



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *02/12/2013* par :

THIBAUT LE GOUIC

Localisation de masse et espaces de Wasserstein

JURY

JEAN-MARC AZAÏS
PHILIPPE BERTHET
BENOÎT CADRE
PATRICK CATTIAUX
EUSTASIO DEL BARRIO
CHARLES SUQUET

Examineur
Directeur de thèse
Président du jury
Examineur
Rapporteur
Rapporteur

Université Paul Sabatier
Université Paul Sabatier
ENS Rennes
Université Paul Sabatier
Universidad de Valladolid
Université Lille 1

École doctorale et spécialité :

MITT : Domaine Mathématiques : Mathématiques appliquées

Unité de Recherche :

Institut de Mathématiques de Toulouse (UMR 5219)

Remerciements

Ces trois années de travail ont été pour moi riches en apprentissage et ont grandement fait mûrir ma vision des mathématiques. L'aboutissement de cette thèse doit beaucoup à ceux qui m'ont entouré pendant cette période.

Mon directeur de thèse Philippe Berthet a su me donner un cadre qui m'a permis de ne pas me perdre longtemps dans les mauvaises directions tout en me laissant une très grande liberté pour développer mes idées. Il est aussi une intarissable source d'idées nouvelles. Pour tout ça et le partage de sa vision du monde, merci Philippe.

Je souhaite aussi remercier Eustasio del Barrio et Charles Suquet d'avoir accepté d'être rapporteurs. Leurs conseils et remarques m'ont permis de corriger plusieurs erreurs et de comprendre comment valoriser certains résultats. Je suis très honoré qu'ils aient accepté. Je remercie aussi Jean-Marc Azaïs, Benoît Cadre et Patrick Cattiaux d'avoir accepté de faire partie de ce jury malgré leurs emplois du temps chargés.

Je remercie aussi l'équipe de Statistiques et Probabilités de l'Institut Mathématiques de Toulouse pour l'environnement extrêmement stimulant qu'elle m'a fourni, et notamment Fabrice Gamboa et Michel Ledoux qui ont su m'accorder un peu de temps quand j'en avais besoin, et tous les doctorants, pour le temps que je leur ai volé en leur posant mille questions. Merci Adil, Maikol, Santiago, Thibault, Rim, Chloé, Paul, Yohann, Hélène, Claire, Julie, Nil, Benoît, Malika, Mélanie et Guillem et Jérémy d'"en face".

Merci à l'équipe administrative de l'Institut Mathématiques de Toulouse pour son aide et notamment Marie-Laure et sa curiosité.

Je remercie aussi mes amis Damien, Carole, Patrick, Marie, Nicolas, Léa, Jérémy, Pauline, Kathleen, Nicko pour tout ce qu'ils ont pu m'apporter ces années.

Je tiens aussi à remercier ma famille pour son soutien inconditionnel.

Et enfin, merci Sophie.

Table des matières

Remerciements	i
Table des matières	iii
Introduction	vii
1 Préliminaires techniques	1
1.1 Recouvrement et packing	1
1.1.1 Définitions	1
1.1.2 Propriétés élémentaires	2
1.1.3 Mesures	3
1.1.4 Dimensions	5
1.1.5 Intégrale de Dudley	6
1.2 Espace de longueurs géodésique	6
1.2.1 Espace de longueurs	6
1.2.2 Espace géodésique	8
1.2.3 Barycentre	9
1.3 Espaces de Wasserstein	10
1.3.1 Définitions	10
1.3.2 Propriétés	11
1.4 Mesures gaussiennes sur un espace de Banach	12
1.4.1 Définition	12
1.4.2 Intégrales à valeurs dans un espace de Banach	13
1.4.3 Espaces à noyau auto-reproduisant	15
1.4.4 Inégalités de transport	18
2 Localisation de masse	21
2.1 Présentation du problème	21
2.1.1 Définitions	24
2.2 Premières propriétés de la localisation de masse	30
2.2.1 Existence	30

2.2.2	Régularité de τ	31
2.3	Consistance de la localisation de masse	32
2.3.1	τ -tension	32
2.3.2	Consistance de τ	34
2.3.3	Consistance du minimiseur	34
2.3.4	Continuité du minimiseur	37
2.3.5	Cas de la mesure empirique	38
2.4	Applications statistiques	39
2.4.1	Localisation de masse	39
2.4.2	Quantification robuste	41
2.5	Démonstrations	49
2.6	Conclusion	56
3	Mesure empirique et distance de Wasserstein	59
3.1	Présentation	59
3.2	Première majoration	61
3.3	Applications	62
3.3.1	Espaces de dimension finie	62
3.3.2	Mesures gaussiennes sur les espaces de Banach	62
3.4	Retour sur la localisation de masse	64
3.5	Déviations	65
3.5.1	Cas des espaces de dimension finie	66
3.5.2	Cas des mesures gaussiennes	66
3.5.3	Cas général	66
3.6	Cas dépendant	67
3.7	Applications statistiques	69
3.7.1	Test statistique	69
3.7.2	Régions de confiance	70
3.8	Démonstrations	70
3.8.1	Première majoration	70
3.8.2	Espaces de dimension finie	74
3.8.3	Mesures gaussiennes	77
3.8.4	Localisation de masse	79
3.8.5	Cas dépendant	83
3.9	Conclusion	86
4	Barycentre dans l'espace de Wasserstein	87
4.1	Introduction	87
4.2	Propriétés du barycentre	88
4.2.1	Existence	88
4.2.2	Continuité	91

4.3	Perspectives d'applications	92
4.3.1	Agrégation d'estimateurs d'une mesure	92
4.3.2	Mélange d'images	93
4.4	Démonstrations	94
4.4.1	Existence	95
4.4.2	Continuité	96
A	Rappels et notations	99
A.1	Rappels et lemmes	99
A.2	Notations	101

Introduction

Le plus étonnant dans le hasard est certainement le nombre de résultats « déterministes » à son sujet. Les plus connus sont la loi des grands nombres, le théorème central limite, ou encore la loi du logarithme itéré, et ils permettent d'avoir des certitudes sur des objets aléatoires tels que la moyenne empirique. L'objet de la statistique est, à l'instar de ces théorèmes, de déduire de l'information sur une mesure de probabilité inconnue au vu de données d'observation dont le comportement est gouverné par cette mesure. La mesure empirique associée à un échantillon est un outil synthétique qui contient toute l'information que contient l'échantillon ; c'est pourquoi elle joue un rôle important en statistique. Ce manuscrit présente trois résultats théoriques sur les mesures de probabilité qui permettent de nouvelles utilisations de la mesure empirique. Deux de ces résultats énoncent chacun une propriété stable d'une mesure de probabilité par approximation par la mesure empirique (chapitres 2 et 4), et le troisième contrôle l'erreur d'approximation par la mesure empirique (chapitre 3).

Premier chapitre

Pour faciliter la lecture de ce manuscrit, un premier chapitre présente les notions et les résultats qui seront utilisés dans les chapitres suivants qui sont consacrés aux résultats obtenus pendant la thèse. Les notions de recouvrements et de packing sont présentées, et quelques propriétés sont présentées, afin de permettre au lecteur de se familiariser si besoin avec ces notions. Quelques usages des outils sont aussi mentionnés. Ensuite, le chapitre définit les espaces de longueurs et en présente les premières propriétés qui seront ensuite utilisées dans le manuscrit. Le cas particulier des espaces de Wasserstein est ensuite introduit, ainsi que quelques propriétés connues. Ces espaces sont l'objet d'une littérature abondante ; seuls les propriétés les plus élémentaires sont décrites ici. Enfin, les mesures gaussiennes sur les espaces de Banach sont définies et quelques propriétés connues et utilisées plus tard dans le manuscrit sont rassemblées, notamment le lien entre la mesure des petites boules et le nombre de recouvrement de la boule unité de l'espace à noyau auto-reproduisant, ainsi que les inégalités de transport que vérifient ces mesures.

Second chapitre

Le chapitre 2 s'intéresse à la localisation de la masse d'une mesure de probabilité μ . Lorsque μ est régulière, les ensembles de niveau de sa densité fournissent une bonne notion pour localiser les zones « denses » de masse, mais perdent leur sens pour les mesures à support fini, comme dans le cas de la mesure empirique. Ce second chapitre définit alors une notion générale de localisation de masse pour les mesures sur un espace métrique. Nous définissons pour cela une fonction τ dite de *taille*, sur les fermés d'un espace métrique, basée sur leur packing. Les ensembles de plus petite τ -taille ayant une masse d'au moins $1 - \alpha$ donnée permettent de localiser les zones denses de μ , même dans les cas irréguliers. Nous montrons que la fonction qui associe à α la plus petite valeur de τ parmi les ensembles de taille au moins $1 - \alpha$ dépend continument de α lorsque la mesure μ est à support connexe. Nous montrons aussi que les ensembles de plus petite τ -taille pour μ et α fixés dépendent aussi continument de μ et de α , pour la distance de Hausdorff. Nous en tirons une nouvelle méthode de quantification de μ , à la fois robuste et stable.

Troisième chapitre

La mesure empirique possède également la propriété de converger vers la mesure μ dont est issu l'échantillon, pour différentes notions de distance sur l'espace des mesures de probabilité sur un espace métrique (E, d) . Le chapitre 3 s'intéresse à la distance de Wasserstein $W(\mu, \mu_n)$ entre une mesure empirique μ_n et la mesure μ . Un premier résultat majore l'espérance de cette distance lorsque l'espace (E, d) est borné, par un majorant fonction du nombre de recouvrement de (E, d) . À partir de ce résultat, nous retrouvons l'ordre de vitesse de convergence de cette espérance qui est de $n^{-1/d}$ où d est la dimension de l'espace E . Ce résultat était déjà connu sur \mathbb{R}^d mais est nouveau pour les espaces non vectoriels de dimension finie, où dimension s'entend dans le sens où le nombre de recouvrement des boules de l'espace est polynomial. Cette nouveauté permet d'appliquer la majoration à des espaces tels que la sphère. La majoration dans le cas des espaces bornés s'étend aussi au cas des mesures définies sur un espace de Banach. Nous montrons alors la majoration suivante

$$EW(\mu, \mu_n) \leq O(\varphi^{-1}(n^{-\zeta})),$$

où $\varphi(t) = \mu(B(0, t))$ est la mesure de la boule centrée en 0 de rayon t , et $\zeta < 1/2$. Cette vitesse de convergence correspond aux vitesses des quantifieurs optimaux (c'est-à-dire à la mesure la plus proche de μ en distance de Wasserstein parmi les mesures à support à n points) dont nous avons pu calculer φ . Ensuite, nous avons obtenu une majoration de cette même espérance pour les mesures définies sur un espace métrique quelconque. Le majorant est alors une fonction du nombre de

recouvrement des sous-ensembles de (E, d) . Des hypothèses classiques de concentration permettent de déduire des ces majorations des inégalités de déviation. Une partie de cet article [Boissard et Le Gouic, 2011] a été écrit en collaboration avec E. Boissard et a donné lieu à un article à paraître dans les Annales de l’Institut Henri Poincaré.

Quatrième chapitre

La chapitre 4 s’intéresse au barycentre d’une famille finie de mesures de probabilité. La moyenne de Fréchet fournit une extension de la notion de barycentre aux espaces métriques, nous permettant de le définir sur les espaces de Wasserstein. Nous montrons son existence, puis, en étudions les propriétés de continuité par rapport aux mesures de probabilité. Nous discutons enfin de l’application pratique de ces résultats en agrégation de mesures empiriques et en mélange d’images. Une partie de ce chapitre a été écrit en collaboration avec E. Boissard et J.M. Loubes et a donné lieu à un article [Boissard et al., 2011] à paraître dans le journal Bernoulli.

Commentaires

Plusieurs résultats classiques qui interviennent dans les démonstrations ont été rassemblés en annexe A avec une liste des notations utilisées

Tout au long du manuscrit, toute mesure est une mesure de probabilité sauf mention du contraire.

Chapitre 1

Préliminaires techniques

Ce chapitre expose des résultats, classiques dans leur discipline respective, qui seront utilisés dans la suite. Les démonstrations des théorèmes importants ne sont pas présentées, mais une référence qui les contient est donnée au début de chaque paragraphe.

1.1 Recouvrement et packing

Cette première section présente les notions de recouvrement et de packing. Ce sont deux approches duales (comme le montre [Saint Raymond et Tricot, 1988]) qui permettent de « mesurer » la taille d'un ensemble, au moyen, *in fine*, de mesures. Nous y présentons des usages de ces deux notions dans le but essentiellement de motiver le choix du packing dans la définition de la notion de localisation de masse du chapitre 2.

1.1.1 Définitions

Le cadre général de définition des recouvrements est celui d'un espace métrique (E, d) . Le diamètre d'un sous-ensemble de E est alors défini par $\text{diam}(B) = \sup\{d(x, y); x, y \in B\}$.

