombres à partir de l'inégalité de Hoeffding (6.3.1). Pour cela, nous allons tout d'abord supposer que $X_1 \in [a, b]$ p.s. Ainsi, d'après l'inégalité de Hoeffding nous avons

$$\mathbb{P}(|S_n| \geq t) \leq 2e^{-\frac{2t^2}{n(b-a)^2}} \quad \text{pour tout } t > 0.$$

Implicite-ment nous avons aussi utilisé l'inégalité de Hoeffding (6.3.1) pour la variable $-X_1$ pour obtenir cette inégalité de concentration. Posons à présent $u = tn$, ceci mène à

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq u\right) \leq 2e^{-\frac{2u^2n}{(b-a)^2}} \quad \text{pour tout } u > 0.$$

En particulier, le membre de droite de cette inégalité est le terme général d'une suite sommable. En conséquence, le lemme de Borel-Cantelli (cf. lemme 1.1.1 ou [22]) nous assure que

$$\lim_{n \rightarrow +\infty} \frac{S_n}{n} = 0 \quad p.s.$$

Traisons à présent le cas général. Puisque $\mathbb{E}[|X|] < +\infty$, par construction de l'intégrale de Lebesgue (les fonctions mesurables peuvent être approchées par des fonctions étagées), il est possible

d'approcher X par une variable aléatoire étagée. Plus précisément, pour tout $\epsilon > 0$ fixé, il existe Y_1 étagée telle que

$$\mathbb{E}[|X_1 - Y_1|] \leq \epsilon.$$

Soit $(Y_i)_{i=1,\dots,n}$ une famille de copies indépendantes de Y_1 ; par la suite, nous utiliserons la notation $\tilde{S}_n = \sum_{i=1}^n Y_i$. Ceci mène à

$$\frac{|S_n|}{n} \leq \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| + \frac{|\tilde{S}_n|}{n}.$$

D'après la première étape de la démonstration nous savons que $\limsup_{n \rightarrow +\infty} \frac{|\tilde{S}_n|}{n} = 0$, pour conclure nous devons alors montrer que

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n |X_i - Y_i|$$

peut-être contrôlée par une quantité arbitrairement petite lorsque ϵ est arbitrairement petit. Pour cela, nous allons employer le lemme suivant (cf. [22] page 134 – 135) :

Lemme 6.3.1. *Soit Z une variable aléatoire positive et intégrable et considérons $(Z_i)_{i \geq 1}$ une famille de copies indépendantes. Alors pour tout $k \geq 0$ et tout $\delta > 0$*

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n Z_i \leq 2\mathbb{E}[Z] \quad p.s.$$

La conclusion s'ensuit aisément en choisissant $Z_i = |X_i - Y_i|$ pour tout $i \geq 1$. En effet, d'après le lemme 6.3.1, nous avons

$$\begin{aligned} \frac{|S_n|}{n} &\leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| + \limsup_{n \rightarrow +\infty} \frac{|\tilde{S}_n|}{n} \\ &\leq 2\mathbb{E}[|X_1 - Y_1|] \\ &\leq 2\epsilon \quad p.s. \end{aligned}$$

ϵ étant arbitraire, ceci conclut la preuve. Il ne reste plus qu'à démontrer le lemme 6.3.1.

Démonstration. (du lemme 6.3.1) La démonstration repose sur un argument de bloc et l'utilisation du lemme de Borel-Cantelli 1.1.1. Soit $k \geq 0$ et $\delta > 0$, en découpant suivant le seuil 2^k nous avons

$$\begin{aligned} \mathbb{P}\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{i=1}^n Z_i \geq 2\mathbb{E}[Z] + \delta\right) &\leq \mathbb{P}\left(\exists i \in \{1, 2, \dots, 2^{k+1}\} : Z_i > 2^k\right) \\ &+ \mathbb{P}\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{i=1}^n Z_i 1_{[0, 2^k]}(Z_i) \geq 2\mathbb{E}[Z] + \delta\right). \end{aligned}$$

En utilisant une borne d'union pour la première probabilité et en ajoutant les 2^k termes manquants pour la deuxième probabilité, le membre de droite de l'inégalité précédente est contrôlé par

$$\begin{aligned} 2^{k+1}\mathbb{P}(Z > 2^k) + \mathbb{P}\left(\sum_{i=1}^{2^{k+1}} Z_i 1_{[0,2^k]}(Z_i) \geq 2^{k+1}\mathbb{E}[Z] + \delta 2^k\right) &\leq 2^{k+1}\mathbb{P}(Z > 2^k) \\ &+ \mathbb{P}\left(\sum_{i=1}^{2^{k+1}} Z_i 1_{[0,2^k]}(Z_i) 2^{k+1} - \mathbb{E}[Z_i 1_{[0,2^k]}(Z_i)] \geq \delta 2^k\right). \end{aligned}$$

Par suite, en appliquant l'inégalité de Tchebychev (6.1.1) au deuxième terme du majorant précédent, nous obtenons que

$$\mathbb{P}\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{i=1}^n Z_i \geq 2\mathbb{E}[Z] + \delta\right) \leq 2^{k+1}\mathbb{P}(Z > 2^k) + \frac{2}{\delta^2 2^k} \mathbb{E}[Z^2 1_{[0,2^k]}(Z)]. \quad (6.3.2)$$

Nous allons maintenant montrer que $\sum_{k \geq 0} 2^{k+1}\mathbb{P}(Z > 2^k) + \frac{2}{\delta^2 2^k} \mathbb{E}[Z^2 1_{[0,2^k]}(Z)] < +\infty$ pour ensuite appliquer le lemme de Borel-Cantelli 1.1.1.

D'une part,

$$\sum_{k \geq 0} 2^{k+1}\mathbb{P}(Z > 2^k) \leq 4\mathbb{E}[Z]. \quad (6.3.3)$$

En effet, pour établir ceci il suffit d'observer que

$$\mathbb{E}[Z] = \int_0^\infty \mathbb{P}(Z > t) dt = \sum_{k \geq 0} \int_{2^k}^{2^{k+1}} \mathbb{P}(Z > t) dt \geq \sum_{k \geq 0} 2^k \mathbb{P}(Z > 2^{k+1})$$

où la dernière minoration s'obtient en utilisant la décroissance de $t \mapsto \mathbb{P}(Z > t)$ sur l'intervalle $[2^k, 2^{k+1}]$. D'autre part, nous avons aussi

$$\sum_{k \geq 0} 2^{-k} \mathbb{E}[Z^2 1_{[0,2^k]}(Z)] = \mathbb{E}\left[Z^2 \sum_{k \geq 0} 2^{-k} 1_{[0,2^k]}(Z)\right] \leq 4\mathbb{E}[Z]. \quad (6.3.4)$$

Pour établir ceci, il faut tirer profit du découpage dyadique :

- s'il existe $l \geq 0$ tel que $2^l < Z \leq 2^{l+1}$ alors

$$Z^2 \sum_{k \geq 0} 2^{-k} 1_{[0,2^k]}(Z) \leq 2^{2l+2} \sum_{k \leq l+1} 2^{-k} \leq 4Z$$

puisque dans la première somme de nombreux termes sont nuls.

- si $0 \leq Z \leq 1$, la conclusion reste la même.

Finalement, en combinant (6.3.2) avec (6.3.3) et (6.3.4), nous avons montré que

$$\mathbb{P}\left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{i=1}^n Z_i \geq 2\mathbb{E}[Z] + \delta\right) \leq 4(1 + 2\delta^{-2})\mathbb{E}[Z] < +\infty.$$

D'après le lemme de Borel-Cantelli 1.1.1, presque sûrement, pour k suffisamment grand, nous avons donc

$$\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{i=1}^n Z_i < 2\mathbb{E}[Z] + \delta.$$

Puisque δ est arbitraire,

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n Z_i \leq 2\mathbb{E}[Z] \quad p.s.$$

ce qui conclut la preuve du lemme. $\square$

Exemple 6.3.3 (Version faible de la loi du logarithme itéré). Considérons $X_1, \dots, X_n$ des variables aléatoires de Rademacher i.i.d.. Nous allons montrer que

$$\limsup_{n \rightarrow +\infty} \frac{S_n}{\sqrt{2n \log n}} \leq 1 \quad p.s.$$

où, à nouveau, $S_n = \sum_{i=1}^n X_i$. Pour cela, il suffit d'utiliser l'inégalité de Hoeffding (6.3.1). En effet, pour tout $\epsilon > 0$, celle-ci entraîne que

$$\mathbb{P}(S_n \geq (1 + \epsilon)\sqrt{2n \log n}) \leq e^{-(1+\epsilon)^2 \log n} = \frac{1}{n^{(1+\epsilon)^2}}.$$

Le membre de droite de l'inégalité précédente est sommable, le lemme de Borel-Cantelli (cf. [22]) implique alors que

$$\limsup_{n \rightarrow +\infty} \frac{S_n}{\sqrt{2n \log n}} \leq 1 + \epsilon \quad p.s.$$

Il suffit ensuite de faire $\epsilon \rightarrow 0$.

Remarque. Le résultat précédent soulève quelques questions auxquelles nous invitons le lecteur à répondre. Nous savons, d'après la loi forte des grands nombres, que

$$\frac{1}{n} S_n \rightarrow \mathbb{E}[X_1] \quad p.s. \quad \text{lorsque } n \rightarrow +\infty.$$

Nous savons aussi que

$$\frac{1}{\sqrt{n}} S_n \rightarrow \mathcal{N}(0, 1) \quad \text{en loi lorsque } n \rightarrow +\infty.$$

Si $1 < p < 2$ que dire de la convergence de $\frac{1}{n^{p/2}} S_n$? A quelle place se trouve la loi du logarithme itéré parmi ces résultats?

Avant de passer à la section suivante, il est important de souligner certaines faiblesses de l'inégalité de Hoeffding (6.3.1) :

- la taille de la variance $\text{Var}(Z)$ n'apparaît pas dans celle-ci.
- Le théorème limite poissonien (cf. théorème 6.2.1) n'est également pas reflété dans cette inégalité.

Comme nous allons le voir, au prix d'hypothèses supplémentaires, il est possible de raffiner et renforcer l'inégalité d'Hoeffding.

6.4 Inégalités de Bernstein et de Bennett

Dans cette section, $X_1, \dots, X_n$ désignent toujours une suite de variables aléatoires mutuellement indépendantes. Nous supposons de plus qu'il existe $b > 0$ telle que, pour tout $i = 1, \dots, n$, $X_i \leq b$ p.s et posons $\sigma^2 = \sum_{i=1}^n \mathbb{E}[X_i^2]$.

Sous les hypothèses précédentes, nous avons l'inégalité de Bernstein suivante.

Théorème 6.4.1 (Bernstein). *Pour tout $t \geq 0$,*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) \leq e^{-\frac{t^2}{2\sigma^2(1+\frac{bt}{3})}}$$

avec $S_n = \sum_{i=1}^n X_i$.

Remarque. 1. Si t est petit (par rapport à σ^2) le membre de droite est équivalent à $e^{-t^2/2\sigma^2}$ et reflète le comportement gaussien du théorème de la limite centrale.