Définition 1.1 (*t*-recouvrement). *Un ensemble $\{B_i\}_{i \in I}$ de sous-ensembles de E est appelé un ***t*-recouvrement** d'un ensemble B si le diamètre de tout B_i n'excède pas $2t$ et*

$$B \subset \bigcup_{i \in I} B_i.$$

*On appelle alors **nombre de recouvrement pour t** , noté $\mathcal{N}(B, t)$, le cardinal du plus petit *t*-recouvrement de B .*

Le logarithme de $\mathcal{N}(B, t)$ est parfois appelé l'entropie métrique de B , ou encore l'entropie de Kolmogorov.

Remarque 1.2. Dans le chapitre 3, la même notation \mathcal{N} est utilisée pour un type de recouvrement un peu différent. Alors qu'ici, un t -recouvrement peut être composé de n'importe quels sous-ensembles de l'espace, dans le chapitre 3, il est question de recouvrement par des boules. Ces deux notions sont liées par la relation

$$\mathcal{N}_b(B, 2t) \leq \mathcal{N}(B, t) \leq \mathcal{N}_b(B, t),$$

mais ne sont pas nécessairement égales. Dans ce paragraphe, \mathcal{N} désignera toujours le recouvrement par un sous-ensemble quelconque.

Dans le chapitre 2 sur la localisation de masse, nous utiliserons plutôt la notion de packing.

Définition 1.3 (t -séparé). Un ensemble B est dit **t -séparé** si la distance entre deux points distincts de B est strictement plus grande que t . On appelle alors **packing pour t** , noté $\mathcal{M}(B, t)$, le cardinal du plus grand sous-ensemble t -séparé de B .

Le logarithme de $\mathcal{M}(B, t)$ est parfois appelé la capacité de B .

1.1.2 Propriétés élémentaires

Les deux notions sont intuitivement très liées, et semblent contenir chacune approximativement la même information sur la taille de l'ensemble. La proposition suivante démontrée dans [Kolmogorov et Tihomirov, 1961] fournit une estimation du lien entre \mathcal{N} et \mathcal{M} .

Proposition 1.4. Pour tout ensemble B d'un espace métrique (E, d) ,

$$\mathcal{M}(B, 2t) \leq \mathcal{N}(\bar{B}, t) = \mathcal{N}(B, t) \leq \mathcal{M}(B, t) = \mathcal{M}(\bar{B}, t).$$

Pour un ensemble B donné, \mathcal{M} peut être vu comme une fonction $t \mapsto \mathcal{M}(B, t)$, dite fonction de packing. La continuité à droite de la fonction de packing joue un rôle important dans les résultats sur la localisation de masse.

Proposition 1.5. Pour tout ensemble B d'un espace métrique (E, d) , la fonction $t \mapsto \mathcal{M}(B, t)$ est décroissante et continue à droite.

1.1.3 Mesures

Le recouvrement et le packing sont des outils naturels pour définir des mesures « uniformes ». La mesure de Hausdorff est sûrement la plus connue des mesures répondant à l'idée d'uniformité. Elle coïncide notamment (à une constante près) avec la mesure de Lebesgue sur \mathbb{R}^d .

Définition 1.6 (Mesure de Hausdorff). *Soit (E, d) un espace métrique. Soit $s > 0$. Pour un ensemble B , on définit la mesure de Hausdorff de dimension s par*

$$\mathcal{H}^s(B) = \liminf_{\delta \rightarrow 0} \left\{ \sum \text{diam}(B_i)^s; \{B_i\}_{i \in I} \text{ } \delta\text{-recouvrement de } B \right\}.$$

La mesure de Hausdorff n'est pas une fonction de $\mathcal{N}(B, \cdot)$ puisque sa définition fait intervenir le diamètre de chacun des sous-ensembles qui recouvrent B , ce qui n'est pas le cas de \mathcal{N} . En revanche, il est possible de définir de manière analogue des mesures à partir de $\mathcal{N}(B, \cdot)$ et $\mathcal{M}(B, \cdot)$.

Pour cela, nous utilisons la construction de Carathéodory, qui permet de construire une mesure à partir d'une fonction sur les ensembles, à l'aide des deux théorèmes suivants et de la notion de mesure extérieure.

Définition 1.7 (Mesure extérieure). *Soit E un ensemble. Une fonction de l'ensemble des parties de E dans $\overline{\mathbb{R}}_+$ est appelée une mesure extérieure si elle nulle pour l'ensemble vide, monotone pour l'inclusion, et σ -sous-additive.*

Le théorème suivant permet alors de construire une mesure extérieure à partir d'une fonction d'ensemble quelconque.

Théoreme 1.8 (Construction d'une mesure extérieure). *Soient E un espace et C une classe de sous-ensembles de E contenant l'ensemble vide. Soit φ une fonction de C à valeurs dans $\overline{\mathbb{R}}_+$. Alors,*

$$\mu^*(B) = \inf \left\{ \sum_{i \geq 1} \varphi(B_i); B \subset \bigcup_{i \geq 1} B_i, \forall i \geq 1, B_i \in C \right\},$$

définit une mesure extérieure sur E muni de la tribu des parties de E .

Il est alors possible de vérifier qu'une mesure extérieure est une mesure borélienne si elle est métrique.

Théoreme 1.9 (Mesure métrique extérieure). *Soit (E, d) un espace métrique. Une mesure extérieure μ^* qui vérifie pour tout $A, B \subset E$ tels que $\inf\{d(a, b); a \in A, b \in B\} > 0$,*

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$$

est une mesure borélienne (on parle de mesure extérieure métrique).

Il devient donc simple de construire une mesure à partir d'une fonction d'ensemble, c'est ce que nous allons faire avec le recouvrement et le packing.

Définition 1.10 (Mesure de recouvrement, mesure de packing). *Soit (E, d) un espace métrique. Pour $s > 0$, la **mesure de recouvrement** est la mesure définie par*

$$\mu_{\mathcal{N}}(B) = \inf \left\{ \sum_{i \geq 1} \liminf_{t \rightarrow 0} \mathcal{N}(B_i, t)t^s; \forall i \geq 1; B_i \subset B \right\},$$

pour tout borélien B . La mesure de packing $\mu_{\mathcal{M}}$ est définie de manière analogue.

Il est assez simple de montrer que ces mesures sont invariantes par translation dans le cas $E = \mathbb{R}^d$, et donc que $\mu_{\mathcal{N}}$ et $\mu_{\mathcal{M}}$ ne sont rien d'autre que la mesure de Lebesgue, à une constante près.

Remarque 1.11. *Le choix de prendre la limite inférieure dans cette définition est arbitraire ; nous aurions pu prendre la limite supérieure et obtenir une mesure $\bar{\mu}$. Cependant, comme $\bar{\mu}$ est aussi invariante par translation, il s'agit aussi de la mesure de Lebesgue.*

Remarque 1.12. *Les définitions ici diffèrent de celles qui portent habituellement les noms de mesure de recouvrement et packing, et qui ont été étudiées dans [Saint Raymond et Tricot, 1988]. La notion qu'ils étudient utilise le diamètre de chaque boule du recouvrement pour $\mu_{\mathcal{N}}$ et l'écart au plus voisin pour $\mu_{\mathcal{M}}$. Ce qui nous importe est surtout le fait que l'on conserve une « bonne » mesure dans le sens où elle correspond à la mesure de Lebesgue sur \mathbb{R}^d .*

Ces constructions de mesures permettent d'obtenir toutes les propriétés des mesures à partir des notions de packing et de recouvrement. Un théorème dû à [Kolmogorov et Tihomirov, 1961] permet de comprendre différemment pourquoi le packing et le recouvrement sont de bonnes notions pour mesurer les ensembles.

Théoreme 1.13. *Sur \mathbb{R}^d , muni la norme $\|\cdot\|_p$ pour ≥ 1 , il existe $\theta_{d,p}$ et $\zeta_{d,p}$ tels que tout ensemble Jordan mesurable (i.e. tel que le bord topologique est de mesure de Lebesgue nulle) vérifie lorsque $t \rightarrow 0$,*

$$\mathcal{M}(B, t)t^d \rightarrow \theta_{d,p}\lambda(B),$$

et

$$\mathcal{N}(B, t)t^d \rightarrow \zeta_{d,p}\lambda(B).$$

Pour $p = 1$, il est facile de voir que $\zeta_{1,p} = \theta_{1,p} = 1$. Pour le packing pour $p = 2$, et $d \leq 2$, les calculs sont relativement simples et donnent $\theta_{1,2} = 1$ et $\theta_{2,2} = \frac{\pi}{3\sqrt{3}}$.

Pour la dimension 3, une démonstration assistée par ordinateur a été fournie par Thomas Hales déterminant la valeur de $\theta_{3,2}$ (il s'agit de la conjecture de Kepler) et pour les dimensions supérieures, la valeur exacte de $\theta_{d,p}$ est toujours inconnue.

1.1.4 Dimensions

Le recouvrement et le packing interviennent aussi en géométrie fractale à travers les dimensions. Ce paragraphe présente trois notions de dimension. Leurs propriétés sont très succinctement décrites ici. Le lecteur pourra se référer à [Falconer, 2003] pour un traitement plus complet.

Il existe différentes notions de dimension pour les espaces métriques, dont la plus connue est probablement celle de Hausdorff. Elle coïncide avec la dimension algébrique dans le cas des espaces vectoriels et est définie à partir de la mesure de Hausdorff.

Définition 1.14 (Dimension de Hausdorff). *Soit (E, d) un espace métrique. La dimension de Hausdorff d'un sous-ensemble B est définie par*

$$\dim_H(B) = \inf\{s; \mathcal{H}^s(B) = 0\} = \sup\{s; \mathcal{H}^s(B) = +\infty\}.$$

Dans \mathbb{R}^d , le recouvrement et le packing d'une boule sont de l'ordre de t^{-d} lorsque t tend vers 0, $\mathcal{N}(B, t) \sim \mathcal{M}(B, t) \sim ct^{-d}$. On a alors envie de définir la dimension comme la limite en 0 du rapport des logarithmes de ces fonctions sur le logarithme de $1/t$. Mais cette limite n'étant pas toujours définie, cette idée donne lieu alors deux définitions ; l'une correspondant à la limite inférieure et l'autre à la limite supérieure.

Définition 1.15 (Dimensions de recouvrement et de packing). *Soit (E, d) un espace métrique. La dimension de recouvrement inférieure d'un sous-ensemble B de E est définie par*

$$\underline{\dim}_{\mathcal{N}}(B) = \liminf_{t \rightarrow 0} \frac{\log \mathcal{N}(B, t)}{-\log t}.$$

On définit, de même, la dimension de recouvrement supérieure $\overline{\dim}_{\mathcal{N}}(B)$ avec la limite supérieure. Les définitions de dimension de packing inférieure et supérieure sont analogues.

Ces dimensions correspondent à la dimension de Hausdorff sur les sous-variétés de \mathbb{R}^d . L'avantage de ces dimensions sur celle de Hausdorff est avant tout qu'elles sont plus facilement calculables. Elles souffrent cependant d'un inconvénient pour la théorie de la géométrie fractale, qui est qu'un ensemble a la même dimension que sa fermeture. Un ensemble dense dénombrable (intuitivement petit) se retrouve alors avec la même dimension que l'espace entier. Cet inconvénient n'a pas d'incidence sur les usages qui seront faits dans le chapitre sur la localisation de masse, puisque nous y considérons des ensembles fermés (support de mesure).

1.1.5 Intégrale de Dudley

On retrouve le recouvrement (et implicitement le packing) dans l'intégrale de Dudley. Soit $(X_t)_{t \in T}$ un processus gaussien (voir [Ledoux et Talagrand, 1991] pour une définition formelle). Alors, on définit sur T une pseudo-distance par

$$d_X(s, t) = \|X_s - X_t\|_2, s, t \in T.$$

L'intégrale de Dudley est alors définie par

$$I(X) = \int_0^{+\infty} \sqrt{\log \mathcal{N}(T, u)} du.$$

Cette intégrale trouve son origine dans [Dudley, 1967], à travers le théorème suivant.

Théorème 1.16. *Dans le cadre défini ci-dessus,*

$$E \sup_{t \in T} X_t \leq 24I(X).$$

Ce théorème est notamment utilisé pour démontrer des critères pour les classes de Donsker. Ce théorème et ses conséquences achèvent notre partie présentant et motivant le packing.

1.2 Espace de longueurs géodésique

Ce paragraphe définit un espace de longueur et en présente quelques résultats classiques. Le livre [Burago et al., 2001] fournit une version détaillée de ces résultats.

La notion d'espace de longueur précise celle d'espace métrique en définissant la métrique comme la longueur du plus court chemin entre deux points donnés. Une illustration pertinente de ce concept est celle d'un randonneur en montagne : partant d'un sommet d'un pic A , le chemin d'un randonneur voulant atteindre le sommet d'un autre pic B distinct sera plus long que la distance à vol d'oiseau entre les deux sommets. Le randonneur devra descendre la pente de A pour ensuite remonter celle du pic B . Il est alors naturel de considérer la métrique qui associe à deux points la distance que parcourt le randonneur suivant le plus court chemin entre ces deux points. Les espaces de longueurs formalisent cette illustration.

1.2.1 Espace de longueurs

Un espace topologique E peut être muni d'une structure d'espace de longueur. Cette structure se compose d'une classe de chemins (fonction continue d'un intervalle I à valeurs dans E) dits admissibles, et d'une longueur L associée à chaque

chemin. La classe de chemins admissibles A est supposée stable par restriction, concaténation, et reparamétrisation continue. Les quatre propriétés suivantes sont supposées vérifiées par les longueurs.

1. La longueur des chemins est additive.
2. La longueur de la restriction d'un chemin défini sur $[a; b]$ à un chemin sur $[a; t]$ est continue par rapport à t .
3. La longueur est invariante par reparamétrisation continue.
4. La structure de longueur coïncide avec la topologie dans le sens où, pour tout $x \in E$ et tout voisinage U_x de x ,

$$\inf\{L(\gamma); \gamma(a) = x, \gamma(b) \in E \setminus U_x\} > 0.$$

Un espace topologique muni d'une telle structure de longueur est appelé un *espace de longueur*. Dans ce contexte, il est possible de définir une distance d_L entre deux points x et y de l'espace comme l'infimum des longueurs L des chemins joignant les deux points x et y :

$$d_L(x, y) = \inf \{L(\gamma); \gamma(a) = x, \gamma(b) = y, \gamma \in A\}. \quad (1.1)$$

De même, étant donnée une distance d , il est possible de définir la longueur L_d d'un chemin continu $\gamma : [a; b] \mapsto E$:

$$L_d(\gamma) = \sup \left\{ \sum_{i=1}^n d(\gamma(x_i), \gamma(x_{i+1})); a = x_1 \leq x_2 \leq \dots \leq x_n = b, n \geq 2 \right\}.$$

Ainsi, un espace métrique (E, d) peut être muni d'une structure d'espace de longueur (pour la classe des chemins continus) qui fournit une nouvelle distance $\hat{d} = d_{L_d}$ dite *intrinsèque*. Une définition synthétique d'un espace métrique de longueur est alors la suivante.

Définition 1.17 (Espace de longueur). *Soit (E, d) un espace métrique. On dit que (E, d) est un **espace de longueur** si*

$$d = d_{L_d}.$$

La longueur L_d possède bien les propriétés demandées à une structure d'espace de longueur et est, de plus, semi-continue inférieurement.

Proposition 1.18. *Soit (E, d) un espace de longueur. Soit $(\gamma_n)_{n \geq 1}$ une suite de chemins continus de même domaine, qui converge ponctuellement vers un chemin γ . Alors,*

$$\liminf_n L_d(\gamma_n) \geq L(\gamma).$$

1.2.2 Espace géodésique

L'infimum dans la définition de d_L (1.1) n'est pas nécessairement atteint, y compris lorsque l'espace est complet. Pour définir certaines notions (notamment le barycentre) des prochains chapitres, il sera nécessaire que ce minimum soit atteint. Dans ces cas, on parle d'espace de longueur géodésique.

Définition 1.19. Soit (E, d) un espace de longueur. E est dit **géodésique** si pour tout $x, y \in E$, il existe un chemin γ tel que $\gamma(a) = x$ et $\gamma(b) = y$ et

$$d(x, y) = L_d(x, y).$$

$\mathbb{R}^2 \setminus \{0\}$ est un exemple trivial d'espace de longueur non géodésique non complet. Afin d'expliquer un cas d'espace de longueur non géodésique complet, nous introduisons les espaces localement compacts.

Définition 1.20. Un espace E est dit **localement compact** si tout point de E possède un voisinage précompact.

Le théorème suivant assure alors l'existence du chemin le plus court.

Théorème 1.21. Soit (E, d) un espace de longueur complet et localement compact. Alors, E est un espace de longueur géodésique.

L'exemple suivant (la banane de Venet) est un cas d'espace de longueur complet non géodésique et montre que l'hypothèse de locale compacité ne peut être simplement omise. Si l'on munit

$$E = \{(x, y) \in \mathbb{R}^2; y = (x - 1)(x + 1)/n, x \in [-1; 1], n \in \mathbb{N}^*\},$$

de la distance d_{L_d} où d est la distance euclidienne, alors (E, d_{L_d}) est un espace de longueur complet, qui n'est pas géodésique. On remarque, en effet, que les points $(-1, 0)$ et $(1, 0)$ n'ont pas de voisinage précompact.

Outre ce critère, il est possible de caractériser un espace géodésique via les points milieux.

Définition 1.22. Soit (E, d) un espace métrique. Un point $z \in E$ est un point milieu des points $x, y \in E$ si

$$d(x, z) = d(z, y) = \frac{1}{2}d(x, y).$$

Le théorème suivant fournit alors une définition alternative d'un espace géodésique.

Théorème 1.23. Soit (E, d) un espace métrique. E est un espace de longueur géodésique si et seulement si tout couple (x, y) de $E \times E$ possède un point milieu.

1.2.3 Barycentre

Pour définir aisément le barycentre, nous introduisons la notion de poids.

Définition 1.24. Soit $J \in \mathbb{N}^*$. On dit que $(\lambda_i)_{1 \leq i \leq J} \in (\mathbb{R}_+^*)^J$ sont des **poids** si

$$\sum_{1 \leq i \leq J} \lambda_i = 1.$$

Le barycentre d'un ensemble $\{x_i\}_{1 \leq i \leq J}$ de J points muni des poids $(\lambda_i)_{1 \leq i \leq J}$ de \mathbb{R}^d est défini par

$$\text{Bar}((\lambda_i)_{1 \leq i \leq J}, \{x_i\}_{1 \leq i \leq J}) = \sum_{1 \leq i \leq J} \lambda_i x_i.$$

Le barycentre a la propriété caractéristique d'être le minimiseur de

$$\sum_{1 \leq i \leq J} \lambda_i \|x - x_i\|^2,$$

parmi les $x \in \mathbb{R}^d$, où $\|\cdot\|$ désigne la norme euclidienne. Cette propriété caractéristique permet une définition du barycentre qui s'étend alors aux espaces métriques.

Définition 1.25 (p -barycentre). Soient (E, d) un espace métrique et $p > 0$. Soient $(\lambda_i)_{1 \leq i \leq J}$ des poids et $(x_i)_{1 \leq i \leq J} \in E^J$. Le p -barycentre de $(x_i)_{1 \leq i \leq J}$ pour les poids $(\lambda_i)_{1 \leq i \leq J}$ est défini comme l'ensemble des minimiseurs de

$$\sum_{1 \leq i \leq J} \lambda_i d(x, x_i)^p.$$

Tout élément de cet ensemble est aussi appelé un barycentre.

Dans la littérature, le barycentre est en général défini pour $p = 2$, alors que le cas $p = 1$ correspond aux médianes.

L'existence du barycentre n'est pas toujours assurée (il se peut que le minimum ne soit pas atteint). Un cadre qui convient pour leur existence est alors celui des espaces de longueur localement compacts. Ceux-ci possèdent la propriété suivante, qui permet de montrer l'existence des barycentres.

Proposition 1.26. Soit (E, d) un espace de longueur complet localement compact, alors toutes les boules fermées de E sont compactes.

Nous pouvons alors démontrer le théorème suivant.

Théorème 1.27. Soient (E, d) un espace de longueur complet localement compact, J un entier et $((\lambda_i)_{1 \leq i \leq J})$ des poids. Alors, il existe au moins un barycentre de tout J -uplet de E .

Démonstration. Soit $(z_n)_{n \geq 1}$ une suite minimisante de

$$A(z) = \sum_{1 \leq i \leq J} \lambda_i d(z, x_i)^p.$$

Alors $A(z_n)$ est bornée, et donc $r = \sup_{n \geq 1} \sup_{1 \leq i \leq J} d(z_n, x_i)$ est fini. Donc $(z_n)_{n \geq 1}$ est inclus dans l'union des boules centrées en x_i et de rayon r , qui est compacte. On peut donc en extraire une sous-suite qui converge vers un certain z , qui est un barycentre par continuité de A . \square

Les espaces de longueur dont nous allons étudier l'existence de barycentre en chapitre 4 sont les espaces de Wasserstein, qui ne sont pas localement compacts, comme nous allons le voir dans le paragraphe suivant.

1.3 Espaces de Wasserstein

Les espaces de Wasserstein sont issus de la théorie du transport optimal qui n'est pas le sujet de ce paragraphe. Les formes duales de l'expression de la distance de Wasserstein ne sont donc pas abordées dans cette thèse. Le livre [Villani, 2009] est une référence dans ce domaine. On pourra y trouver, notamment, les démonstrations des théorèmes énoncés dans ce paragraphe.

1.3.1 Définitions

L'ensemble des mesures de probabilité d'un espace métrique (E, d) peut être muni de différentes métriques, donnant lieu à différentes topologies. La topologie la plus classique est celle de la convergence faible, qui peut être métrisée par la distance de Prokhorov ou la distance bornée-Lipschitz. La distance de Wasserstein métrise la même topologie lorsque (E, d) est compact, mais diffère dans les autres cas.

Nous utiliserons la notation suivante : étant données J mesures $(\mu_i)_{1 \leq i \leq J}$, on note $\Gamma(\mu_1, \dots, \mu_J)$ l'ensemble des mesures définies sur la tribu borélienne de E^J , ayant pour i ème marginale μ_i , pour $1 \leq i \leq J$.

Définition 1.28 (Distance de Wasserstein). *Soient (E, d) un espace polonais (métrique, complet et séparable) et $p > 1$. Pour deux mesures μ et ν définies sur (E, d) , on note*

$$W_p(\mu, \nu) = \inf \left\{ \int d(x, y)^p d\pi(x, y); \pi \in \Gamma(\mu, \nu) \right\}^{1/p}. \quad (1.2)$$

W_p définit une métrique appelée **distance de Wasserstein** (d'ordre p) sur l'espace des mesures telles qu'il existe $x_0 \in E$ tel que

$$\int d(x_0, x)^p d\mu(x) < \infty.$$

Cet espace est appelé l'**espace de Wasserstein** associé à (E, d) et est noté $\mathcal{P}_p(E)$.

L'infimum de l'équation 1.2 est toujours atteint par une mesure π . Cette mesure est appelée **plan de transport optimal** entre μ et ν .

Remarque 1.29. Les distances de Wasserstein pour différentes valeurs de p sont reliées entre elles par la relation $W_p \leq W_q$ pour $p \leq q$. En effet, si l'on note Π_p et Π_q les plans de transport optimaux entre μ et ν pour respectivement p et q , l'inégalité de Jensen donne pour deux mesures μ et ν ,

$$\begin{aligned} W_p(\mu, \nu) &= \left(\int d(x, y)^p d\Pi_p(x, y) \right)^{1/p} \\ &\leq \left(\int d(x, y)^p d\Pi_q(x, y) \right)^{1/p} \\ &\leq \left(\int d(x, y)^q d\Pi_q(x, y) \right)^{1/q} \\ &= W_q(\mu, \nu). \end{aligned}$$

1.3.2 Propriétés

Un des premiers intérêts de l'espace de Wasserstein est sa topologie. La distance de Wasserstein métrise la topologie de la convergence faible sur les espaces bornés, et le théorème suivant s'applique pour les espaces non bornés.

Théorème 1.30 (Topologie de l'espace de Wasserstein). *La convergence en distance de Wasserstein d'ordre p est équivalente à la convergence faible et la convergence des moments d'ordre p .*

Remarque 1.31. Les moments d'ordre p d'une mesure μ sur un espace métrique (E, d) sont les valeurs $\int d(x_0, x)^p d\mu_n(x)$ pour tout $x_0 \in E$. Il est alors possible de montrer que les moments d'ordre p d'une suite de mesures convergent pour tout $x_0 \in E$ dès qu'ils convergent pour un x_0 .

L'espace (E, d) s'injecte de manière isométrique dans son espace de Wasserstein :

$$i : x \mapsto \delta_x,$$

où δ_x est la mesure de Dirac en $x \in E$. En plus de cette injection, l'espace de Wasserstein conserve certaines propriétés de son espace sous-jacent, notamment la complétude, la séparabilité, et la qualité d'être un espace de longueur ou géodésique.