2. Au contraire, lorsque t est grand (par rapport à σ^2), le membre de droite est équivalent à e^{-t} . Cette partie ne reflète pas convenablement le théorème 6.2.1.

3. Bien entendu, si $|X_i| \leq b$ pour tout $i = 1, \dots, n$, il est possible d'obtenir l'inégalité de concentration

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2(1+\frac{bt}{3})}} \quad \text{pour tout } t \geq 0.$$

Démonstration. Le début de la démonstration repose sur l'inégalité de Chernoff 6.1.2 et l'hypothèse d'indépendance. Ainsi, pour tout $t \geq 0$ et tout $\lambda > 0$, nous obtenons

$$\begin{aligned} \mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) &\leq e^{-\lambda t} \prod_{i=1}^n \mathbb{E}[e^{\lambda(X_i - \mathbb{E}[X_i])}] \\ &= e^{-\lambda t + \sum_{i=1}^n \log(\mathbb{E}[e^{\lambda X_i}]) - \lambda \mathbb{E}[S_n]} \\ &\leq e^{-\lambda t + \sum_{i=1}^n \mathbb{E}[\phi(\lambda X_i)]} \end{aligned}$$

avec $\phi(u) = e^u - u - 1$ puisque $\log(u) \leq u - 1$ pour tout $u > 0$.

La suite de la démonstration va reposer sur différentes manières de contrôler la famille $(\phi(\lambda X_i))_{i=1, \dots, n}$ suivant le signe des variables indépendantes $(X_i)_{i=1, \dots, n}$.

Puisque $\phi(u) \leq \frac{u^2}{2}$ lorsque $u < 0$, nous avons, pour tout $i = 1, \dots, n$,

$$\phi(\lambda X_i) \leq \lambda^2 \frac{X_i^2}{2} \quad \text{lorsque } X_i < 0.$$

Lorsque $u > 0$, un développement de Taylor de la fonction ϕ entraîne que, pour tout $i = 1, \dots, n$, nous avons (lorsque $X_i > 0$)

$$\begin{aligned} \phi(\lambda X_i) &\leq \frac{\lambda^2}{2} X_i^2 + \frac{\lambda^3}{3!} X_i^3 + \dots \\ &\leq \frac{\lambda^2}{2} X_i^2 \left(1 + \frac{\lambda b}{3} + \frac{\lambda^2 b^2}{12} + \dots \right). \end{aligned}$$

Notons que nous avons utilisé l'hypothèse « $X_i \leq b$ pour tout $i = 1, \dots, n$, » afin d'obtenir la dernière inégalité. En outre, lorsque $\frac{\lambda b}{3} < 1$, nous avons aussi

$$1 + \frac{\lambda b}{3} + \frac{\lambda^2 b^2}{12} + \dots \leq \frac{1}{1 - \frac{\lambda b}{3}}.$$

D'où, pour tout $t \geq 0$ et pour tout $\frac{\lambda b}{3} < 1$, nous obtenons

$$\begin{aligned} \mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) &\leq e^{-\lambda t + \frac{\lambda^2}{2(1 - \frac{\lambda b}{3})} \times \sum_{i=1}^n \mathbb{E}[X_i^2]} \\ &= e^{-\lambda t + \frac{\lambda^2 \sigma^2}{2(1 - \frac{\lambda b}{3})}}. \end{aligned}$$

Il ne reste plus qu'à optimiser en $\lambda > 0$ (sous la contrainte $\frac{\lambda b}{3} < 1$) : cela revient à choisir $\lambda = \frac{t}{\sigma^2 + \frac{bt}{3}}$. $\square$

Sous les mêmes hypothèses et par des méthodes similaires, il est possible d'affiner l'inégalité de Bernstein. Cette précision est contenue dans l'inégalité de Bennett ; cette inégalité permet alors de quantifier le théorème de la limite centrale poissonnienne 6.2.1.

Théorème 6.4.2 (Bennett). *Sous les hypothèses décrites au début de la section 6.4, nous avons, pour tout $t \geq 0$,*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) \leq e^{-\frac{\sigma^2}{b^2} h\left(\frac{bt}{\sigma^2}\right)}$$

avec $h(u) = (1 + u) \log(1 + u) - u$ pour $u > 0$.

Remarque. 1. Si $|X_i| \leq b$ pour tout $i = 1, \dots, n$, nous obtenons

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) \leq 2e^{-\frac{\sigma^2}{b^2} h\left(\frac{bt}{\sigma^2}\right)}.$$

2. Si u est petit $h(u) \sim \frac{u^2}{2}$, le membre de droite de l'inégalité de Bennett reflète toujours le comportement gaussien du théorème de la limite centrale.
3. Si u est grand $h(u) \sim u \log u$, le membre de droite de l'inégalité de Bennett reflète alors le comportement poissonnien du théorème de la limite centrale.

Pour clore ce chapitre, en guise de piste de réflexion, nous posons la question suivante :

Comment serait-il possible de généraliser les inégalités de concentration précédentes lorsque les variables $(X_i)_{i=1, \dots, n}$ ne sont plus à valeurs réelles mais à valeurs dans un espace de Banach ? Quel sens donner à X_1^2 dans un tel cadre ?

6.5 Références historiques

Ces quelques notes sont tirées de [44].

Les inégalités de concentration exponentielles pour les sommes de variables aléatoires sont connues depuis quelques temps déjà. Bernstein a présenté son inégalité en 1946, Chernoff la sienne en 1952, Bennett en 1962 et Hoeffding en 1963. De nombreuses extensions de ces inégalités sont maintenant disponibles, Azuma a par exemple étendu en 1967 l'inégalité d'Hoeffding à des martingales. Dans le chapitre suivant, nous présenterons le cas de fonctions dites « aux différences bornées » introduit par Mc Diarmid en 1989.

La loi du logarithme itéré (pour une somme de variables aléatoires i.i.d.) date des travaux de Khinchin (1924) puis ceux de Kolmogorov en 1929. De nombreuses extensions ont été obtenues depuis. Citons par exemple les travaux de Hartman-Wintner (1940) concernant des marches aléatoires, ceux de Strassen (1964) qui aborda ce sujet du point de vue des principes d'invariances. Mentionnons également les travaux de Ledoux et Talagrand [143] : dans cet ouvrage, ils obtiennent ce genre de résultat dans des espaces de Banach généraux par le biais d'inégalités de concentration.

Troisième partie

Superconcentration

Chapitre 11

Introduction

Cette dernière partie est un peu particulière car elle a pour objet de présenter une thématique apparue récemment dans le monde de la recherche mathématique (~ 2008). Celui-ci évoluant très rapidement, il y a fort à parier qu'en peu de temps certains des points que nous allons présenter dans les chapitres suivants seront devenus incomplets voir obsolètes. Néanmoins, cet inconvénient peut avoir un rôle pédagogique intéressant pour un élève de master 2 qui débute son initiation à la recherche. En parcourant les chapitres à venir, il sera alors à même de mesurer à quelle vitesse progresse la recherche mathématique, de quelle manière la communauté mathématique digère certaines théories pour n'en garder que les parties essentielles mais aussi de quelles manières certains théorèmes sont transformés et modelés pour devenir plus efficaces et plus simples d'emploi.

Les chapitres précédents ont pu mettre en valeur la théorie de la concentration de la mesure. En effet, celle-ci s'est imposée comme un outil incontournable dans de nombreux domaines des mathématiques. Cependant une telle généralité et polyvalence risque de manquer de précision dans certains cas particuliers. Lorsque cela arrive nous parlerons de superconcentration.

Chatterjee fut le premier à essayer de formaliser cette notion et à collecter différents exemples pour lesquels la théorie de la concentration est sous-optimale (cf.[61]). Les chapitres qui vont suivre ont donc pour objectif de présenter plus en détails le phénomène de superconcentration. Nous nous focaliserons, la plupart du temps, sur le cas de la mesure gaussienne γ_n sur $\mathbb{R}^n$. Dans un premier temps, nous allons utiliser un exemple simple pour illustrer ce phénomène. Ensuite, nous aborderons plusieurs modèles présentant un phénomène de superconcentration ; nous verrons à ce moment-là que ces modèles sont de nature très différente et que les méthodes mises en place pour exhiber de la superconcentration dépendent fortement de ces derniers. Nous parlerons en particulier de matrices aléatoires, de marches aléatoires branchantes, de percolation, de théorie des extrêmes et d'un modèle de verres de spin issu de la physique statistique.

La plupart du temps, les méthodes permettant d'atteindre un résultat optimal sur un modèle précis sont assez délicates et inhérentes à celui-ci. Toutefois, pour certains d'entre eux, il est possible d'élaborer des méthodes plus souples permettant d'améliorer, même modestement, les résultats fournis par la théorie classique de la concentration de la mesure. Dans certains cas, ces méthodes parviennent même à être optimales. A présent, présentons ce phénomène.

11.1 Phénomène de superconcentration

Parmi les manifestations de la concentration de la mesure satisfaite par γ_n , la mesure gaussienne standard sur $\mathbb{R}^n$, nous avons montré qu'elle satisfaisait une inégalité de Poincaré (cf. proposition 8.2.1). Rappelons le contenu de ce résultat et voyons certaines de ses conséquences.

Pour $f : \mathbb{R}^n \rightarrow \mathbb{R}$ suffisamment régulière, nous avons

$$\mathrm{Var}_{\gamma_n}(f) \leq \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma_n(x). \quad (11.1.1)$$

Le choix de la fonction $f(x) = \max_{i=1,\dots,n} x_i$ entraîne alors le résultat suivant.

Proposition 11.1.1. *Soit X un vecteur gaussien dans $\mathbb{R}^n$ de matrice de covariance Γ , alors l'inégalité suivante est vérifiée*

$$\mathrm{Var}(M_n) \leq \max_{i=1,\dots,n} \mathrm{Var}(X_i), \quad (11.1.2)$$

où $M_n = \max_{i=1,\dots,n} X_i$.

Cette inégalité est optimale et ne peut être améliorée (le cas d'égalité est atteint lorsque $n = 1$ ou lorsque $X = (X_1, \dots, X_1)$ est un vecteur gaussien ayant toutes ses coordonnées égales).

Soulignons-le, cette inégalité ne dépend pas de la matrice de covariance. Elle ne prend donc pas en compte la structure des corrélations du vecteur gaussien. Ainsi, il est possible que la généralité de l'inégalité (11.1.2) fournisse des résultats sous-optimaux pour des cas particuliers. Par exemple, voyons ce que nous apprend l'inégalité (11.1.2) lorsque $\Gamma = I_d$ la matrice identité de $\mathbb{R}^n$. Nous trouvons que

$$\mathrm{Var}(M_n) \leq 1.$$

Cependant, pour un tel choix de matrice de covariance, toutes les coordonnées du vecteur gaussien X sont indépendantes. Il est donc possible de calculer directement la variance de M_n et d'obtenir

$$\mathrm{Var}(M_n) \leq \frac{C}{\log n} \quad (11.1.3)$$

avec $C > 0$ une constante numérique. Il apparaît donc que l'inégalité (11.1.1), valable pour une large classe de fonction, provenant de la théorie, dite « classique », de la concentration, fournit une borne sous-optimale pour le choix particulier de la fonction $f(x) = \max_{i=1,\dots,n} x_i$. Selon la terminologie introduite par Chatterjee (cf. [61]), il s'agit du phénomène de *superconcentration*.

Cet exemple simple fournit la première question, abordée dans [61], que nous développerons dans les chapitres qui vont suivre : « si la borne fournie par l'inégalité de Poincaré (11.1.1) est sous-optimale, comment obtenir le bon ordre de grandeur ? ».

De manière plus formelle, la définition de la notion de superconcentration s'énonce comme suit

Définition 11.1.1. Soit μ une mesure de probabilité sur $\mathbb{R}^n$ satisfaisant une inégalité de Poincaré de constante C_p . Alors, une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *superconcentrée* s'il existe $0 < \epsilon < 1$ tel que

$$\mathrm{Var}_\mu(f) \leq \epsilon C_p \int_{\mathbb{R}^n} |\nabla f|^2 d\mu.$$

Remarque. Il faut comprendre ceci de la manière suivante : ϵ est petit et peut dépendre de la dimension de l'espace sous-jacent ; il est même envisageable que $\epsilon = \epsilon_n \rightarrow 0$ lorsque $n \rightarrow \infty$. C'est notamment le cas dans l'exemple (11.1.3) où $\epsilon_n = \frac{C}{\log n}$.

Dans l'étude qui va suivre, nous nous focaliserons principalement sur le cas de la fonction maximum, néanmoins d'autres fonctionnelles, comme la médiane, les normes l_p , ou encore l'énergie libre $F_\beta(x)$ (dont la définition sera donnée ultérieurement) sont d'autres exemples de fonctions qui exhibent un phénomène de superconcentration sous une mesure gaussienne.

Soulignons une différence notable entre la théorie classique de la concentration et ce nouveau phénomène. Une inégalité de Poincaré concerne une famille de fonctions tandis que le phénomène de superconcentration n'est valable, *a priori*, que pour une fonction particulière.

Le lecteur curieux pourrait se demander si des propriétés plus fortes de concentration satisfaites par la mesure μ peuvent améliorer la borne obtenue via (11.1.1) et fournir le bon ordre de grandeur de la variance.

Par exemple, la mesure gaussienne sur $\mathbb{R}^n$ vérifie l'inégalité de Sobolev logarithmique. En particulier, rappelons que cette inégalité, via l'argument de Herbst, entraîne la propriété de concentration suivante (cf. corollaire 4.2.1) : soient $X_1, \dots, X_n$ des variables aléatoires gaussiennes standards i.i.d. et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction lipschitzienne avec $\|f\|_{\mathrm{Lip}} \leq 1$. Alors, pour tout $t \geq 0$, nous avons

$$\mathbb{P}\left(|f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]| \geq t\right) \leq 2e^{-t^2/2}. \quad (11.1.4)$$

A nouveau, comme pour l'inégalité de Poincaré, constatons que cette inégalité est optimale dans le sens où le cas d'égalité est atteint lorsque f est une fonction linéaire. De manière similaire, la grande généralité de (11.1.4) risque de fournir des résultats sous-optimaux pour des choix de fonctions f particuliers. Ceci est notamment visible au niveau de la variance. En effet, pour toute fonction lipschitzienne f avec $\|f\|_{\mathrm{Lip}} \leq 1$, (11.1.4) implique (après intégration) que

$$\mathrm{Var}_{\gamma_n}(f) \leq 4.$$

En particulier, si nous considérons la fonction $f(x_1, \dots, x_n) = \max_{i=1, \dots, n} x_i$ nous obtenons la majoration

$$\mathrm{Var}(M_n) \leq 4.$$

Comme nous l'avons déjà vu plus haut, cette borne est sous-optimale. En conclusion, l'inégalité de concentration (11.1.4) ne reflète pas le véritable comportement du maximum au niveau de la variance.

Mettons de côté un instant ces questions à propos de la variance et posons nous la question suivante : *l'inégalité de concentration (11.1.4) pourrait-elle être pertinente vis-à-vis d'un théorème de convergence en loi ?* Il est aisé d'illustrer ce questionnement par l'exemple de la fonction maximum et la théorie des extrêmes (cf. [81, 134, 94]).

Soit $X = (X_1, \dots, X_n)$ est un vecteur gaussien standard. Il est bien connu que $M_n = \max_{i=1, \dots, n} X_i$ peut être renormalisé afin d'obtenir un résultat de convergence en loi. Plus précisément, si l'on choisit $a_n = \sqrt{2 \log n}$ et $b_n = a_n - \frac{\log \log n + \log(4\pi)}{2a_n}$ (avec $n \geq 1$) alors

$$a_n(M_n - b_n) \rightarrow \mathcal{G} \quad \text{en loi} \quad \text{lorsque} \quad n \rightarrow \infty.$$

Ici, $\mathcal{G}$ désigne la loi de Gumbel laquelle admet pour fonction de répartition

$$\mathbb{P}(\mathcal{G} \leq t) = e^{-e^{-t}} \quad \text{avec} \quad t \in \mathbb{R}.$$

Il serait donc également intéressant d'obtenir une inégalité de concentration non-asymptotique reflétant cette convergence en loi (et donc l'ordre correct de la variance). Autrement dit, nous aimerions obtenir des inégalités de la forme

$$\mathbb{P}(a_n(M_n - b_n) \geq t) \leq \theta_1(t) \quad \text{et} \quad \mathbb{P}(a_n(M_n - b_n) \leq -t) \leq \theta_2(t) \quad \text{pour tout} \quad t \geq 0,$$

telles que les asymptotiques des fonctions $\theta_i : \mathbb{R}_+ \rightarrow \mathbb{R}$ (pour $i = 1, 2$) reflètent les queues de distribution de la loi de Gumbel. C'est-à-dire,

$$\theta_1(t) \simeq e^{-t} \quad \text{et} \quad \theta_2(t) \simeq e^{-e^t} \quad \text{lorsque} \quad t \rightarrow +\infty.$$

Il est naturel mais naïf de voir ce qu'il serait possible d'obtenir grâce à l'inégalité de concentration (11.1.4). Celle-ci entraîne que, pour tout $t \geq 0$,

$$\mathbb{P}(a_n |M_n - b_n| \geq t) \leq 2e^{-t^2/2a_n^2}.$$

Cette inégalité se comporte très mal vis-à-vis de la dépendance en n et devient inutile lorsque $n \rightarrow \infty$. De plus, elle ne reflète pas les asymptotiques de la loi de Gumbel. En revanche, elle reste pertinente pour les grandes déviations de la forme $t = u\sqrt{a_n}$ avec $u \geq 1$.

Cet exemple introduit la deuxième problématique à laquelle nous allons proposer des éléments de réponse durant les prochains chapitres :

Supposons que nous ayons réussi à obtenir le bon ordre de grandeur de la variance et que M_n satisfasse une convergence en loi (après une éventuelle renormalisation). Est-il possible d'obtenir une inégalité de concentration reflétant à la fois le bon ordre de grandeur de la variance et les asymptotiques des queues de distribution de la loi limite ?

Comme nous allons le voir, ceci n'est pas simple et il ne semble pas être possible de trouver une méthode générale permettant de traiter de manière optimale un modèle gaussien quelconque. Le chapitre suivant est dédié à la présentation d'un certain nombre de modèles gaussiens vérifiant un phénomène de superconcentration. Nous insistons sur le fait que chaque modèle présenté possède

des structures de corrélations fondamentalement différentes les unes des autres et que des méthodes *ad hoc* sont utilisées pour chacun d'entre eux. Ceci amène le problème ouvert suivant, déjà présent dans [61] :

Est-il possible de trouver des conditions sur la structure de covariance qui seraient équivalentes au phénomène de superconcentration ?

Après avoir effectué un rapide survol de la littérature et des différents modèles (gaussiens, pour la plupart) qui présentent ce type de phénomène, nous allons présenter plusieurs inégalités, portant sur la variance, qui améliorent l'inégalité de Poincaré (11.1.1). Nous illustrerons ces améliorations sur plusieurs exemples et expliqueront également les limitations de leur champ d'application.

Dans un ultime chapitre, nous développerons les arguments permettant d'étendre les améliorations obtenues au niveau de la variance en inégalité de superconcentration.

Chapitre 12

Modèles

Dans cette section, nous allons présenter différents modèles présentant de la superconcentration. Nous esquisserons, pour chaque thématique, les outils utilisés pour obtenir de la superconcentration. Il sera alors évident pour le lecteur que chaque exemple nécessite des outils mathématiques spécifiques, inhérents au modèle et parfois très techniques, et qu'une méthodologie générique ne semble pas envisageable. Le panorama de la littérature existante à ce sujet est vaste, nous n'abordons d'un nombre restreint d'exemples dans ce qui va suivre. Plus précisément, nous allons nous concentrer sur les domaines suivants :

- théorie des extrêmes,
- matrices aléatoires,
- marche aléatoire branchante,
- percolation,
- verres de spin.

Pour compléter cette présentation, nous renvoyons le lecteur vers [61, 199].

12.1 Théorie des extrêmes

Dans cette section, nous présenterons des modèles provenant la théorie des extrêmes. Nous nous focaliserons principalement sur la convergence en loi du maximum dans le cadre suivant : étant donné une famille $(X_i)_{i=1,\dots,n}$ de variables aléatoires, est-il possible de trouver des constantes a_n et b_n telles que

$$a_n(M_n - b_n) \quad \text{avec} \quad M_n = \max_{i=1,\dots,n} X_i$$

converge en loi vers une limite non-triviale ? Pour plus de détails sur la théorie des extrêmes, nous renvoyons vers les ouvrages de référence suivants : [134, 81, 94].

12.1.1 Famille de variables aléatoires indépendantes et de même loi

Débutons par le cas le plus simple, celui où toutes les variables de la famille $(X_i)_{i=1,\dots,n}$ sont supposées indépendantes et de même loi. Dans un tel cadre, les lois limites possibles (vers lesquelles le maximum renormalisé converge) sont complètement caractérisées (cf. [134, 81, 94]) et sont restreintes au nombre de trois. Nous présentons, ci-dessous, deux d'entre-elles via leurs fonctions de répartitions.

- La loi de Gumbel :

$$\mathcal{G}(x) = \exp(-e^{-x}) \quad \text{pour } x \in \mathbb{R}.$$

- La loi de Weibull de paramètre $\alpha > 0$:

$$\mathcal{W}_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{lorsque } x < 0, \\ 1 & \text{sinon.} \end{cases}$$

Remarque. La loi manquante est celle de Fréchet, cette dernière apparaît naturellement lorsque les variables $(X_i)_{i=1,\dots,n}$ sont des lois à queues lourdes.

Des conditions nécessaires et suffisantes existent pour déterminer si la famille $(X_i)_{i=1,\dots,n}$ appartient au domaine d'attraction d'une de ces trois lois limites. Par exemple, si F_μ désigne la fonction de répartition de la loi commune (μ) , ces conditions peuvent s'exprimer en terme de la queue de distribution de $1 - F_\mu$ (cf. [134, 81, 94]) et les constantes de renormalisation a_n et b_n peuvent être déterminées à partir de cette même fonction.