Théoreme 1.32 (Espace métrique complet (séparable)). *Soit (E, d) un espace métrique complet (séparable). Alors l'espace de Wasserstein associé à (E, d) est un espace métrique complet (séparable).*

Théoreme 1.33 (Espace de longueur (géodésique)). *Soit (E, d) un espace de longueur (géodésique). Alors l'espace de Wasserstein associé à (E, d) est un espace de longueur (géodésique).*

Théoreme 1.34 (Espace métrique compact). *Soit (E, d) un espace métrique compact. Alors l'espace de Wasserstein associé à (E, d) est compact.*

Cependant, la propriété de locale compacité, qui assure notamment l'existence des barycentres n'est pas conservée, comme le montre l'exemple suivant - l'existence des barycentres est cependant assurée par ailleurs dans le cas d'un espace sous-jacent de longueur géodésique, comme nous le montrerons au théorème 4.4.

Exemple 1.35. *On considère l'espace de Wasserstein associé à \mathbb{R} . En posant alors $\mu = \delta_0$ et $\mu_n = \delta_0(1 - \varepsilon^p/n) + \delta_{n^{1/p}}\varepsilon^p/n$, de sorte que $W_p(\mu, \mu_n) = \varepsilon$, il vient $\mu_n \rightarrow \mu$ pour la convergence faible, mais clairement pas pour la topologie de la distance de Wasserstein, et donc aucune sous-suite de $(\mu_n)_{n \geq 1}$ ne peut converger en distance de Wasserstein, et aucune boule centrée en μ ne peut alors être compacte.*

1.4 Mesures gaussiennes sur un espace de Banach

Ce paragraphe présente les mesures gaussiennes, de leur construction (dont un rappel sur les notions d'intégrale sur les espaces de Banach) à quelques propriétés qui seront utilisées dans le chapitre 3.

Elles jouent un rôle central en probabilité, ne serait-ce que par le théorème central limite. Elles sont définies sur \mathbb{R}^d comme les mesures dont la mesure image par une forme linéaire sur \mathbb{R}^d est toujours une gaussienne sur \mathbb{R} . Cette définition peut-être étendue à un espace de Banach.

1.4.1 Définition

Pour un espace de Banach B , on note B' le dual topologique de B .

Définition 1.36. Soit $(B, \|\cdot\|)$ un espace de Banach. Une mesure μ est dite gaussienne (centrée) si pour toute forme linéaire continue $f \in B'$, la mesure image $f_{\#}\mu$ est une mesure gaussienne centrée sur \mathbb{R} .

Dans cette définition, la mesure δ_0 de Dirac en 0 est considérée comme une mesure gaussienne centrée (de variance nulle).

On supposera toujours que ces mesures gaussiennes sont centrées.

Les espaces de Banach de dimension infinie sont souvent plus facilement abordés par leur espace dual comme c'est le cas dans cette définition. Les propriétés présentées dans ce chapitre utilisent aussi cet espace dual, notamment dans la définition de l'intégrale de Pettis, un outil qui permet de définir les espaces à noyau auto-reproduisant.

1.4.2 Intégrales à valeurs dans un espace de Banach

Ce paragraphe s'inspire du cours de Charles Suquet [Suquet, 2006].

L'intégration par rapport à une mesure est une opération qui associe à une fonction f à valeurs réelles, une valeur dans \mathbb{R} . Une application à valeurs dans \mathbb{R}^d se présente sous la forme (f_1, \dots, f_d) où chaque f_i est une fonction. Intégrer une application à valeurs dans \mathbb{R}^d revient alors à intégrer chaque composante f_i et à former le vecteur composé de ces intégrales. En revanche, la tâche devient plus complexe pour intégrer une application à valeurs dans un espace de Banach de dimension infinie.

Il existe deux types d'intégrales à valeurs dans un espace de Banach. Lorsque l'espace est de dimension finie, elles coïncident toutes les deux avec la notion d'intégrale à valeurs dans \mathbb{R}^d présentée ci-dessus.

Le première intégrale que l'on va présenter est celle de Pettis, appelée parfois intégrale au sens faible. Elle s'appuie sur l'espace dual, et se définit à partir de l'intégrale de chaque forme linéaire continue.

Définition 1.37 (Intégrale de Pettis). Soient (Ω, μ) un espace mesuré et $(B, \|\cdot\|)$ un espace de Banach. Soit X une application mesurable de Ω dans B (pour B muni de la topologie engendrée par la topologie faible $\sigma(B, B')$) telle que pour toute forme linéaire continue $f \in B'$,

$$\int |f(X)| d\mu < \infty.$$

Alors, s'il existe $y \in B$ tel que pour tout $f \in B'$

$$f(y) = \int f(X) d\mu,$$

on dit que X est μ -Pettis-intégrable, et

$$\int X d\mu = y.$$

Il est facile de vérifier que l'intégration de Pettis est une opération linéaire.

Remarque 1.38. *On peut montrer que l'application qui associe à une forme linéaire continue $f \in B'$, le réel $\int f(X)d\mu$ est un élément du bidual (la continuité est une conséquence du théorème du graphe fermé). L'intégrale de Pettis est alors définie dès que cet élément est dans l'image de l'injection canonique de B dans son bidual.*

La seconde intégrale que l'on présente est celle de Bochner, dite intégrale au sens fort.

La construction de l'intégrale de Bochner suit celle de l'intégrale de Lebesgue. Une application $X : (\Omega, \mu) \rightarrow B$ est dite *simple* si l'on peut l'écrire sous la forme

$$X = \sum_{1 \leq i \leq n} x_i \mathbf{1}_{A_i},$$

où $x_i \in B$ et A_i mesurable. On définit alors naturellement l'intégrale de Bochner d'une telle fonction par

$$\int X d\mu = \sum_{1 \leq i \leq n} \mu(A_i) x_i.$$

On peut ensuite définir l'intégrale de Bochner plus généralement.

Définition 1.39 (Intégrale de Bochner). *Soient (Ω, μ) un espace mesuré et $(B, \|\cdot\|)$ un espace de Banach. Soit X une application mesurable de Ω dans B (pour B muni de la tribu borélienne engendrée par la norme $\|\cdot\|$). S'il existe une suite de fonctions simples $(X_n)_{n \geq 1}$ telle que*

$$X_n \rightarrow X, \text{ p.s. et } \int \|X_n - X\| d\mu \rightarrow 0,$$

alors X est dite Bochner intégrable et

$$\int X d\mu = \lim_n \int X_n d\mu.$$

Dans les espaces de Banach séparables, il existe une caractérisation concise de la Bochner intégrabilité d'une application mesurable.

Théoreme 1.40. *Soient (Ω, μ) un espace mesuré et $(B, \|\cdot\|)$ un espace de Banach séparable. Soit X une application mesurable de Ω dans B (pour B muni de la topologie engendrée par la norme $\|\cdot\|$). Alors X est μ -Bochner intégrable si et seulement si*

$$\int \|X\| d\mu < \infty.$$

Enfin, la proposition suivante permet de comparer les deux intégrales.

Proposition 1.41. *Soient (Ω, μ) un espace mesuré et $(B, \|\cdot\|)$ un espace de Banach séparable. Soit X une application mesurable de Ω dans B (pour B muni de la tribu borélienne engendrée par la norme $\|\cdot\|$). Si X est Bochner intégrable alors elle est Pettis intégrable, et les deux intégrales sont égales.*

Remarque 1.42. *La réciproque est fausse.*

1.4.3 Espaces à noyau auto-reproduisant

Sur un espace de Banach séparable muni d'une mesure gaussienne, il existe une construction d'un sous-espace de Hilbert qui caractérise la mesure et possède des propriétés intéressantes. La construction de cet espace présentée ici est reprise de [Ledoux et Talagrand, 1991].

Pour $(B, \|\cdot\|, \mu)$ un espace de Banach séparable mesuré gaussien (autrement dit, μ est gaussienne),

$$\sigma = \sup_{f \in B', \|f\| \leq 1} (E(f^2))^{1/2}$$

est fini. Pour s'en convaincre, il suffit de remarquer que l'opérateur $A : B' \rightarrow L^2(B, \mu)$ défini par $Af = f$ est continu par le théorème du graphe fermé (rappelé en A.3) et $\|A\| = \sigma$.

Ensuite, nous voulons montrer que l'image de l'opérateur adjoint $A^* : L^2(B, \mu) \rightarrow B''$ est incluse dans B . Pour cela, nous choisissons d'abord une suite de compacts $(K_n)_{n \geq 1}$ telle que $\mu(K_n) \rightarrow 1$ (son existence est assurée par la tension de μ , et la tension de μ est assurée par la séparabilité de $(B, \|\cdot\|)$). Ensuite, pour $\xi \in L^2(B, \mu)$,

$$A^*(\xi \mathbf{1}_{K_n}) = \left(f \mapsto \int f \xi \mathbf{1}_{K_n} d\mu \right) = \int x \xi(x) \mathbf{1}_{K_n}(x) d\mu(x).$$

L'intégrale est définie au sens de l'intégrale de Bochner, d'après le théorème 1.40 et le fait que

$$\int \|x\| |\xi(x)| \mathbf{1}_{K_n}(x) d\mu(x) \leq \sup_{x \in K_n} \|x\| |\xi(x)| < \infty,$$

et donc, $\int x \xi(x) \mathbf{1}_{K_n}(x) d\mu(x) \in B$. Comme le précise la remarque 1.38, on peut définir $\int x \xi(x) d\mu(x)$ comme un élément du bidual. L'inclusion canonique $x \in B \mapsto (f \mapsto f(x))$ de B dans B'' est une isométrie, et donc

$$\left\| \int x \xi(x) d\mu(x) - \int x \xi(x) \mathbf{1}_{K_n}(x) d\mu(x) \right\|_{B''} = \sup_{\|f\| \leq 1} f \left(\int x \xi(x) \mathbf{1}_{K_n^c}(x) d\mu(x) \right).$$

Alors, la convergence

$$\sup_{\|f\| \leq 1} f \left(\int x\xi(x) \mathbf{1}_{K_n^c}(x) d\mu(x) \right) \leq \sigma \left(\int \xi(x)^2 \mathbf{1}_{K_n^c}(x) d\mu(x) \right)^{1/2} \rightarrow 0,$$

assure $\int x\xi(x) d\mu(x) \in B$, puisque B est fermé dans B'' et $A^*(\xi) = \int x\xi(x) d\mu(x) \in B$ par continuité de l'opérateur adjoint. On a donc montré que l'image de A^* est bien incluse dans B .

On peut alors définir un produit scalaire sur $H = \text{Im}(A^*)$. Pour $\xi, \zeta \in L^2(B, \mu)$, on pose

$$\langle A^*\xi, A^*\zeta \rangle_H = \int \xi\zeta d\mu.$$

Il est clair que le produit scalaire ainsi défini sur H , ne dépend pas du choix de ξ et ζ dans $A^{*-1}(A^*\xi)$ et $A^{*-1}(A^*\zeta)$.

Définition 1.43 (Espace à noyau auto-reproduisant). *($H, \langle \cdot, \cdot \rangle_H$) ainsi défini est un espace de Hilbert appelé espace à noyau auto-reproduisant associé à $(B, \|\cdot\|, \mu)$.*

Exemple 1.44 (Espace euclidien). *Si (E, μ) est un espace euclidien de dimension finie mesuré par une mesure gaussienne non dégénérée, de matrice de covariance Σ , son espace à noyau auto-reproduisant associé est le même espace euclidien muni du produit scalaire défini par*

$$\langle a, b \rangle_H = \langle \Sigma^{-1}a, \Sigma^{-1}b \rangle,$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire sur E .

Exemple 1.45 (Espace de Wiener). *Lorsque (E, μ) est l'espace des fonctions continues sur $[0, 1]$ muni de la mesure de Wiener, l'espace à noyau auto-reproduisant associé est l'espace de Cameron-Martin.*

La construction de cet espace est valable pour une mesure non-gaussienne, mais les mesures gaussiennes (centrées) étant entièrement définies par leur structure de covariance, cet espace présente un intérêt particulier pour ces mesures. Les mesures gaussiennes vérifient $\int \|x\|^2 d\mu(x) < \infty$, la proposition suivante permet donc d'assurer la compacité de la boule unité de leur espace à noyau auto-reproduisant.

Proposition 1.46. *Soit $(B, \|\cdot\|, \mu)$ un espace de Banach séparable mesuré par une mesure μ . Si*

$$\int \|x\|^2 d\mu(x) < \infty,$$

alors la boule unité de l'espace à noyau auto-reproduisant est compacte dans $(B, \|\cdot\|)$.

Pour le cas des mesures gaussiennes, il existe un résultat plus précis, démontré dans [Kuelbs et Li, 1993].