Ci-dessous, nous proposons deux résultats classiques de convergence : le premier concernant la loi uniforme sur $[0, 1]$ et le deuxième, la mesure gaussienne standard.

Théorème 12.1.1. *La loi uniforme appartient au domaine d'attraction de la loi de Weibull. Plus précisément, si X_1 suit une loi uniforme sur $[0, 1]$, alors*

$$a_n(M_n - 1) \rightarrow \mathcal{W}_1$$

lorsque $n \rightarrow \infty$ avec $a_n = n$.

Plus tôt, nous avons déjà évoqué le cas de la mesure gaussienne standard.

Théorème 12.1.2. *La mesure gaussienne standard γ_n appartient au domaine d'attraction de la loi de Gumbel. Autrement dit, si X_1 suit une loi normale standard, alors*

$$a_n(M_n - b_n) \rightarrow \mathcal{G} \tag{12.1.1}$$

lorsque $n \rightarrow \infty$ avec $a_n = \sqrt{2 \log n}$ et $b_n = a_n - \frac{\log(4\pi) + \log \log n}{2a_n}$.

Comme nous l'avons vu dans le cas gaussien, les outils classiques de la concentration de la mesure ne sont pas efficaces pour obtenir des bornes pertinentes sur la variance de M_n ou sur les queues de déviation (à droite et à gauche) du maximum renormalisé. Ce constat est également valable pour la loi uniforme (bien qu'il soit possible de calculer directement la variance du maximum dans ce cas de figure).

12.1.2 Famille de variables aléatoires gaussiennes

La théorie des extrêmes ne se limite pas uniquement au cas de variables indépendantes de mêmes lois. En effet, il existe de nombreux résultats concernant les suites stationnaires (cf. [134, 81, 94]). Il se trouve que le cas gaussien est plus facile à traiter grâce à des outils de comparaison (propres aux variables gaussiennes) comme l'inégalité de Slepian (cf. [134, 143]) et ses variantes. Les hypothèses assurant la convergence en loi des extrêmes, pour ce cas particulier d'aléa, s'expriment facilement à l'aide de la fonction de covariance. Ci-dessous, nous présentons un résultat dans cet esprit.

Soit $(X_i)_{i \geq 0}$ une suite gaussienne stationnaire centrée telle que $\mathbb{E}[X_i^2] = 1$, pour tout $i \geq 0$, avec pour fonction de covariance

$$\text{Cov}(X_i, X_j) = \phi(|i - j|) \quad \text{avec } i, j \geq 0,$$

où $\phi : \mathbb{N} \rightarrow \mathbb{R}$. De nombreux résultats ont été obtenus dans ce cadre afin d'assurer la convergence du maximum $M_n = \max_{i=1, \dots, n} X_i$ (cf. e.g. [134, 81, 94]). Ces résultats s'obtiennent en étudiant le comportement asymptotique la fonction de covariance ϕ .

Théorème 12.1.3 (Berman). *Soit $(X_i)_{i \geq 0}$ une suite gaussienne stationnaire avec, pour fonction de covariance ϕ telle que $\phi(n) \log n \rightarrow 0$ lorsque $n \rightarrow \infty$. Alors*

$$a_n(M_n - b_n) \xrightarrow[n \rightarrow \infty]{} \mathcal{G} \quad \text{en loi}$$

où $a_n = (2 \log n)^{1/2}$ et

$$b_n = (2 \log n)^{1/2} - \frac{1}{2}(2 \log n)^{-1/2}(\log \log n + \log 4\pi).$$

Remarque. Il existe d'autres théorèmes expliquant ce qui se produit lorsque $\phi(n) \log n = O(1)$ ou $\phi(n) \log n = +\infty$; dans ces cas de figures, la limite n'est plus la loi de Gumbel. Il est également possible d'étendre le résultat précédent à des processus gaussiens stationnaires mais nous n'aborderons pas cet aspect dans ce cours. Pour tout ceci, nous renvoyons le lecteur vers [134, 81, 94].

Tout comme le cas des variables i.i.d., la théorie classique de la concentration s'avère inefficace pour produire des bornes pertinentes. Dans un chapitre ultérieur, nous verrons de quelle manière des arguments hypercontractifs permettent d'atteindre les ordres de grandeur optimaux.

12.2 Matrices aléatoires

Dans cette section nous allons brièvement décrire des résultats concernant les matrices aléatoires. Nous porterons une attention toute particulière à l'ensemble de matrices gaussiennes du GUE (pour *Gaussian Unitary Ensemble*). La plupart du temps, les résultats présentés sont valables pour des modèles plus généraux que nous omettrons et renvoyons le lecteur vers les ouvrages [6, 15] pour plus de détails.

L'un des résultats majeurs de cette théorie concerne l'étude de la plus grande valeur propre $\lambda_{\max}$ d'une matrice aléatoire. Sous certaines conditions de régularité, cette valeur propre converge presque sûrement vers une valeur déterministe (cf. [6, 141]). Un résultat de fluctuations pour $\lambda_{\max}$

(sorte d'analogie du théorème de la limite centrale pour des matrices aléatoires) est, par la suite, obtenu pour différents modèles de matrices aléatoires. Nous allons présenter ce résultat pour des matrices provenant de l'ensemble GUE.

Considérons, pour tout entier $n \geq 1$, la matrice (de taille $n \times n$) auto-adjointe gaussienne centrée $X = X^n = (X_{ij}^n)_{1 \leq i, j \leq n}$, de variance σ^2 . Ceci revient à considérer une matrice hermitienne X de taille $n \times n$ dont les entrées, au-dessus de la diagonale, sont des variables aléatoires gaussiennes complexes (respectivement réelles sur la diagonale) centrées et de variance σ^2 (autrement dit, la partie réelle et imaginaire sont des variables aléatoires gaussiennes centrées de variance $\sigma^2/2$). De manière équivalente, la matrice aléatoire X possède la densité de probabilité suivante

$$\mathbb{P}(dX) = \frac{1}{Z} \exp(-\operatorname{Tr}(X^2)/2\sigma^2) dX$$

sur l'espace des matrices hermitiennes $\mathcal{H}_n \cong \mathbb{R}^{n^2}$ de taille $n \times n$, avec

$$dX = \prod_{1 \leq i \leq n} dX_{ii} \prod_{1 \leq i < j \leq n} d\operatorname{Re}(X_{ij}) d\operatorname{Im}(X_{ij})$$

la mesure de Lebesgue sur $\mathcal{H}_n$ et $Z = Z_n$ est une constante de renormalisation. Nous désignerons par $\lambda_{\max}$ la plus grande valeur propre d'une matrice du GUE.

Théorème 12.2.1 (Tracy-Widom). *Soit $X = X^n$ une matrice du GUE avec $\sigma^2 = \frac{1}{4n}$, alors la suite $(\lambda_{\max})_{n \in \mathbb{N}}$ converge en loi vers une distribution indépendante du modèle. Plus précisément,*

$$n^{2/3}(\lambda_{\max} - 1) \rightarrow \mathcal{TW} \quad \text{lorsque } n \rightarrow \infty$$

où $\mathcal{TW}$ désigne la loi de Tracy-Widom.

Remarque. La fonction de répartition $F_{\mathcal{TW}}$ de cette loi peut s'exprimer de plusieurs manières explicites (cf. [141]). L'une d'entre-elles fait intervenir un déterminant de Fredholm associé au noyau d'Airy, tandis qu'une autre montre qu'il est possible d'exprimer la fonction de répartition de la manière suivante :

$$F_{\mathcal{TW}}(s) = \exp\left(-\int_{2s}^{\infty} (x - 2s)u(x)^2 dx\right) \quad \text{avec } s \in \mathbb{R}$$

où $u(x)$ désigne la solution de l'équation différentielle de Painlevé II

$$u'' = 2u^3 + xu$$

sous la contrainte asymptotique $u(x) \sim \frac{1}{2\sqrt{\pi}x^{1/4}}e^{-\frac{2}{3}x^{3/2}}$ lorsque $x \rightarrow \infty$. Cette formulation permet notamment de déterminer les comportements asymptotiques de $F_{\mathcal{TW}}$. En effet, il est possible d'obtenir les encadrements suivants : pour tout $s \geq 0$, nous avons

$$C^{-1}e^{-Cs^3} \leq F_{\mathcal{TW}}(-s) \leq Ce^{-s^3/C}$$

et

$$C^{-1}e^{-Cs^{3/2}} \leq 1 - F_{\mathcal{TW}}(s) \leq Ce^{-s^{3/2}/C}$$

pour des constantes universelles $C > 0$.

Afin d'employer les outils classiques de la concentration de la mesure, il est possible d'exprimer $\lambda_{\max}$ sous la forme d'un supremum. C'est-à-dire,

$$\lambda_{\max} = \sup_{|u|=1} uX\bar{u}.$$

Sous cette forme, il devient alors facile de voir que $\lambda_{\max}$ est une application 1-lipschitzienne des n^2 variables aléatoires gaussiennes indépendantes. Par suite, la concentration classique nous assure que

$$\mathbb{P}\left(\left|\lambda_{\max} - \mathbb{E}[\lambda_{\max}]\right| \geq t\right) \leq 2e^{-2nt^2} \quad \text{avec } t \geq 0.$$

Bien que cette inégalité de concentration soit pertinente vis-à-vis du comportement gaussien des grandes déviations, elle ne reflète nullement le cas des petites déviations ; cette inégalité montre également que

$$\text{Var}(\lambda_{\max}) \leq \frac{C}{n}$$

au lieu de l'ordre de grandeur attendu en $\frac{C}{n^{4/3}}$.

De nombreuses méthodes ont été mises en place pour étudier ce modèle, la plus satisfaisante (de part sa généralité) d'entre-elles est due à Ledoux et Rider. En effet, ces derniers proposent une approche unifiée permettant de traiter un grand nombre de modèles matriciels (les β -ensembles, les matrices de Wishart, cf. [142]). Voici l'énoncé, dans le cadre du GUE, de l'un des résultats majeurs de cet article.

Théorème 12.2.2 (Ledoux-Rider). *Dans le cadre de la section 12.2, il est possible d'exhiber des constantes numériques $C > 0$ telles que*

$$\mathbb{P}(\lambda_{\max} \geq 1+t) \leq Ce^{-n \max(t^2, t^{3/2})/C} \quad \text{pour tout } t \geq 0$$

et

$$\mathbb{P}(\lambda_{\max} \leq 1-t) \leq Ce^{-n^2 t^3 / C} \quad \text{pour tout } t \geq 0.$$

Remarque. Ils ont également montré qu'il s'agissait du bon ordre de grandeur en proposant une minoration du même ordre. Ces inégalités de concentrations reflètent le comportement asymptotique de la loi de Tracy-Widom et, une fois intégrées, permettent de retrouver le bon ordre du grandeur de la variance de $\lambda_{\max}$.

La démonstration repose sur une formulation variationnelle de la loi de Tracy-Widom combinée à l'introduction de matrice tridiagonale. Les arguments étant assez techniques, nous ne les présenterons pas ici et renvoyons le lecteur vers l'article [142].

Des résultats plus généraux ont été obtenus depuis et seront mentionnés à la fin de ce chapitre.