Lemme 1.47 (Kuelbs et Li). *Soient $(B, \|\cdot\|, \mu)$ un espace de Banach séparable mesuré par une mesure μ gaussienne. On note \mathcal{K} la boule unité de l'espace à noyau auto-reproduisant associé. On pose $\varphi(t) := \mu(B(0, t))$ pour tout $t > 0$. Alors, pour tout $\lambda > 0$ et $\varepsilon > 0$,*

$$\log \mathcal{N}(\lambda \mathcal{K}, 2\varepsilon) \leq \lambda^2/2 - \log \varphi(\varepsilon).$$

En posant $\psi(t) = \frac{\sqrt{2}t}{\sqrt{-\log \varphi(t)}}$, il vient alors

$$\mathcal{N}(\lambda \mathcal{K}, \varepsilon) \leq \varphi(\psi^{-1}(\varepsilon/\lambda))^{-2}.$$

Démonstration. L'article [Kuelbs et Li, 1993] donne une démonstration de la première inégalité. Pour la seconde inégalité, on remarque d'abord que $\mathcal{N}(\lambda \mathcal{K}, \varepsilon) = \mathcal{N}(\mathcal{K}, \varepsilon/\lambda)$. En posant, $\lambda^2/2 = -\log \varphi(\varepsilon)$, il vient alors,

$$\begin{aligned} \mathcal{N}(\mathcal{K}, \psi(\varepsilon)) &= \mathcal{N}(\mathcal{K}, 2\varepsilon/\lambda) \\ &= \mathcal{N}(\lambda \mathcal{K}, 2\varepsilon) \\ &\leq \exp(\lambda^2/2 - \log \varphi(\varepsilon)) \\ &= \varphi(\varepsilon)^{-2}. \end{aligned}$$

Choisir $\varepsilon = \psi^{-1}(t)$ permet alors de conclure la démonstration. □

Remarque 1.48. *Il est possible de raffiner le résultat en choisissant $\lambda^2/2 = -c \log \varphi(\varepsilon)$, pour $c > 0$, de sorte que pour $\psi_c(t) = \frac{\sqrt{2}t}{\sqrt{-c \log \varphi(t)}}$, la majoration devienne*

$$\mathcal{N}(\lambda \mathcal{K}, \varepsilon) \leq \varphi(\psi_c^{-1}(t))^{-1-c}.$$

Cet espace à noyau auto-reproduisant permet aussi d'étendre l'inégalité isopérimétrique des mesures gaussiennes sur \mathbb{R}^d à celles sur les espaces de Banach. Ce résultat est démontré dans [Borell, 1975].

Théoreme 1.49 (Inégalité isopérimétrique). *Soit $(B, \|\cdot\|, \mu)$ un espace de Banach séparable mesuré par une mesure μ gaussienne. On note \mathcal{K} la boule unité de l'espace à noyau auto reproduisant associé et Φ la fonction de répartition de la mesure gaussienne standard sur \mathbb{R} . Alors, pour tout borélien A , et $\varepsilon > 0$,*

$$\mu(A + \varepsilon \mathcal{K}) \geq \Phi(\Phi^{-1}(\mu(A)) + \varepsilon).$$

1.4.4 Inégalités de transport

Dans le chapitre 3, nous utiliserons d'autres inégalités vérifiées par les mesures gaussiennes, qui sont en lien avec la concentration des mesures gaussiennes. Ces liens sont développés dans [Ledoux et Talagrand, 1991]. L'article [Gozlan et Léonard, 2007] est un excellent résumé dont sont extraits certains des résultats présentés ici.

Pour présenter les inégalités de transport, il nous faut d'abord définir l'entropie relative.

Définition 1.50 (Entropie relative ou distance de Kullback-Leibler). *Soient μ et ν deux mesures sur un espace polonais (E, d) . L'entropie relative de ν par rapport à μ est définie par*

$$H(\nu|\mu) = \begin{cases} \int \log \left(\frac{d\nu}{d\mu} \right) d\nu, & \text{si } \nu \ll \mu, \\ +\infty & \text{sinon.} \end{cases}$$

Une mesure vérifie alors une inégalité de transport si le carré de la distance de Wasserstein est contrôlé par l'entropie métrique.

Définition 1.51 (Inégalités de transport). *Soit (E, d, μ) un espace polonais mesuré. On dit que μ vérifie une inégalité de transport $T_p(C)$, pour $1 \leq p \leq 2$, $C > 0$, si*

$$W_p^2(\nu, \mu) \leq CH(\nu|\mu),$$

pour toute mesure ν sur (E, d) .

Il existe un théorème qui permet de mieux comprendre les conséquences d'une inégalité de transport $T_p(C)$.

Théorème 1.52 (Caractérisation de $T_p(C)$). *Soit (E, d, μ) un espace polonais mesuré. Alors les deux assertions suivantes sont équivalentes.*

1. *il existe $C > 0$ tel que μ vérifie $T_p(C)$*
2. *il existe $a > 0$ tel que*

$$\int e^{ad(x,y)^2} d\mu(x) d\mu(y) < \infty.$$

Une conséquence des inégalités de transport qui va nous intéresser est celle du contrôle de la déviation de la distance de Wasserstein d'une mesure à sa version empirique démontré dans [Gozlan et Léonard, 2007].

Théoreme 1.53. Soit (E, d, μ) un espace de espace polonais mesuré. On suppose que μ vérifie $T_p(C)$, pour un $C > 0$. On note

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

la mesure empirique associée à une suite i.i.d. de variables aléatoires $(X_i)_{i \geq 1}$ de loi commune μ . Alors, pour tout $t > 0$

$$P(W_p(\mu, \mu_n) \geq EW_p(\mu, \mu_n) + t) \leq e^{-nt^2/C}.$$

Nous utiliserons ce résultat pour les mesures gaussiennes, car celles-ci vérifient une inégalité de transport.

Théoreme 1.54. Soit $(B, \|\cdot\|, \mu)$ un espace de Banach séparable mesuré par une mesure μ gaussienne. On note

$$\sigma = \sup_{f \in B', \|f\| \leq 1} (E(f^2))^{1/2} < \infty.$$

Alors μ vérifie une inégalité $T_2(2\sigma^2)$.

À l'aide de la remarque 1.29, il est clair que les mesures gaussiennes vérifient alors les inégalités $T_p(2\sigma^2)$, pour $1 \leq p \leq 2$.

Chapitre 2

Localisation de masse

Ce chapitre présente des résultats originaux sur une nouvelle notion dont le but est de localiser les zones concentrant le plus de masse d'une mesure donnée, dans un espace métrique.

Une première partie présente les notions et le problème considérés, puis deux parties successives énoncent les résultats obtenus. Une quatrième partie présente des applications statistiques, puis viennent les démonstrations dans une dernière partie.

2.1 Présentation du problème

La localisation de masse d'une mesure (de probabilité) μ consiste à retirer une proportion α du support de la mesure, afin d'obtenir une nouvelle mesure, dite α -localisée. Une définition formelle est donc le conditionnement d'une mesure par un borélien de masse $1 - \alpha$. Ce chapitre propose une définition du *plus petit* borélien de masse $1 - \alpha$ afin d'obtenir une version localisée avec un support le *plus petit* possible.

Ce *petit* borélien représente intuitivement la partie « essentielle » de la mesure. Cependant, il semble difficile d'obtenir un choix satisfaisant de ce *plus petit* borélien pour toutes les mesures : alors qu'il semble évident qu'un choix idéal pour la mesure gaussienne standard sur \mathbb{R}^d est celui d'une boule centrée sur l'origine, il n'est pas évident de comprendre ce que serait un *plus petit* borélien lorsque la mesure n'est pas uni-modale, ou lorsqu'elle n'est pas symétrique par rapport à l'origine, ou encore, lorsqu'elle n'est pas définie sur un espace euclidien.

Une propriété essentielle que l'on souhaite retrouver dans une méthode pour choisir cet ensemble est sa consistance. Lorsque la mesure μ , souvent inconnue, est approchée par une suite de mesures $(\mu_n)_{n \geq 1}$, le *plus petit* borélien pour μ_n doit se rapprocher de celui de μ . Plusieurs méthodes ont été proposées pour choisir ce

plus petit borélien.

Une première méthode consiste à choisir une classe d'ensembles \mathcal{F} partiellement ordonnés par leur aire, et à choisir un des plus petits ensembles (pour la relation d'ordre sur les aires) de cette classe avec une masse supérieure à $1 - \alpha$. Cet ensemble correspond aux ensembles de niveau de la densité f lorsque μ est régulière et que la classe \mathcal{F} contient les ensembles de niveau. Une autre façon de définir ces ensembles consiste à choisir un maximiseur B_β de

$$\mu(B) - \beta\lambda(B),$$

sur $B \in \mathcal{F}$, où λ désigne la mesure de Lebesgue et $\mu(\{f \geq \beta\}) = 1 - \alpha$, et de même, on définit B_β^n comme le maximiseur de

$$\mu_n(B) - \beta\lambda(B),$$

où μ_n désigne la mesure empirique. On parle alors d'*excès de masse*. L'article [Hartigan, 1987] se place dans le cadre où \mathcal{F} est l'ensemble des convexes et démontre qu'en dimension 2, la distance de Hausdorff $d_H(B_n, B)$ entre B_β^n et B_β converge vers 0 et vérifie

$$d_H(B_\beta^n, B_\beta) = O\left(\frac{\log n}{n}\right)^{2/7}.$$

L'article [Nolan, 1991] s'intéresse au cas où la classe \mathcal{F} correspond aux ellipsoïdes. La consistance de B_β^n y est démontrée, ainsi qu'un théorème limite que l'on peut énoncer comme suit. Si l'on note c_n et c les centres des ellipsoïdes B_β^n et B_β respectivement, σ_n et σ un vecteur contenant les entrées de la matrice permettant de définir B_β^n et B_β respectivement, alors, si les lignes de niveau de μ sont des ellipsoïdes,

$$n^{1/3}(c_n - c, \sigma_n - \sigma)$$

converge en loi vers le maximum d'un processus gaussien. [Polonik, 1997] étudie un cadre plus général et montre la consistance de B_β^n pour la pseudo-distance

$$d_\mu(F, G) = \mu(F \triangle G),$$

où \triangle désigne la différence symétrique, lorsque la classe \mathcal{F} considérée est une classe de Glivenko-Cantelli. Sous plusieurs hypothèses incluant notamment que les ensembles de niveau de la mesure μ appartiennent à \mathcal{F} et des conditions de régularité sur μ , il obtient la vitesse de convergence

$$d_\mu(B_\beta, B_\beta^n) = O(n^{-\delta}),$$

où δ est une constante qui dépend de la régularité de μ . Cette approche par excès de masse permet donc d'obtenir des résultats de convergence précis dans de nombreux

cas. Elle présente cependant quelques inconvénients. Le fait de devoir choisir une classe \mathcal{F} contenant les ensembles de niveau demande un a priori important sur la mesure. Il est donc aussi nécessaire que la mesure μ soit suffisamment régulière pour définir les ensembles de niveau (il faut que μ soit absolument continue par rapport à la mesure de Lebesgue). Enfin, cette approche se limite aux cas où l'espace E est de dimension finie (et même $E = \mathbb{R}^d$ dans la plupart des cas).

Une seconde méthode s'inspire de la notion de *trimming* sur \mathbb{R} et l'étend au cas multivarié. Sur \mathbb{R} , il s'agit d'approcher l'intervalle $[F^{-1}(\alpha/2); F^{-1}(1 - \alpha/2)]$, où F désigne la fonction de répartition de la mesure μ par le même intervalle où F est remplacé par F_n , la fonction de répartition de la mesure empirique. Le passage au cas multivarié se fait de la manière suivante : C_α désigne l'intersection de tous les demi-espaces fermés de μ -mesure supérieure à $1 - \alpha$. C_α est alors un ensemble convexe non vide lorsque $\alpha < 1/2$ et que la mesure μ est suffisamment régulière. [Nolan, 1992] s'intéresse alors à la convergence de C_n , défini de la même façon à l'aide de la mesure empirique μ_n et montre la consistance de celui-ci. Afin de quantifier la vitesse de convergence de cet ensemble C_n vers C_α , l'article introduit la fonction aléatoire

$$r_n(u) = \inf\{r \geq 0; ru \notin C_n\},$$

et de même

$$r_\alpha(u) = \inf\{r \geq 0; ru \notin C_\alpha\},$$

et établit la convergence en loi vers un processus gaussien défini sur la sphère unité \mathbb{S}^{d-1} du processus

$$\sqrt{n}(r_n - r_\alpha),$$

sous des hypothèses de régularité de la densité de μ .