12.3 Marche aléatoire branchante

Pour présenter ce nouveau modèle, nous suivrons partiellement l'exposition faite dans le cours de Zeitouni [216]. Pour des raisons pédagogiques, nous présenterons uniquement le modèle de marche aléatoire branchante (« branching random walk »). Celui-ci fait partie des modèles gaussiens dit « log-corrélés » les plus simples à exposer et sert de version simplifiée du champ libre gaussien discret sur $\mathbb{Z}^2$ ou DGFF (pour « Discrete Gaussian Free Field ») en abrégé.

Fixons tout d'abord les notations. Soit $\mathcal{T}$ un arbre binaire ayant pour racine un sommet o . Nous noterons par V l'ensemble des sommets et par E l'ensemble des arêtes de cet arbre. Nous désignerons par $|v|$ la distance d'un sommet à la racine. Autrement dit, il s'agit de la longueur du chemin géodésique reliant v à o et nous désignerons par $o \leftrightarrow v$ l'ensemble des sommets composant cette géodésique (par abus de notation, nous noterons également par $o \leftrightarrow v$ l'ensemble des arêtes présentes sur la géodésique reliant v à o). Similairement, pour $v, w \in V$, nous noterons par $|v \leftrightarrow w|$ la longueur de l'unique géodésique reliant v à w . Pour $n \geq 1$, la n -ième génération de l'arbre correspond à l'ensemble $\{v \in V; |v| = n\}$ que nous noterons D_n .

Posons ensuite $S_\pi = \sum_{e \in \pi} X_e$ pour tout $\pi \in D_n$. La marche aléatoire branchante sur $\mathcal{T}$ correspond à la famille $(S_\pi)_{\pi \in D_n}$ et nous allons nous focaliser sur

$$M_n = \max_{\pi \in D_n} S_\pi \quad \text{pour tout } n \geq 1.$$

Remarque. Comme présenté dans le cours de Zeitouni, mentionnons rapidement qu'il est possible de considérer ce modèle dans un contexte plus général en imposant des conditions sur les poids $(X_e)_{e \in E}$ et la forme de l'arbre. Pour plus de détails, nous renvoyons le lecteur vers [216].

Ce modèle de marche aléatoire branchante est étudié à l'aide de méthodes dites de « second moment » (liées à l'inégalité de Paley et Zygmund) et de résultats du type « Ballot's theorem ». Suivant ces techniques, il est possible de montrer que

Théorème 12.3.1. *Sous les hypothèses précédentes, l'estimation suivante est vérifiée*

$$\mathbb{E}[M_n] = C_1 n - C_2 \log n + O(1) \quad \text{pour tout } n \geq 1,$$

avec $C_1, C_2 > 0$ des constantes numériques dépendant uniquement de la loi μ .

Remarque. Une étape importante pour établir ce genre de résultat est de déterminer une barrière au-delà de laquelle les configurations extrémales $(S_v)_{v \in D_n}$ ne peuvent aller. Il sera aussi important de suivre l'évolution des incréments $(S_v^k)_{k=1, \dots, n}$, composant la « trajectoire » de S_v (ces incréments sont définis par $S_v = \sum_{k=1}^n S_v^k$), lesquels ne peuvent également pas dépasser un certain seuil.

Le résultat précédent, via l'argument de Dekking-Host [216], permet de montrer que la suite $(M_n - \mathbb{E}[M_n])_{n \geq 1}$ est tendue. Il a été montré depuis que cette même suite converge en loi vers une loi de probabilité (cf. [51, 50]).

Bien que la preuve explicite semble être absente, à notre connaissance, de la littérature, le fait que $\text{Var}(M_n) = O(1)$ fait dorénavant partie du folklore. Il est intéressant de mentionner que les arguments utilisés pour démontrer le théorème ci-dessus, bien qu'élémentaires, reposent sur des

calculs assez techniques et précis. Une fois de plus, l'inégalité de Poincaré (11.1.1) fournit une majoration sous-optimale de la variance :

$$i.e. \quad \text{Var}(M_n) \leq n$$

ne reflétant pas la réalité. Nous reviendrons plus tard sur ce modèle en expliquant de quelle manière des arguments hypercontractifs permettent d'améliorer, sans pour autant atteindre un résultat optimal, la borne précédente au prix de calculs relativement élémentaires.

Comme mentionné plus tôt, la marche aléatoire branchante est une version plus simple du champ libre gaussien discret. Ce modèle a donné naissance à de nombreux travaux de recherche qui représentent une littérature conséquente. Comme nous le présenterons rapidement dans la dernière section de ce chapitre, un grand nombre de résultats (impliquant des calculs délicats et techniques) sont disponibles à ce sujet.

12.4 Percolation

Les modèles de percolation sont parmi les plus célèbres exemples présentant de la superconcentration. Cette célébrité est notamment due au fait que certains modèles de percolation dirigée ont de profonds liens avec la théorie des matrices aléatoires (cf. [141]).

Présentons l'un de ces modèles. Considérons le réseau $\mathbb{Z}^d$ (avec $d \geq 2$) et notons par E l'ensemble des arêtes de $\mathbb{Z}^d$. Ensuite, associons à chacune de ces arêtes $e \in E$ un poids ω_e provenant d'une famille de variables aléatoires positives i.i.d. $(\omega_e)_{e \in \mathbb{Z}^d}$. Intuitivement, ces poids modélisent le temps nécessaire à un fluide pour traverser une arête.

Définissons à présent la notion de temps de passage. Soient $x, y \in \mathbb{Z}^d$ et π un chemin reliant ces deux sommets, le temps de passage $T_\pi(x, y)$ est alors défini par

$$T_\pi(x, y) = \sum_{e \in \pi} \omega_e.$$

Le premier temps de passage $T(x, y)$ d'un sommet x à un sommet y est alors défini comme le minimum (sur les chemins π) des temps de passage $T_\pi(x, y)$.

Soient $x \in \mathbb{R}^d$ et $n \in \mathbb{N}^*$, nous allons étudier

$$T_n(x) := T(0, [nx])$$

où 0 désigne l'origine de $\mathbb{Z}^d$ et $[nx]$ le point du réseau le plus proche de nx . Ce nouvel objet a été étudié par de nombreux auteurs, lesquels ont pu établir un certain nombre de résultats (cf. [12]). Par exemple,

$$\lim_{n \rightarrow \infty} \frac{T_n(x)}{n}$$

existe et est une fonction déterministe de x .

Considérons $x = e_1 = (1, 0, \dots, 0) \in \mathbb{Z}^d$ et abrégeons $T_n(x)$ en T_n . Grâce à des arguments de martingales, Kesten fut le premier à prouver que

$$\text{Var}(T_n) \leq Cn$$

avec $C > 0$ une constante numérique. Bien que cela n'atteigne pas encore l'ordre de fluctuations attendu par les physiciens ($n^{2/3}$), ce résultat fut renforcé par Benjamini, Kalai et Schramm (cf. [25]). Dans cet article, les auteurs supposent que les poids ω_e sont des variables aléatoires de Bernoulli et exhibent un phénomène de superconcentration. Chatterjee a étendu ce résultat lorsque les poids sont des fonctions bornées de variables aléatoires gaussiennes standard (cf. [61]), c'est le contenu du théorème suivant.

Théorème 12.4.1 (Chatterjee). *Si les poids des arêtes $(\omega_e)_{e \in E}$ peuvent s'exprimer comme la loi d'une fonction F lipschitzienne, uniformément bornée, de variables aléatoires gaussiennes :*

i.e. $\omega_e = F(X_e)$ avec $(X_e)_{e \in E}$ une famille de variables aléatoires gaussiennes standard i.i.d., alors, pour tout $n \geq 2$, nous avons

$$\text{Var}(T_n) \leq \frac{Cn}{\log n}$$

avec $C > 0$ est une constante numérique ne dépendant que de la loi des poids et de la dimension d .

Dans le chapitre prochain, nous proposerons une démonstration de ce théorème.

Benaïm et Rossignol (cf. [24]) furent les premiers à obtenir un résultat généralisant le théorème 12.4.1. En fait, ils furent les premiers à proposer une approche plus abstraite du problème afin de considérer des poids plus généraux. Pour ce faire, ils se restreignent à une classe de mesures de probabilités dite « nearly gamma ». Ceci signifiant que ces lois de probabilités ont des propriétés de concentration proches des lois gamma. Leur article apporte des arguments novateurs dans ce domaine :

- le premier argument provient de la théorie du transport optimal et se retrouve combiné avec une inégalité alternative à celle de Sobolev logarithmique ;
- le deuxième argument consiste à utiliser une inégalité de Poincaré (à l'échelle exponentielle) avec des constantes dépendant de n .

Pour être plus précis, ils considèrent leurs poids comme étant la mesure image de la mesure gaussienne γ_n par une application T , ensuite ils utilisent l'inégalité de Falik et Samorodnisky (cf. [97]) pour enfin appliquer l'inégalité de Sobolev logarithmique satisfaite par la mesure gaussienne γ_n . L'autre argument est résumé dans le lemme suivant.

Lemme 12.4.1. *Soit Z une variable aléatoire et $K > 0$. Supposons que pour tout $|\theta| \leq 2/\sqrt{K}$,*

$$\text{Var} \left(e^{\theta Z/2} \right) \leq K \frac{\theta^2}{4} \mathbb{E} \left[e^{\theta Z} \right].$$

Alors,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| > t) \leq 6e^{-ct/\sqrt{K}} \quad \text{pour tout } t \geq 0$$

avec $c > 0$ une constante numérique.

Ils utilisent ce résultat en faisant dépendre K de n . Grossièrement, ils suivent la méthodologie mise en place par Benjamini-Kalai et Schramm au niveau exponentiel pour ensuite appliquer le résultat précédent et montrer que K_n est de l'ordre de $\frac{n}{\log n}$. Ils obtiennent donc une inégalité de superconcentration (pour T_n), laquelle renforce alors (au niveau exponentiel) le théorème 12.4.1. Le principal défaut de cette approche est visible sur la condition « nearly-gamma ». Celle-ci est trop restrictive et exclue des lois usuelles de probabilités (la mesure gaussienne standard par exemple). Nous reviendrons sur l'utilisation du lemme 12.4.1 lorsque nous étudierons les suites gaussiennes stationnaires.

Les travaux de Benaïm et Rossignol ont été poursuivis par d'autres mathématiciens et ont permis de faire de nombreux progrès dans ce domaine des mathématiques. Nous dirons quelques mots à ce sujet à la fin du chapitre.

12.5 Verres de spin

La littérature concernant les verres de spin est très vaste, nous renvoyons le lecteur vers les ouvrages de référence (dont nous nous inspirons) suivants [195, 196, 158, 46] pour plus de détails. Nous choisissons de nous restreindre au modèle de Sherrington et Kirkpatrick (SK en abrégé) que nous présentons ci-dessous.

Débutons par un point de vue informel issu de [193]. Considérons une grande famille d'individus numérotés de 1 à n et supposons que chaque individu connaisse le reste de la population mise en jeu. Nous mesurons les sentiments d'un individu i envers un autre j par un nombre X_{ij} qui peut être négatif ou positif; par la suite X_{ij} sera remplacé par une variable aléatoire. Nous supposons également que $X_{ij} = X_{ji}$. Bien que cela soit irréaliste, nous supposons aussi que la famille $(X_{ij})_{i < j}$ est composée de variables aléatoires réelles indépendantes.

Ce modèle vérifie une propriété curieuse : si les individus i et j sont amis (i.e. $X_{ij} > 0$), tout comme les individus j et k (i.e. nous ne pouvons rien dire à priori du signe de X_{ik}). Autrement dit, les individus i et k peuvent être aussi bien amis qu'ennemis. Cette propriété est désignée sous l'appellation de « frustrations ». Finalement, les interactions $(X_{ij})_{i < j}$ décrivent une situation sociale très complexe.

Etant donné une réalisation typique de la famille $(X_{ij})_{i < j}$, nous nous proposons dès lors l'objectif suivant : séparer du mieux possible cette population en deux groupes de sorte que les amis restent ensemble et soient séparés de leurs ennemis. Ceci étant un moyen de soulager les tensions créées par les frustrations. Il apparaît rapidement que cette partition risque d'être effectuée de façon imparfaite : des amis seront séparés et des ennemis devront cohabiter. Pour mesurer de manière quantitative la qualité de notre partition, il est pratique d'assigner à chaque individu i un nombre $\sigma_i \in \{-1, 1\}$, et ainsi définir deux classes différentes. La qualité de la partition se mesure alors grâce à la quantité suivante :

$$\sum_{i < j} X_{ij} \sigma_i \sigma_j.$$

Ainsi, plus grande est cette quantité, meilleure est la partition.

Malgré la simplicité apparente de ce nouvel objet $\sum_{i < j} X_{ij} \sigma_i \sigma_j$, le problème d'optimisation revenant à déterminer le maximum de cette fonction pour une réalisation typique des X_{ij} est extrêmement difficile. De façon équivalente nous pouvons aussi nous intéresser à la fonction

$$-\sum_{i < j} X_{ij} \sigma_i \sigma_j.$$

Déterminer le minimum d'une telle fonction de configurations est appelé en physique *un problème à température zéro*. Cette appellation s'explique par le fait suivant : lorsque la température est nulle, un système physique se trouve toujours dans sa configuration d'énergie minimale (l'*état fondamental* du système).

D'un point de vue physique, les verres de spin modélisent le phénomène suivant : considérons des matériaux non magnétiques (comme le cuivre) dans lequel nous introduisons aléatoirement des impuretés magnétiques (du manganèse par exemple). Nous retrouvons alors les deux propriétés essentielles décrites plus haut :

- le désordre dû à la répartition aléatoire des impuretés ;
- les frustrations engendrées par les interactions des impuretés entre elles.

De manière plus formelle, notre problème revient alors à considérer n particules, chacune d'entre-elles portant un spin valant $+1$ ou -1 . Une configuration de spin $\sigma = (\sigma_1, \dots, \sigma_n)$ correspond alors à un élément de $C_n = \{-1; 1\}^n$. Considérons ensuite $(X_{ij})_{1 \leq i < j \leq n}$ une collection de variables aléatoires gaussiennes standards indépendantes ; cette famille modélise le désordre de notre système physique.

Etant donné un tel désordre, nous pouvons définir l'énergie d'une configuration de spin σ par l'hamiltonien :

$$H_n(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1 \leq i < j \leq n} X_{ij} \sigma_i \sigma_j.$$

Elle permet aussi de définir une mesure de probabilité (dite de Gibbs) sur C_n en associant à chaque configuration σ le poids suivant : $\exp(-\beta H_n(\sigma))$ où $\beta > 0$ représente l'inverse de la température. Le poids que nous associons à chaque configuration se justifie physiquement par le fait qu'il permet de favoriser les configurations ayant une faible énergie.

Connaissant le désordre, la probabilité d'observer une configuration de spin $\sigma \in S_n$ est donc

$$\frac{\exp(-\beta H_n(\sigma))}{Z_{n,\beta}}$$

où $Z_{n,\beta} = \sum_{\sigma \in C_n} e^{-\beta H_n(\sigma)}$ est une constante de renormalisation (permettant d'obtenir une mesure de probabilité). Dans le monde des verres de spins, cette constante $Z_{n,\beta}$ est appelée *fonction de partition du système*.

En résumé, nous avons donc à notre disposition un champ gaussien centré $(H_n(\sigma))_{\sigma \in C_n}$ indexé par le cube discret. Celui-ci définit le modèle des verres de spin de Sherrington et Kirkpatrick SK.

Cette fois-ci, nous allons étudier la fonction $F_{n,\beta}$ définie par

$$F_{n,\beta} = \frac{1}{\beta} \log(Z_{n,\beta}) \quad \text{avec } \beta > 0.$$

Cette fonction porte le nom *d'énergie libre* du système, il s'agit d'une version régulière de la fonction minimum. En effet, il n'est pas difficile de vérifier que

$$\lim_{\beta \rightarrow +\infty} F_{n,\beta} = \min_{\sigma \in C_n} H_n(\sigma).$$

Remarque. Il existe de nombreux modèles de verres de spins (REM, GREM, ...) que nous ne pouvons pas présenter dans ce cours. Nous renvoyons le lecteur vers les références suivantes [195, 196, 46].

À partir d'une formule de représentation (que nous présenterons dans le chapitre suivant) de la variance le long d'un semi-groupe, Chatterjee a obtenu dans [61] une série d'inégalités différentielles. Celles-ci lui ont permis d'obtenir une majoration de la variance de l'énergie libre pour le modèle SK (ainsi que pour d'autres modèles plus compliqués, dit de p -spin). Mentionnons au passage l'acte de conférence [202] qui propose une relecture des travaux de [61] et permet de clarifier certains arguments proposés par Chatterjee. Son théorème dit essentiellement la chose suivante

Théorème 12.5.1 (Chatterjee). *Dans le modèle SK, la majoration suivante est vérifiée pour l'énergie libre, pour tout $n \geq 2$,*

$$\text{Var}(F_{n,\beta}) \leq \frac{C(\beta)n}{\log n} \quad \text{pour tout } \beta \geq 0.$$

avec $C(\beta)$ une constante numérique dépendant uniquement du paramètre β .

Remarque. Cependant cette borne n'est pas optimale, d'après [61] il est attendu que $\text{Var}(F_{n,\beta}) = O(1)$ pour toute température $\beta > 0$. Soulignons tout de même que les travaux de Chatterjee sont les premiers à obtenir une borne sous-linéaire de la variance de l'énergie libre (améliorant ainsi la majoration fournie par l'inégalité de Poincaré). Mentionnons également que sa méthodologie fournit aussi des bornes pertinentes pour des modèles plus simples de verres de spin, comme le REM (« Random Energy Model »).

Les différents modèles que nous venons de présenter dans ces dernières sections permettent de saisir la variété et la complexité du phénomène de superconcentration. Ainsi, il semble très difficile d'obtenir une méthode générale permettant d'améliorer la théorie classique de la concentration pour chacun d'entre eux. Nous verrons cependant qu'il est possible de développer des méthodes assez souples et plus simples d'emploi (que celles mentionnées dans ce chapitre) et qui permettront de renforcer les résultats connus. Le prochain chapitre a pour objectif d'expliquer ce qu'il est possible de faire au niveau de la variance ; dans certains cas, nous verrons que ces méthodes sont optimales.

12.6 Références bibliographiques

Les mots qui vont suivre sont principalement issus de la thèse [199].

Matrices aléatoires et gaz de Coulomb

Historiquement la théorie des matrices aléatoires est née durant les années 20 sous l'impulsion des mathématiciens Wigner et Wishart. Dès lors, ce domaine des mathématiques a connu un développement intense.

Le théorème de Tracy-Widom reste vrai si lorsque des ensembles de matrices plus généraux sont considérés, il s'agit des matrices dites de Wigner, en imposant des conditions sur les moments de ces dernières. À ce sujet citons par exemples les travaux de Péché-Soshnikov [164], Ruzmaikina [173], Khorunzhiy [129] ou encore Lee-Yin [144].

Par souci de concision, nous n'avons pas abordé le cas des matrices de covariance empirique. Nous renvoyons le lecteur vers les références suivantes [121, 119]. Tout comme pour les matrices du GUE, des résultats d'universalité sont obtenus par Soshnikov [179] et Péché [163].

Les premières tentatives d'étude non-asymptotique d'inégalités de déviation pour la plus grande des valeurs propres du GUE sont obtenues par Aubrun dans [10], ce résultat peut être également obtenu à partir des travaux de Johansson [120], comme annoncé dans [141]. L'inégalité obtenue s'exprime de la manière suivante, pour tout $t \geq 0$

$$\mathbb{P}(\lambda_{\max} \geq 1 + t) \leq C e^{-n \max(t^2, t^{3/2})/C}.$$

Même si nous pouvons citer un article de Tao et Vu [205] qui fournit une inégalité de déviation, reflétant le taux de convergence en $n^{-2/3}$, pour la plus grande valeur propre ; celle-ci ne reflète pas les asymptotiques de la fonction de répartition de la loi de Tracy-Widom et n'est pas valable pour les petites déviations. Il faudra attendre les travaux de Ledoux et Rider [142] qui obtiennent, par des démonstrations unifiées, pour les β -ensembles (qui contiennent les ensembles GUE et GOE), des inégalités de concentration optimales. En effet, ils retrouvent le résultat de Johansson ainsi qu'une inégalité de déviation à gauche de la forme

$$\mathbb{P}(\lambda_{\max} \leq 1 - t) \leq C e^{-n^2 t^3 / C} \quad \text{avec } t \geq 0.$$

En intégrant ces inégalités de déviation, ces résultats permettent d'obtenir le bon ordre de grandeur de la variance de la plus grande valeur propre d'une matrice du GUE et GOE.

Remarque. En fait, leur article fournit également les mêmes inégalités de déviation optimales (et donc sur la variance) à droite et à gauche pour la plus grande valeur propre d'une matrice provenant d'un ensemble de Laguerre.

Il faudra attendre les travaux de Dallaporta pour obtenir des résultats sur la variance du $\lambda_{\max}$ pour des ensembles de matrices plus généraux. Dans [78, 77], elle obtient notamment des majorations, de la variance de $\lambda_{\max}$, du bon ordre de grandeur pour des matrices de Wigner ou de covariances satisfaisant certaines hypothèses. À notre connaissance, il n'existe que très peu de travaux proposant des inégalités de déviations non asymptotiques pour des matrices de Wigner

ou de Wishart. Citons toutefois l'article de Feldheim et Sodin [99] dans lequel ils obtiennent des inégalités de déviation (à droite) optimales pour des petites déviations.

Le lecteur pourra obtenir d'autres informations concernant des résultats non-asymptotiques, s'inscrivant dans le cadre du phénomène de superconcentration, connus à ce jour dans la théorie des matrices aléatoires, en s'orientant vers la thèse de Dallaporta [78] ou les articles de survols [141, 210].