[Cuesta-Albertos et al., 1997] présente une méthode, dite de k -means α -trimmé, qui n'introduit que très peu d'arbitraire. Elle consiste à choisir le support de la mesure ν α -localisée de μ avec la plus petite distorsion à son meilleur k -quantifieur. Formellement, pour une certaine fonction Φ , et un certain entier k , la méthode consiste à choisir

$$B_\alpha \in \arg \min \left\{ \inf_{\{m_1, \dots, m_k\} \subset \mathbb{R}^d} \int_B \Phi \left(\inf_{1 \leq i \leq k} \|X - m_i\| \right) d\mu; \mu(B) \geq 1 - \alpha \right\}.$$

Après avoir montré l'existence d'un tel minimiseur, [Cuesta-Albertos et al., 1997] montre la consistance de B_α : si $(\mu_n)_{n \geq 1}$ converge faiblement vers une mesure μ absolument continue par rapport à la mesure de Lebesgue, alors, pour n'importe quel choix de

$$B_\alpha^n \in \arg \min \left\{ \inf_{\{m_1, \dots, m_k\} \subset \mathbb{R}^d} \int_B \Phi \left(\inf_{1 \leq i \leq k} \|X - m_i\| \right) d\mu_n; \mu_n(B) \geq 1 - \alpha \right\},$$

la suite $(B_\alpha^n)_{n \geq 1}$ converge vers B_α (lorsque celui-ci est unique) en distance de Hausdorff. Ces résultats sont valables sur \mathbb{R}^d .

Toutes ces techniques présentent certains inconvénients, que nous allons chercher à atténuer à travers une nouvelle méthode.

2.1.1 Définitions

Ce chapitre définit une notion de *plus petit* borélien et en présente des propriétés. Le cadre qui se veut le plus général possible est le suivant. (E, d) est un espace polonais (métrique, complet et séparable) et μ est une mesure borélienne de probabilité sur (E, d) . Un *plus petit* borélien sera défini comme minimiseur d'une fonction τ définie sur une classe \mathcal{F} de sous-ensembles de E et ne dépendant de ces sous-ensembles qu'à travers leur *packing*. Pour comprendre cette notion, on introduit les deux définitions suivantes.

2.1.1.a Ensemble stable

La classe d'ensembles que l'on va considérer ne peut pas être quelconque si l'on veut que le *plus petit* borélien existe. Il est nécessaire d'imposer une certaine **stabilité** de cette classe.

Pour un ensemble B et un $\varepsilon > 0$, on introduit la notation B^ε pour le ε -**épaississement** de B :

$$B^\varepsilon := \{x \in E; \exists y \in B, d(x, y) < \varepsilon\}.$$

Définition 2.1 (Ensemble stable). *Pour une suite d'ensembles fermés $(B_n)_{n \geq 1}$, on note $\lim_n B_n$ l'ensemble*

$$\lim_n B_n := \bigcap_{\varepsilon > 0} \bigcup_{k \geq 1} \bigcap_{n \geq k} B_n^\varepsilon.$$

*Soit \mathcal{F} une classe d'ensembles fermés de E . \mathcal{F} est dite **stable** si $E \in \mathcal{F}$ et*

$$(B_n)_{n \geq 1} \subset \mathcal{F} \implies \exists (n_k)_{k \geq 1}, n_k \rightarrow \infty, \lim_k B_{n_k} \in \mathcal{F}.$$

Cette notion de stabilité est proche de celle de la stabilité par convergence de Hausdorff. Elle est équivalente dans le cas des espaces compacts, comme nous allons le voir dans les remarques suivantes.

Mais d'abord, comme nous avons supposé qu'une classe stable \mathcal{F} est un sous-ensemble des fermés, vérifions que cette notion de stabilité conserve bien la propriété d'être un fermé.

Remarque 2.2. Étant donnée une suite d'ensembles $(B_n)_{n \geq 1}$, $\lim_n B_n$ est toujours un fermé. En effet, en notant $B(x, \varepsilon/2)$ la boule de centre x et de rayon $\varepsilon/2$,

$$x \notin \lim_n B_n \Leftrightarrow \exists \varepsilon > 0, \forall k \geq 1, \exists n \geq k, x \notin B_n^\varepsilon \quad (2.1)$$

$$\Rightarrow \exists \varepsilon > 0, \forall k \geq 1, \exists n \geq k, B(x, \varepsilon/2) \cap B_n^{\varepsilon/2} = \emptyset \quad (2.2)$$

$$\Rightarrow \exists \varepsilon > 0, B(x, \varepsilon/2) \cap \lim_n B_n = \emptyset. \quad (2.3)$$

Autrement dit, $(\lim_n B_n)^c$ est un ouvert, et $\lim_n B_n$ est donc fermé.

Pour bien comprendre ce qu'est une classe stable, nous lions cette notion à la distance de Hausdorff dans les deux remarques suivantes.

Remarque 2.3. Lorsque $(B_n)_{n \geq 1}$ est une suite d'ensembles qui converge en distance de Hausdorff vers un ensemble limite B_∞ , alors

$$\lim_n B_n = B_\infty.$$

En effet, si l'on note ε_k le plus petit $\varepsilon > 0$ tel que $B_n \subset B_\infty^\varepsilon$ et $B_\infty \subset B_n^\varepsilon$ pour tout $n \geq k$, alors,

$$B_\infty = \bigcap_{\varepsilon > 0} \bigcup_{k \geq 1} \bigcap_{n \geq k} B_\infty \subset \bigcap_{\varepsilon > 0} \bigcup_{k \geq 1} \bigcap_{n \geq k} B_n^{\varepsilon_k} = \lim_n B_n \subset \bigcap_{\varepsilon > 0} \bigcup_{k \geq 1} \bigcap_{n \geq k} B_\infty^{\varepsilon + \varepsilon_k} = B_\infty.$$

D'une manière plus générale, pour une suite $(B_n)_{n \geq 1}$, s'il existe un ensemble B_∞ tel que pour tout compact K , la suite $(B_n \cap K)_{n \geq 1}$ converge en distance de Hausdorff vers $B_\infty \cap K$, alors

$$\lim_n B_n = B_\infty.$$

On parlera alors de **convergence de Hausdorff généralisée**. Pour s'en convaincre, il suffit de remarquer que $\lim_n (B_n \cap K) = (\lim_n B_n) \cap K$.

Remarque 2.4. Dans le cas d'un espace métrique (E, d) totalement compact (i.e. tel que toutes les boules fermées sont compactes, c'est par exemple le cas d'un espace de longueur localement compact), la notion de convergence de Hausdorff généralisée se comprend mieux. Pour toute suite $(B_n)_{n \geq 1}$ de fermés de E , il existe une sous-suite (renommée $(B_n)_{n \geq 1}$) qui admet une limite B_∞ pour la convergence de Hausdorff généralisée, et

$$\lim_n B_n = B_\infty.$$

Dans ce cas, une classe stable au sens de la définition 2.1 est en fait une classe stable par limite de Hausdorff généralisée.

En effet, dans un espace totalement compact, il existe une suite croissante de compacts $(K_k)_{k \geq 1}$ telle que $\cup_{k \geq 1} K_k = E$, par exemple une suite croissante de boules. La convergence de Hausdorff généralisée est alors équivalente à la convergence de $(B_n \cap K_k)_{n \geq 1}$ pour tout $k \in \mathbb{N}$. Par compacité de l'ensemble des fermés de K_k pour la distance de Hausdorff, il existe une sous-suite de $(B_n \cap K_k)_{n \geq 1}$ qui converge vers un certain B_∞^k . Par un argument diagonal, on peut extraire une sous-suite de $(B_n)_{n \geq 1}$ telle que pour tout $k \in \mathbb{N}$, $(B_n \cap K_k)_{n \geq 1}$ converge en distance de Hausdorff vers un B_∞^k . Il est évident que les B_∞^k sont croissants en k , et donc que $B_\infty := \cup_k B_\infty^k$ est bien la limite de la sous-suite construite de $(B_n)_{n \geq 1}$ au sens de la convergence généralisée de Hausdorff.

Pour voir que le sens d'ensemble stable n'est pas vide, nous présentons une série d'exemples d'ensemble stable.

Exemple 2.5. *Le premier exemple est trivial, c'est celui de l'ensemble de tous les fermés. La stabilité est alors évidente puisque toute limite pour la convergence de Hausdorff généralisée est fermée, d'après la remarque 2.2.*

Exemple 2.6. *L'ensemble des boules n'est pas toujours stable, mais le plus petit ensemble stable qui contient toutes les boules de \mathbb{R}^d est l'ensemble des boules et des demi-espaces. Ce n'est généralement pas l'ensemble des fermés. Une démonstration de cette affirmation passe par une paramétrisation des centres des boules en coordonnées sphériques et utilise la compacité de la sphère.*

Exemple 2.7. *D'autres formes d'ensembles de \mathbb{R}^d forment des ensembles stables. Les ellipsoïdes, les rectangles, les convexes de diamètre majoré par $R < \infty$ sont encore des ensembles stables. Il est possible de s'affranchir de la majoration du diamètre, mais il faut alors agrandir l'ensemble pour qu'il devienne stable.*

Exemple 2.8. *Si \mathcal{F} est un ensemble stable d'ensembles connexes d'un espace métrique (E, d) totalement compact, tels que leur intersection à toute boule est encore connexe, alors*

$$\mathcal{F}_\varepsilon := \{\cup_{F \in \mathcal{G}} F; \mathcal{G} \subset \mathcal{F}, \forall F, G \in \mathcal{G} \inf_{x \in F, y \in G} d(x, y) \geq \varepsilon\},$$

est aussi un ensemble stable (voir lemme A.2).

2.1.1.b Fonction de taille

La fonction τ - correspondant intuitivement à la *taille* d'un ensemble - que l'on cherche à minimiser pour localiser la masse d'une mesure, est supposée dépendre du packing de l'ensemble considéré. Notons alors que deux ensembles isométriques auront la même taille. Nous appellerons τ la **fonction de taille**. Elle est définie

sur tous les ensembles, mais pour tout ensemble B , $\tau(B) = \tau(\bar{B})$ (où \bar{B} désigne la fermeture topologique de B). Nous aurons donc tendance à considérer seulement la taille des ensembles fermés.

Définissons le **contraste de Hausdorff** pour pouvoir énoncer simplement les hypothèses qui seront faites sur τ .

Définition 2.9 (Contraste de Hausdorff). *Soient A et B deux sous-ensembles fermés d'un espace polonais (E, d) . Le **contraste de Hausdorff** entre A et B est défini par*

$$\text{Haus}(A|B) := \inf\{\varepsilon > 0 | A \subset B^\varepsilon\}.$$

Nous pouvons maintenant définir formellement une fonction de taille.

Définition 2.10 (Fonction de taille). *Soit (E, d) un espace métrique. Une fonction $\tau : \mathcal{P}(E) \rightarrow \mathbb{R}^+$ est appelée **fonction de taille** si elle est de la forme*

$$\tau(B) = \Phi(\mathcal{M}(B, \cdot))$$

où Φ est une fonction de l'ensemble des fonctions décroissantes de \mathbb{R}^+ à valeurs dans \mathbb{N} et vérifie les trois hypothèses suivantes :

- (H1) τ est croissante, i.e. $A \subset B \implies \tau(A) \leq \tau(B)$,
- (H2) pour toute suite de fermés $(A_n)_{n \geq 1}$ décroissante telle que $\tau(A_1) < \infty$ et $\text{Haus}(A_n | \cap_k A_k) \rightarrow 0$, la convergence suivante a lieu $\tau(A_k) \rightarrow \tau(\cap_k A_k)$,
- (H3) Φ satisfait $\Phi(\liminf_{n \rightarrow \infty} f_n) \leq \liminf_{n \rightarrow \infty} \Phi(f_n)$ pour toute suite $(f_n)_{n \geq 1}$ de fonctions décroissantes de \mathbb{R}^+ à valeurs dans \mathbb{N} .

Une fonction de taille particulière est à l'origine de la formulation de ces hypothèses, qui ont permis de définir une notion plus générale de fonction de taille. C'est, cependant, l'exemple que nous garderons en tête.

Exemple 2.11. *L'exemple canonique de fonction de taille τ est*

$$\tau(B) = \int \varphi(\mathcal{M}(B, t), t) dt,$$

pour une fonction φ croissante en sa première variable et continue. L'hypothèse (H1) est alors assurée par la croissance de φ . La continuité de φ assure les hypothèses (H2) et (H3) par le théorème de convergence dominée de Lebesgue et le lemme de Fatou.