D'autres résultats provenant de la théorie des matrices aléatoires fournissent des exemples où la théorie de la concentration est sous-optimale. Citons par exemple l'article de Chafaï et Piché [60] ainsi que celui de Rider [169] qui s'intéresse à un gaz de particules chargées $\{z_1, \dots, z_n\}$ dans le plan complexe, où chaque particule est confinée par un champ externe Q ainsi qu'une interaction répulsive (avec les autres particules) coulombienne. La densité de probabilité d'un tel objet est proportionnelle à

$$(z_1, \dots, z_n) \mapsto \prod_{j=1}^n e^{-nQ(z_j)} \prod_{1 \leq j < k \leq n} |z_j - z_k|^\beta,$$

avec $\beta > 0$ un paramètre fixé et Q une fonction régulière donnée. Dans la suite, nous supposons que Q est radialement symétrique, au sens où

$$Q(z) = V(|z|),$$

avec $V : \mathbb{R}_+ \rightarrow \mathbb{R}$ est un potentiel régulier tel que $t \mapsto tV'(t)$ est croissante ou V est convexe. Les auteurs des articles mentionnés ci-dessus s'intéressent aux statistiques d'ordre associées à ces particules $z_1, \dots, z_n$, que l'on désignera par $|z_{(1)}| \geq \dots \geq |z_{(n)}|$, celles-ci ont une signification particulière vis à vis des matrices aléatoires (voir [60]).

Lorsque $\beta = 2$ (et $V(r) = r^2$), Rider a obtenu un théorème de représentation de la loi de ces statistiques d'ordre :

$$(|z_{(1)}|, \dots, |z_{(n)}|) = (R_{(1)}, \dots, R_{(n)}) \quad \text{en loi}$$

où les $R_{(1)} \geq \dots \geq R_{(n)}$ sont les statistiques d'ordre de variables indépendantes $R_1, \dots, R_n$ avec R_k de densité proportionnelle à

$$t \mapsto t^{2k-1} e^{-nV(t)} 1_{[0, \infty)}(t).$$

Remarque. Notons que ce choix particulier de β et du potentiel V pour le gaz de Coulomb correspond au modèle de Ginibre complexe en matrice aléatoire.

Rider obtient alors un théorème de convergence pour la particule ayant le plus grand module, ce résultat est ensuite prolongé par Chafaï et Piché pour des potentiels $V : \mathbb{R}_+ \rightarrow \mathbb{R}$ plus généraux. Plus précisément,

Théorème 12.6.1. *Avec les notations précédentes (avec $\beta = 2$), soit $|z_{(1)}| = \max_{k=1, \dots, n} |z_k|$. Supposons que $V(t) = t^\alpha$ pour tout $t \geq 0$ et $\alpha \geq 1$. Posons $c_n = \log n - 2 \log \log n - \log 2\pi$ et*

$$a_n = 2 \left(\frac{\alpha}{2} \right)^{1/\alpha + 1/2} \sqrt{nc_n} \quad \text{et} \quad b_n = \left(\frac{2}{\alpha} \right)^{1/\alpha} \left(1 + \frac{1}{2} \sqrt{\frac{2c_n}{\alpha n}} \right).$$

Alors, $(a_n(|z_{(1)}| - b_n))_{n \geq 1}$ converge en loi, lorsque $n \rightarrow \infty$, vers la loi de Gumbel, i.e., pour tout $t \in \mathbb{R}$

$$\mathbb{P}(a_n(|z_{(1)}| - b_n) \leq t) \rightarrow e^{-e^{-t}}.$$

Remarque. Ce théorème reste valable en supposant certaines conditions de stricte convexité sur $t \mapsto V(t)$ (voir [60]).

Marche aléatoire branchante et champ libre gaussien discret

La marche aléatoire branchante est une version plus simple du champ libre gaussien discret. Nous définirons ce nouveau modèle ci-dessous, puis nous tenterons de résumer l'évolution des progrès obtenus durant son étude. Enfin, nous mentionnerons brièvement les articles de recherche, s'inspirant des techniques utilisées pour le champ libre gaussien discret, portant sur des modèles gaussiens logarithmiquement corrélés plus généraux.

Considérons $\mathcal{T} = (V, E)$ un graphe (fini) non-orienté et connexe. Choisissons un sommet particulier $v \in V$ et désignons-le comme « la racine » du graphe, le champ libre gaussien discret correspond alors à la collection de variables aléatoires gaussiennes $(X_v)_{v \in V}$ avec $X_o = 0$ et la densité de probabilité, par rapport à la mesure de Lebesgue, des variables aléatoires restantes est donnée par

$$f((x_v)_{v \in V, v \neq o}) = \frac{1}{Z_{\mathcal{T}}} e^{-\frac{1}{2} \sum_{(v,w) \in V} (x_v - x_w)^2}, \quad (12.6.1)$$

où $Z_{\mathcal{T}}$ est une constante de renormalisation.

Remarque. Il est important de faire quelques commentaires après cette définition. Tout d'abord, si $\mathcal{T}$ est un arbre, le champ libre gaussien discret consiste à assigner à chaque arête $e \in E$ une variable aléatoire gaussienne standard Y_e et de poser $X_v = \sum_{e \in o \leftrightarrow v} Y_e$, $v \in V$. En particulier, si $\mathcal{T}$ correspond à l'arbre binaire, alors les valeurs de champ libre gaussien discret (noté $(X_v^T)_{v \in D_n}$) suivent le même comportement que M_n de la marche aléatoire branchante. Il est donc possible d'obtenir le résultat suivant

$$\mathbb{E}[\max_{v \in D_n} X_v^T] = \sqrt{2 \log 2n} - \frac{3}{2\sqrt{2 \log 2}} \log n + O(1)$$

Ensuite, le champ libre gaussien possède des liens avec certaines marches aléatoires. En effet, considérons la matrice de transition suivante

$$L_{v,w} = \begin{cases} 1, & (v,w) \in E, \\ -d_v, & v = w \neq o, \\ 0, & \text{sinon,} \end{cases}$$

où d_v désigne le degré du sommet v . A partir de (12.6.1), nous remarquons alors que la structure de corrélation du champ libre gaussien discret $(X_v)_{v \in V, v \neq o}$ est obtenue grâce à la matrice $(\tilde{L})^{-1}$, où $\tilde{L}$ est construite à partir de la matrice L à laquelle nous avons supprimé la colonne et la ligne correspondant au sommet $v = o$. Ceci coïncide avec la fonction de Green d'une marche aléatoire simple (à temps continu) $(S_t)_{t \geq 0}$ sur $\mathcal{T}$ qui serait tuée lorsqu'elle atteint o . Autrement dit, en notant τ le temps d'arrêt défini par $\tau = \min\{t \geq 0, S_t = o\}$, nous obtenons

$$\mathbb{E}[X_v X_w] = \mathbb{E}^v \left[\int_0^\tau 1_{\{S_t=w\}} dt \right],$$

avec $\mathbb{E}^v[\cdot]$ désignant l'espérance conditionnelle sachant $S_0 = v$. Dans le cas particulier où $d_v = d$ pour tout $v \neq o$, le même raisonnement montre que la fonction de covariance du champ libre gaussien discret correspond à la fonction de Green d'une marche aléatoire simple (à temps discret) tuée en o . Autrement dit,

$$\mathbb{E}[X_v X_w] = \mathbb{E}^v \left[\frac{1}{d} \sum_{n=1}^{\tau-1} 1_{\{S_n=w\}} \right].$$

En fait, il s'agit d'un champ libre gaussien avec des conditions de Dirichlet au bord. Le comportement de la fonction de covariance est donc déterminé par celui de la marche aléatoire simple sur le graphe. Il est bien connu que la marche aléatoire symétrique sur $\mathbb{Z}^d$ présente des comportements différents suivant la dimension. Lorsque $d = 2$ la fonction de Green associée à cette marche aléatoire admet un comportement logarithmique tandis qu'en dimension supérieure elle se comporte comme l'inverse d'une fonction puissance. Tandis que la majeure partie de la littérature se concentre sur le cas difficile de la dimension deux, les articles de [74, 73, 72] s'intéressent aux comportements du champ libre gaussien lorsque $d \geq 3$. En fait, il est possible de construire un champ libre gaussien discret sur un groupe infini $\mathcal{S}$ de type fini. Dans [74, 73, 72] les auteurs considèrent un tel objet sur $\mathbb{Z}^d$, $d \geq 3$. Le comportement de la marche aléatoire symétrique sur $\mathcal{S}$ conditionne le comportement de la fonction de covariance.

Il est possible de considérer un champ libre gaussien discret $(X_s)_{s \in \mathcal{S}}$ indexé par $\mathcal{S}$, un groupe infini de type fini. En effet, pour cela il suffit de considérer une marche aléatoire symétrique $(S_n)_{n \geq 0}$ sur Σ une partie génératrice symétrique de $\mathcal{S}$. Sous certaines hypothèses (cf [174]) sur la croissance de volume de boule, la fonction de covariance de ce champ libre gaussien discret admet un comportement similaire à celui étudié par Cipriani *et al.* dans [74, 73, 72].

Nous proposons ci-dessous un bref survol historique de la littérature concernant les travaux portant sur le champ-libre gaussien discret et les modèles liés. Les mathématiciens Bramson, Bolthausen, Deuschel, Giacomini, Kumagai et Zeitouni sont à la base d'une série d'articles étudiant le maximum du champ libre gaussien discret sur $\mathbb{Z}^2$, beaucoup d'effort sont faits pour prouver que la suite du maximum (recentré) est tendue [38, 39, 132, 52]. Dans les articles de Ding [84], puis Ding-Zeitouni [87], des inégalités de déviations (à droite et à gauche) précises du maximum recentré sont obtenues. Outre les méthodes de second moment, les principaux arguments utilisés sont alors des outils de comparaison comme l'inégalité de Slepian et ses variantes ainsi que des résultats classiques propres à la concentration de famille de variables aléatoires gaussiennes. Ding et Zeitouni proposent des modèles hiérarchiques plus manipulables, présentant une structure d'arbre sous-jacente, auxquels une comparaison avec le champ libre gaussien discret est possible. Signalons que ce genre d'arguments est repris dans l'article d'Acosta [1]. Enfin, Bramson, Ding et Zeitouni obtiennent un résultat de convergence en loi du maximum recentré vers une variable aléatoire non triviale, correspondant à une loi de Gumbel translatée aléatoirement [51]. Ils étendent ensuite leurs travaux à des marches aléatoires branchantes dans [50]. Ces résultats font écho aux travaux d'Aïdekon et Madaule, sur des modèles aléatoires dit « log-corrélé » [4, 150], notamment sur l'interprétation de la loi limite en tant que dérivée de martingale. Nous renvoyons le lecteur curieux vers les articles de survols de Duplantier, Rhodes, Sheffield et Vargas [90], ainsi que celui

d'Arguin [8] pour plus de détails à ce sujet. Récemment, Ding, Roy et Zeitouni ont aussi étendu leurs résultats de convergence, obtenus pour le champ libre gaussien discret sur $\mathbb{Z}^2$, à des modèles log-corrélés plus généraux [86].