En effet, pour une suite $(A_n)_{n \geq 1}$ telle que dans (H2) ,

$$\mathcal{M}(A_n, t) \xrightarrow{n \rightarrow \infty} \mathcal{M}(\cap_n A_n, t), \tag{2.4}$$

pour tout point t de continuité de $t \mapsto \mathcal{M}(\cap_n A_n, t)$. Pour s'en convaincre, il suffit de remarquer que, pour $\varepsilon > 0$ et n suffisamment grand, un ensemble t -séparé $\{x_1, \dots, x_k\} \subset A_n \subset (\cap_n A_n)^{\varepsilon/2}$ permet de construire $\{y_1, \dots, y_k\} \subset \cap_n A_n$ tel que pour tout $1 \leq i \leq k$, $d(x_i, y_i) < \varepsilon/2$. L'ensemble $\{y_1, \dots, y_k\}$ est alors $t - \varepsilon$ séparé, de sorte que pour tout $\varepsilon > 0$ et n suffisamment grand,

$$\mathcal{M}(A_n, t) \leq \mathcal{M}(\cap_n A_n, t - \varepsilon),$$

ce qui montre bien (2.4), l'autre inégalité étant évidente. Les points de discontinuité étant au plus dénombrables, le théorème de convergence dominée de Lebesgue s'applique par domination de $\mathcal{M}(A_n, \cdot)$ par $\mathcal{M}(A_1, \cdot)$ et grâce à (H1), et il vient alors

$$\lim_n \int \varphi(\mathcal{M}(A_n, t), t) dt = \int \lim_n \varphi(\mathcal{M}(A_n, t), t) dt = \int \varphi(\mathcal{M}(\cap_n A_n, t), t) dt.$$

De même, pour tout $n \geq 1$ et tout $t > 0$,

$$\mathcal{M}(\cap_k A_k, t) \leq \liminf_n \mathcal{M}(A_n, t).$$

Le lemme de Fatou ainsi que les hypothèses de croissance et de continuité de φ permettent alors d'affirmer

$$\begin{aligned} \int \varphi(\mathcal{M}(\cap_n A_n, t), t) dt &\leq \int \liminf_n \varphi(\mathcal{M}(A_n, t), t) dt \\ &\leq \liminf_n \int \varphi(\mathcal{M}(A_n, t), t) dt. \end{aligned}$$

Autrement dit, (H3) est vérifiée.

L'intégration se fait par rapport à dt , dans cet exemple, pour que les points de discontinuité de $t \mapsto \mathcal{M}(\cap_n A_n, t)$ soient de masse nulle. Il est possible (mais a priori peu utile) d'intégrer par rapport à une autre mesure qui ne charge pas les ensembles dénombrables. Dans la section des applications statistiques, nous ferons un choix particulier de ce φ .

L'hypothèse (H2) faite sur la fonction de taille fait intervenir le contraste de Hausdorff. Ce choix un peu particulier permet d'alléger l'hypothèse dans les cas où la fonction de taille peut être finie pour des ensembles non compacts (il faut notamment que l'espace ne soit pas totalement borné). Dans le cas où tous les ensembles de taille finie sont compacts, la remarque suivante montre que la condition sur le contraste de Hausdorff est inutile.

Remarque 2.12 (Sur l'hypothèse (H2)). Dans l'hypothèse (H2), la suite $(A_n)_{n \geq 1}$ est supposée vérifier $\text{Haus}(A_n | \cap_k A_k) \rightarrow 0$. En fait, ceci est toujours vrai dès que A_n est compact, pour un certain $n \geq 1$. En effet, pour tout $t > 0$, $\mathcal{M}(A_n, t)$ est

décroissante et devient donc constante à partir d'un certain rang. Si $\{x_1^n, \dots, x_j^n\}$ est un ensemble t -séparé de A_n , où $j = \lim_n \mathcal{M}(A_n, t)$, alors, par compacité, une sous-suite $(x_1^n, \dots, x_j^n)_{n \geq 1}$ converge vers une limite que l'on note (x_1, \dots, x_j) . L'ensemble $\{x_1, \dots, x_j\}$ est t -séparé et est inclus dans $\cap_k A_k \subset A_n$ pour tout $n \geq 1$. Ainsi, par maximalité de cet ensemble dans A_n , pour n assez grand,

$$A_n \subset \{x_1, \dots, x_j\}^t \subset (\cap_k A_k)^t.$$

Comme t est arbitraire, ceci montre bien $\text{Haus}(A_n | \cap_k A_k) \rightarrow 0$.

Remarque 2.13 (Espaces totalement bornés). *Pour les espaces dont les compacts coïncident avec les fermés bornés, si tous les A_n sont non compacts, alors la limite $\cap_k A_k$ ne peut pas être compacte, puisque $A_n \subset (\cap_k A_k)^\varepsilon$, pour tout $\varepsilon > 0$ et n suffisamment grand. Et donc, $\mathcal{M}(A_n, t) = \mathcal{M}(\cap_k A_k, t) = \infty$ pour tout $t > 0$ et $n \geq 1$. Ainsi, pour ces espaces, d'après la remarque précédente, la condition $\text{Haus}(A_n | \cap_k A_k) \rightarrow 0$ demandée à la suite (A_n) peut-être omise (puisque pour les cas où aucun A_n est compact, $\tau(A_n) = \Phi(\infty) \rightarrow \tau(\cap_k A_k) = \Phi(\infty)$ est vérifié trivialement).*

Le choix de faire dépendre la *taille* d'un ensemble à travers le packing n'est pas fortuit et provient du fait que le packing apparaît dans différentes notions de *taille*.

La mesure de packing définie dans [Saint Raymond et Tricot, 1988] est une autre définition de la mesure de Lebesgue (à une constante près) et est, dans le cas plus général d'un espace métrique, une mesure qui joue un rôle dual à celui de la mesure de recouvrement ; toutes les deux étant aussi liées à la mesure de Hausdorff, plus connue. D'une manière analogue, le packing permet aussi de définir une dimension. Ces liens sont établis dans [Saint Raymond et Tricot, 1988].

Nous aurions pu aussi prendre une définition faisant dépendre la taille du recouvrement plutôt que du packing, mais le packing s'est avéré plus facile à manier dans les démonstrations. Il se peut cependant que la plupart, voire l'ensemble, des résultats restent vrais pour le recouvrement.

Concluons ce paragraphe de présentation avec une remarque légère sur la fonction φ de l'exemple 2.11.

Remarque 2.14 (Fonction φ universelle). *Il n'existe pas de fonction φ telle que $\lim_{x \rightarrow \infty} \varphi(x) = \infty$ pour laquelle $\inf\{\tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha\} < \infty$ pour tout espace polonais mesuré (E, d, μ) , où $\tau(A) = \int \varphi(\mathcal{M}(A, t)) dt$.*

En effet, soit (E, d) un espace polonais tel qu'il existe une suite décroissante $(t_i)_{i \geq 1}$ convergeant vers 0, telle que toute boule de rayon $t_i/2$ contient une infinité de boules disjointes de rayon t_{i+1} , quel que soit $i \geq 1$. On choisit alors une fonction

décroissante $f : \mathbb{R}_+^* \rightarrow \mathbb{N}$ telle que $k_i := f(t_i)/f(t_{i-1}) \in \mathbb{N}$ et on définit $(x_n^0)_{n \leq f(t_0)}$ un ensemble t_0 -séparé puis récursivement, $(x_n^i)_{n \leq f(t_i)}$ t_i -séparé, de sorte que chaque $B(x_n^{i-1}, t_{i-1}/2)$ contienne un nombre k_i de ces points.

On pose ensuite $\mu_n = \sum_{n \leq f(t_i)} \delta_{x_n^i} / f(t_i)$. La suite $(\mu_n)_{n \geq 1}$ est de Cauchy dans l'espace de Wasserstein, et elle possède donc une sous-suite convergente vers une mesure μ .

La mesure μ satisfait $\mu(B(x_n^i, t_i/2)) = 1/f(t_i)$ pour tout $i \geq 1, 1 \leq n \leq f(t_i)$. Donc, pour tout borélien A tel que $\mu(A) \geq 1 - \alpha$, nécessairement, $A \cap B(x_n^i, t_i/2) \neq \emptyset$ pour au moins $(1 - \alpha)f(t_i)$ boules et donc $\mathcal{M}(A, t_i) \geq (1 - \alpha)f(t_i)$.

Ainsi,

$$\int \varphi(\mathcal{M}(A, t)) dt \geq \varphi((1 - \alpha)f(t_i)) t_i \xrightarrow{i \rightarrow \infty} \infty,$$

pour un choix approprié de f fonction de φ .

2.2 Premières propriétés de la localisation de masse

2.2.1 Existence

Le cadre est toujours le suivant. (E, d) est un espace polonais et μ est une mesure borélienne de probabilité sur (E, d) . Étant donné une fonction τ , une classe stable \mathcal{F} de sous-ensembles de E et un niveau α , nous définissons le support B^α de la mesure μ^α α -localisée de μ par (lorsque c'est possible)

$$B^\alpha \in \arg \min \{ \tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha \},$$

et posons alors

$$\mu^\alpha = \mu(\cdot | B^\alpha).$$

La première question d'intérêt est celle de l'existence de B^α .

Le théorème suivant établit l'existence du minimum. Notons

$$\tau^\alpha = \inf \{ \tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha \}.$$

Théorème 2.15 (Existence d'un minimum). *Soient (E, d) un espace polonais, \mathcal{F} un ensemble stable et μ une mesure de probabilité sur $(E, \mathcal{B}(E))$. Soit $0 < \alpha < 1$. On suppose (H3). Alors, il existe un ensemble B tel que*

$$B \in \arg \min \{ \tau(A); A \in \mathcal{F}(E), \mu(A) \geq 1 - \alpha \}.$$

Remarque 2.16. *L'hypothèse (H3) ne peut être simplement omise. En effet, pour $E = \mathbb{R}^d$ et $\tau(B) = \liminf_{t \rightarrow 0} \mathcal{M}(B, t).t^d$, si l'on pose*

$$\mu = \alpha\gamma_d + (1 - \alpha)q,$$

où q est une mesure de probabilité de support \mathbb{Q}^d et γ_d est la mesure gaussienne standard sur \mathbb{R}^d , alors la suite $(B_n)_{n \geq 1}$ définie par

$$B_n := \{x_k\}_{1 \leq k \leq n} \cup B(0, r_n),$$

où $\{x_n\}_{n \geq 1} = \mathbb{Q}^d$ et $r_n \rightarrow 0$ de sorte que $\mu(B_n) = 1 - \alpha$, est une suite minimisante. Et $\tau(B_n) = \tau(B(0, r_n))$ de sorte que $\tau^\alpha = 0$ mais $\tau(\mathbb{Q}^d) = +\infty$.

Le *plus petit* borélien n'est pas nécessairement unique. Cependant, il est assez naturel qu'il y ait plusieurs *plus petits* boréliens, dans le cas, par exemple, d'une probabilité uniforme sur le carré unité. Toute translation (suffisamment petite) de *plus petit* borélien serait aussi de taille optimale. Malgré cette non unicité, le corollaire 2.34 viendra nous rassurer en montrant la compacité de l'ensemble des *plus petits* boréliens.

2.2.2 Régularité de τ

Il est assez naturel de vouloir que la fonction $\alpha \mapsto \tau^\alpha$ soit continue pour des mesures assez régulières. Nous imaginons bien la taille de B^α grandir continûment lorsque α tend vers zéro, pour des mesures uni-modales par exemple. L'objet de ce paragraphe est l'étude de la continuité de cette fonction. La première proposition en établit la continuité à droite, sous peu d'hypothèses.

Proposition 2.17 (Continuité à droite). *Soient (E, d) un espace polonais, μ une mesure de probabilité sur $(E, \mathcal{B}(E))$ et \mathcal{F} un ensemble stable. Soit $0 < \alpha < 1$. Alors, sous (H3), $\alpha \mapsto \tau^\alpha$ est continue à droite.*

Ce résultat encourageant ne permet pas de répondre à la question de la continuité, mais la non continuité dans le cas général de cette fonction est en fait assez évidente.

Remarque 2.18. *Il n'est pas vrai en général que τ^α est continue. En prenant, par exemple $\mu = (\delta_x + \delta_y)/2$ et $\alpha = 1/2$, il n'est pas difficile de choisir une fonction de taille τ qui ne soit pas continue en α .*

La continuité de la fonction de taille nécessite une certaine régularité de la mesure à localiser par rapport à l'ensemble \mathcal{F} des supports admissibles de la version localisée. La notion de \mathcal{F} -régularité permet cette continuité.

Définition 2.19 (\mathcal{F} -régularité). Une mesure μ est dite **\mathcal{F} -régulière** si pour tout $B \in \mathcal{F}$, pour tout $\delta > 0$ et pour tout $C \in \mathcal{F}$ tels que $B \subset C$ et $\mu(B) < \mu(B^\delta \cap C)$, il existe $A \in \mathcal{F}$ tel que

$$\begin{aligned} A &\subset B^\delta \cap C, \\ \mu(B) &< \mu(A). \end{aligned}$$

Cette notion de régularité n'est introduite que pour la continuité de la fonction de taille. Elle n'est restrictive sur μ que si l'ensemble \mathcal{F} n'est pas assez riche. En choisissant \mathcal{F} comme l'ensemble des fermés, toute mesure est \mathcal{F} -régulière. En effet, puisque $\mu(B^\delta \cap C) = \lim_n \mu(B^{\delta-1/n} \cap C)$, il existe $n \geq 1$ tel que $\mu(B) < \mu(B^{\delta-1/n} \cap C)$ et alors $A := \overline{B^{\delta-1/n} \cap C}$ convient. D'un autre côté, si \mathcal{F} n'est pas assez riche de sorte que $\tau(\mathcal{F})$ est non connexe, il est facile d'imaginer une mesure μ non \mathcal{F} -régulière.