Faisant suite à cet intense développement autour du champ libre gaussien discret sur $\mathbb{Z}^2$, Cipriani *et al.* proposent dans une série d'articles [74, 73, 72] sur l'étude du champ libre gaussien discret sur $\mathbb{Z}^d$, $d \geq 3$. Le comportement de la marche aléatoire symétrique sous-jacente est alors différent et rend la structure de covariance plus manipulable. En plus des outils de comparaison déjà présents dans la littérature, les auteurs de [74, 73, 72] utilisent la méthode de Stein-Chen pour obtenir des convergences de processus ponctuel de Poisson associés aux extrêmes du champ libre gaussien discret.

Percolation

Le modèle de premier temps de passage en percolation que nous avons présenté dans ce chapitre a été introduit par Hammersley et Welsh (cf. [61]).

L'approche abstraite des travaux de Benaïm et Rossignol a permis d'ouvrir la voie à Damron, Hanson et Sosoe qui, dans une série de plusieurs articles [80, 79], clarifient les arguments de [24] et étendent leurs résultats. Ils montrent, dans l'article [80], que les résultats de Chatterjee sur la variance sont toujours valides si l'on suppose uniquement que les poids sont dans l'espace d'Orlicz $L^2 \log L$. Dans leur deuxième article, ils obtiennent une inégalité de superconcentration exponentielle, identique à celle de Benaïm et Rossignol, sous la condition que les poids possèdent un moment exponentiel. Les outils qu'ils utilisent sont essentiellement les mêmes que ceux de [24]. En revanche, au lieu « d'encoder » leurs poids comme provenant d'une mesure gaussienne, ils proposent un encodage à partir de variables aléatoires de Bernoulli qui satisfont également une inégalité de Sobolev logarithmique. Cette observation accorde beaucoup plus de souplesse dans le choix des poids et permet d'étendre considérablement les travaux de [24]. Cependant cela induit beaucoup plus de technicité lors du calcul des dérivées discrètes intervenant dans l'inégalité de Sobolev logarithmique. Pour plus de détails sur ces modèles de percolation, nous renvoyons le lecteur vers l'article de survol [12].

Verres de spins

Concernant les différents résultats de superconcentration connus pour les modèles de verres de spins. À notre connaissance, il n'existe essentiellement que deux approches mathématiques pour étudier ces modèles : l'une utilise des méthodes d'interpolations tandis que l'autre utilise des méthodes de « second moment modifié ». Nous choisissons donc de partager notre présentation en deux, suivant que l'on utilise l'une ou l'autre des deux méthodes évoquées ci-dessus.

L'approche de la superconcentration développée par Chatterjee dans les modèles de verres de spin a donné naissance à toute une série d'articles. Bien qu'il s'agisse essentiellement de la même chose, les auteurs parlent de phénomène de chaos par rapport au désordre gaussien (« chaos disorder ») au lieu de superconcentration. L'équivalence entre ces deux notions est détaillée dans l'ouvrage [61].

De nombreux travaux ont été fait pour répondre à la question suivante :

dans un modèle de verres de spin donné, quelle est l'influence de la présence d'un champ externe sur le phénomène de superconcentration ?

Plus précisément, cela revient à ajouter un vecteur $h \in \mathbb{R}^n$ à l'hamiltonien. L'énergie d'une configuration est alors mesurée par

$$-H_n(\sigma) + \sum_{i=1}^n h_i \sigma_i \quad \text{avec} \quad \sigma = (\sigma_1, \dots, \sigma_n) \in C_n$$

et où H_n est l'hamiltonien initial du système. Chen montre dans [63] que le phénomène de superconcentration, obtenu par Chatterjee dans le modèle SK, est toujours valable en présence d'un champ externe. Dans cet article, ainsi que les suivants, il s'appuie sur les travaux de Chatterjee et sur la formule de représentation de la variance pour étudier l'énergie libre $F_{n,\beta}$. Cependant, il utilise également des arguments propres à la théorie des verres de spin, comme les répliques symétriques de Guerra ou des résultats pointus comme la formule de Parisi pour le modèle considéré. Chen démontre alors la présence de chaos pour les modèles de p -spin pairs mixtes dans [64]. Chen, Dey et Panchenko étudient le modèle des p -spin mixtes avec présence d'un champ externe dans [65] et démontrent que la présence du champ externe empêche le phénomène de superconcentration ; en d'autres termes, ils montrent que l'inégalité de Poincaré fournit le bon ordre de grandeur de la variance, pour ce modèle, en présence d'un champ externe. Citons également, les articles [68, 69, 70, 11] de Chen, avec différents co-auteurs, dans la même thématique. Chen, Hsieh, Hwang et Sheu montrent la présence de chaos, en le désordre, pour le modèle de champ moyen sphérique dans [67]. Dans [66], Chen, Handschy et Lerman montrent que l'état fondamental (le minimum de l'hamiltonien, que l'on désigne en anglais par « ground state ») du modèle des p -spins mixtes paires est superconcentré si et seulement si le champ externe est nul ; en présence d'un champ externe l'inégalité de Poincaré fournit le bon ordre de grandeur pour la variance de l'état fondamental et un théorème de convergence en loi vers une gaussienne est obtenu. Enfin, dans [16] Baik et Lee étudient l'énergie libre pour le modèle SK sphérique, pour lequel les spins $\sigma \in \mathbb{S}^n$ au lieu du cube discret C_n , et démontrent un phénomène de transition de phase. Ils obtiennent le résultat de convergence en loi suivant

Théorème 12.6.2 (Baik-Lee). *Si $0 < \beta < \frac{1}{2}$, alors,*

$$n(F_{n,\beta} - F_\beta) \rightarrow \mathcal{N}(f, \alpha) \quad \text{lorsque} \quad n \rightarrow \infty$$

où $\mathcal{N}(f, \alpha)$ désigne une gaussienne de moyenne f et de variance α . Tandis que si $\beta > \frac{1}{2}$, la loi limite est différente

$$\frac{1}{\beta - \frac{1}{2}} n^{2/3} (F_{n,\beta} - F_\beta) \rightarrow \mathcal{TW} \quad \text{lorsque} \quad n \rightarrow \infty,$$

où $\mathcal{TW}$ correspond à la loi de Tracy-Widom provenant de la théorie des matrices aléatoires.

L'expression détaillée des paramètres de renormalisation est fournie dans [16], les auteurs précisent également le lien entre l'état fondamental et la plus grande valeur propre d'une matrice du GOE.

Parallèlement aux travaux qui font suite aux résultats de Chatterjee et à son approche, d'autres auteurs ont étudié des modèles de verres dans lesquels une structure d'arbre est présente. Une grande partie de ces travaux est antérieure à ceux de Chatterjee et les questions abordées sont légèrement différentes. L'étude des extrêmes et de la convergence en loi de l'énergie libre font partie de leurs questionnements, cependant les auteurs cherchent à obtenir des résultats plus précis comme la convergence de processus ponctuels de Poisson associés aux extrêmes. Les résultats obtenus sont, pour la plupart, démontrés via la méthode de second moment (et ses variantes). Celle-ci repose sur l'inégalité de Paley-Zygmund et une étude fine, rendue possible grâce à la structure d'arbres sous-jacente, des trajectoires des particules. Nous renvoyons le lecteur vers l'ouvrage de Bovier [46] ou l'excellent cours de Kistler [130] pour plus de détails. Comme nous le verrons dans la section suivante, cette méthodologie de second moment reposant sur la structure hiérarchique de l'arbre sera également un outil essentiel dans l'étude de champ libre gaussien discret sur $\mathbb{Z}^2$. Bien qu'efficace, ces techniques sont assez lourdes en calcul. Comme présenté dans [202], il est possible de retrouver les résultats obtenus dans ces articles par des arguments d'interpolation.

Dans [49], Bovier, Kurkova et Löwe obtiennent des résultats de fluctuations pour les modèles de verres de spin suivants : le REM (« Random Energy Model ») et le p -spin SK. Leurs résultats de fluctuations concernent des convergences en loi de la fonction de partition ou de convergence de processus ponctuel de Poisson. Nous allons présenter brièvement le REM et les résultats obtenus le concernant. Signalons que ce modèle fait partie des exemples pour lesquels les méthodes d'interpolation par semi-groupes sont efficaces pour retrouver l'ordre de grandeur de la variance de la fonction de partition.

Le REM est sans doute le plus simple des modèles de verres de spins existant et est construit comme suit : on considère $(X_\sigma)_{\sigma \in S_n}$ une famille de gaussiennes réelles standards indépendantes et de même loi. On définit l'hamiltonien du système H_n par

$$H_n(\sigma) = \sqrt{n}X_\sigma, \quad \sigma \in C_n.$$

Les auteurs de [49] s'intéressent à la fonction de partition $Z_{n,\beta}$ définie par $Z_{n,\beta} = 2^{-n} \sum_{\sigma \in C_n} e^{\beta H_n(\sigma)}$, où $\beta > 0$ joue le rôle de l'inverse de la température. Parmi les résultats obtenus dans cet article mentionnons ceux qu'il est possible de redémontrer à l'aide des méthodes d'interpolations.

Théorème 12.6.3 (Bovier, Kurkova, Löwe). *Avec les notations précédentes, les convergences en loi suivantes sont vérifiées*

1. Si $\beta < \sqrt{\log 2/2}$, alors

$$\exp \frac{n}{2} (\log 2 - \beta^2) \log \frac{Z_{n,\beta}}{\mathbb{E}[Z_{n,\beta}]} \rightarrow \mathcal{N}(0, 1) \quad \text{lorsque } n \rightarrow \infty$$

2. Si $\beta > \sqrt{2 \log 2}$, alors

$$\log Z_{n,\beta} - \mathbb{E}[\log Z_{n,\beta}] \rightarrow \log \int_{\mathbb{R}} e^{\alpha z} \mathcal{P}(dz) - \mathbb{E} \log \int_{\mathbb{R}} e^{\alpha z} \mathcal{P}(dz) \quad \text{lorsque } n \rightarrow \infty$$

où $\mathcal{P}$ désigne le processus ponctuel de Poisson sur $\mathbb{R}$ de mesure d'intensité $e^{-x} dx$.

Remarque. L'article contient beaucoup d'autres résultats, complétant l'énoncé du théorème précédent. Notons également qu'après une renormalisation et un changement d'échelle, on peut retrouver la fonction d'énergie libre $F_{n,\beta}$ introduite plus tôt à partir de $Z_{n,\beta}$.

Dans la continuité de cet article, Bovier et Kurkova publient deux articles sur le GREM [47, 48], contenant le même type de résultats de convergence et, implicitement, montrent que la fonction de partition est superconcentrée dans certains cas. Toujours avec des méthodes de second moment (modifié), Kistler et Schmidt obtiennent un théorème de convergence pour des processus ponctuels de Poisson dans [131]. Ils insistent sur la présence d'une structure hiérarchique et de la présence d'un arbre dans leur modèle pour mettre en place leurs démonstrations. Ils parviennent notamment à traiter la convergence d'un processus ponctuel de Poisson N_N^α . Celui-ci permettant une interpolation entre le modèle du REM et de la marche aléatoire branchante. Signalons également l'article de Subag et Zeitouni [181] qui obtiennent un résultat de convergence en loi de l'état fondamental (renormalisé) vers une loi de Gumbel (négative).