Proposition 2.20 (Continuité). Soit μ une mesure de probabilité sur un espace polonais (E, d) . Sous les hypothèses (H1), (H2) et (H3), et les hypothèses supplémentaires que μ est \mathcal{F} -régulière, que son support est connexe et que τ^α est finie pour tout $\alpha > 0$, l'application $\alpha \mapsto \tau^\alpha$ est continue.

Remarque 2.21. La condition de finitude de τ^α ne sert qu'à éviter un cas dégénéré de discontinuité.

Ces propriétés étant énoncées, intéressons nous maintenant à la consistance de la localisation de masse de mesure.

2.3 Consistance de la localisation de masse

2.3.1 τ -tension

Une propriété importante souhaitée dans la localisation est celle de la consistance : on veut pouvoir approcher les zones denses en masse d'une mesure par les zones denses d'une mesure approchante. Nous verrons que cette propriété est vérifiée en dehors de cas dégénérés. La définition de suite τ -tendue permet d'exclure certains de ces cas. Elle est donnée aussi pour des mesures aléatoires afin de pouvoir y inclure - entre autres - la mesure empirique.

Définition 2.22 (τ -tension). Une suite de mesures aléatoires de probabilité $(\mu_n)_{n \geq 1}$ convergeant faiblement presque sûrement vers une mesure μ est dite **τ -tendue** si pour tout $\delta > 0$ et tout $B \in \mathcal{F}$ tels que $\tau(B) < \infty$, presque sûrement, pour tout $C \in \mathcal{F}$ tel que $B \subset C$ et $\mu(B) \leq \liminf_n \mu_n(C)$, il existe $A \in \mathcal{F}$ tel que

les trois conditions suivantes sont vérifiées :

$$\begin{aligned}\mu(B) &\leq \liminf_n \mu_n(A), \\ B &\subset A \subset B^\delta \cap C, \\ \tau(A) &< \infty.\end{aligned}$$

Un première remarque importante sur cette définition est qu'une mesure aléatoire τ -tendue n'est pas nécessairement, presque sûrement τ -tendue (c'est-à-dire en considérant la mesure déterministe une fois le tirage aléatoire effectué), ce qui peut être le cas de la mesure empirique par exemple. Ceci réside dans le fait que le « *presque sûrement* » de la définition intervient une fois les choix de B et δ fait. Une seconde remarque est la suivante. L'inégalité $\mu(B) \leq \liminf_n \mu_n(C)$ n'est pas une conséquence directe de $B \subset C$; le théorème du portemanteau n'indique en effet que $\limsup_n \mu_n(C) \leq \mu(C)$ et $\limsup_n \mu_n(B) \leq \mu(B)$ puisque B et C sont fermés. La condition de τ -tension pour les $B \in \mathcal{F}$ tels que $\mu(B) = \lim_n \mu_n(B)$ est trivialement vérifiée pour $A := B$. La définition de τ -tension s'explique alors ainsi : dès lors que $(\mu_n)_{n \geq 1}$ ne capture pas toute la masse de B (i.e. $\liminf_n \mu_n(B) < \mu(B)$) et qu'il existe un ensemble C contenant B dont $(\mu_n)_{n \geq 1}$ capture suffisamment la masse (i.e. $\mu(B) \leq \liminf_n \mu_n(C)$), alors \mathcal{F} contient un élément A dont la masse est elle aussi suffisamment capturée, de taille finie (i.e. $\tau(A) < \infty$), est qui est « coïncé » entre B , et un épaississement arbitrairement petit de B intersecté avec C .

La première application recherchée de ce chapitre, concerne les mesures empiriques. La proposition suivante affirme que la mesure empirique est τ -tendue.

Proposition 2.23 (τ -tension de la mesure empirique). *Soit μ une mesure de probabilité sur E telle que $\tau^\alpha < \infty$ pour tout $0 < \alpha < 1$. Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. de loi commune μ . Soit $\mu_n = \frac{1}{n} \sum_{1 \leq i \leq n} \delta_{X_i}$. Alors, $(\mu_n)_{n \geq 1}$ est τ -tendue.*

La fin de la démonstration de cette proposition 2.23 permet d'obtenir directement le corollaire suivant, qui éclaire beaucoup sur la notion de τ -tension.

Corollaire 2.24. *Soit $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité aléatoires sur E convergeant faiblement presque sûrement vers une mesure μ , telle que $\tau^\alpha < \infty$, pour tout $0 < \alpha < 1$. Si pour tout $B \in \mathcal{F}$, presque sûrement, $\mu(B) \leq \liminf_n \mu_n(B)$, alors $(\mu_n)_{n \geq 1}$ est τ -tendue.*

Ce corollaire permet de comprendre que la τ -tension est une implication de la convergence presque sûre de $\mu_n(B)$ pour chaque $B \in \mathcal{F}$, et donc que retirer le « presque sûr » rend la τ -tension beaucoup plus restrictive.

Nous pouvons maintenant nous intéresser à la consistance de la localisation de masse.

2.3.2 Consistance de τ

Dans l'idée de l'appliquer à la mesure empirique, nous voulons montrer que la taille du plus petit borélien de niveau α fixé pour une mesure μ_n tend vers la taille du plus petit borélien de la mesure limite, pour le même niveau. Comme annoncé précédemment, c'est la τ -tension qui permet d'établir cette consistance de la localisation, que nous montrons dans le résultat suivant.

Théoreme 2.25 (Consistance). *Soient (E, d) un espace polonais, \mathcal{F} une classe stable et $(\mu_n)_{n \geq 1}$ une suite τ -tendue de mesures aléatoires de probabilité sur $(E, \mathcal{B}(E))$ convergeant presque sûrement faiblement vers μ . Soit $0 < \alpha < 1$. Soient $B_n^\alpha \in \arg \min\{\tau(A); A \in \mathcal{F}, \mu_n(A) \geq 1 - \alpha\}$ et $\mu_n^\alpha = \mu_n(\cdot|B_n^\alpha)$, pour tout $n \geq 1$. Alors, sous les hypothèses (H1), (H2) et (H3), la suite $(\mu_n^\alpha)_{n \in \mathbb{N}}$ est presque sûrement relativement compacte pour la topologie de la convergence faible.*

Pour n'importe quel choix de sous-suite convergente $(\mu_{n_k}^\alpha)_{k \geq 1}$ de $(\mu_n^\alpha)_{n \geq 1}$, on pose $B_\infty^\alpha := \lim_k B_{n_k}^\alpha$. Alors B_∞^α satisfait $\mu(B_\infty^\alpha) \geq 1 - \alpha$ et presque sûrement

$$\tau^\alpha \leq \tau(B_\infty^\alpha) \leq \liminf_{n \rightarrow \infty} \tau(B_n^\alpha) \leq \limsup_{n \rightarrow \infty} \tau(B_n^\alpha) \leq \lim_{\varepsilon \rightarrow 0^+} \tau^{\alpha - \varepsilon}.$$

De plus, si μ est \mathcal{F} -régulière et de support connexe, alors les cinq termes ci-dessus sont égaux.

Remarque 2.26. *La condition de τ -tension de la suite n'est requise dans la démonstration que pour la dernière inégalité du résultat.*

Il est aussi clair que lorsque la fonction $\alpha \mapsto \tau^\alpha$ n'est pas continue pour la mesure μ il est difficile d'espérer une consistance de la taille du minimiseur. Ce premier résultat encourageant ne répond pas encore aux espérances : nous voudrions que les minimiseurs convergent, c'est l'objet du prochain paragraphe.

2.3.3 Consistance du minimiseur

Nous avons vu que sous certaines conditions, la taille du minimiseur B_n^α convergeait vers la taille du minimiseur de la mesure limite. L'unicité du minimiseur n'étant pas garantie, nous contournons ce problème en considérant le contraste de Hausdorff sur l'espace des ensembles de fermés de (E, d) muni de la distance de Hausdorff. Le résultat obtenu est énoncé dans le théorème 2.28.

Pour énoncer ce théorème, trois hypothèses supplémentaires - qui vont être discutées, sont faites, dont l'une d'elles nécessite la définition suivante.

Définition 2.27 (Stricte croissance). *On note $\mathcal{F}(\mathbb{R}^+, \mathbb{N})$ l'ensemble des fonctions de \mathbb{R}^+ à valeurs dans \mathbb{N} . Une fonction $\Phi : \mathcal{F}(\mathbb{R}^+, \mathbb{N}) \rightarrow \mathbb{R}$ est dite strictement croissante si elle est croissante et si pour tout $f, g \in \mathcal{F}(\mathbb{R}^+, \mathbb{N})$ tels que*

$$\forall t > 0, f(t) \leq g(t) \text{ et } \exists s > 0, \forall t < s, f(t) < g(t),$$

elle satisfait

$$\Phi(f) < \Phi(g).$$

Dans le cadre de la fonction de taille τ , les fonctions f jouent le rôle du packing. Cette définition se comprend alors ainsi : si l'on ajoute un élément à un ensemble compact, alors sa taille est strictement supérieure. En effet, ajouter un élément x à un ensemble compact K , augmente $\mathcal{M}(K, t)$ de 1 dès lors que $t < d(x, K)$.

Les trois hypothèses supplémentaires que nous ferons sont alors les suivantes.

(H4) $\cdot \mapsto \tau \cdot$ est continue en $\alpha \in]0, 1[$,

(H5) Φ définie dans (H3) est strictement croissante,

(H6) $\tau(A) < \infty \implies A$ précompact.

Ces hypothèses ne sont pas très restrictives. La continuité de la fonction de taille (H4) est une conséquence de la proposition 2.20 : un support connexe de la mesure μ suffit. En reprenant l'exemple 2.11 de fonction $\tau : \tau(B) = \int \varphi(\mathcal{M}(B, t)) dt$, l'hypothèse (H5) revient à demander la stricte croissance de φ . Enfin, dans ce même contexte de fonction de taille canonique, (H6) est toujours vérifiée.

Rappelons que pour deux ensembles A et B , le contraste de Hausdorff entre A et B est défini par

$$\text{Haus}(A|B) := \inf\{\varepsilon > 0 | A \subset B^\varepsilon\},$$

et posons les notations suivantes pour $0 < \alpha < 1$, une suite $(\mu_n)_{n \geq 1}$ et une mesure μ ;

$$\mathcal{S}_n^\alpha = \arg \min\{\tau(A); A \in \mathcal{F}, \mu_n(A) \geq 1 - \alpha\},$$

et

$$\mathcal{S}^\alpha = \arg \min\{\tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha\}.$$

\mathcal{S}^α et \mathcal{S}_n^α sont alors deux sous-ensembles de \mathcal{F} .

Nous pouvons maintenant énoncer le théorème de consistance en distance de Hausdorff.

Théoreme 2.28 (Consistance en distance de Hausdorff). *Soient (E, d) un espace polonais et $(\mu_n)_{n \geq 1}$ une suite τ -tendue de mesures aléatoires de probabilité sur $(E, \mathcal{B}(E))$ presque sûrement faiblement convergente vers une mesure μ . Soit $0 < \alpha < 1$. On équipe l'espace des ensembles compacts de (E, d) de la distance de Hausdorff notée d_H .*

On suppose, en plus de (H2) et (H3), les conditions (H4) (pour la mesure μ), (H5) et (H6).

Alors, presque sûrement,

$$\text{Haus}(\mathcal{S}_n^\alpha | \mathcal{S}^\alpha) \rightarrow 0.$$

Discutons un instant de la nécessité des hypothèses faites dans ce théorème.

Remarque 2.29 (Sur la condition (H4)). *Le résultat n'est pas vrai en général sans la continuité de $\cdot \mapsto \tau$ en α . En effet, si l'on pose $d\mu_n = (1_{[0;1-\alpha-1/n]} + 1_{[2;2+\alpha+1/n]})d\lambda$, alors pour $\tau(B) = \int_0^1 \mathcal{M}(B, t)tdt$ (par exemple), $\text{Supp } \mu_n^\alpha = [0; 1 - \alpha - 1/n] \cup [2; 2 + 1/n]$ et $\text{Supp } \mu_\infty^\alpha = [0; 1 - \alpha]$.*

Remarque 2.30 (Sur la condition (H5)). *Si τ n'est pas strictement croissante alors le résultat n'est pas vrai en général. En effet, pour un espace polonais (E, d) tel qu'aucune boule n'est compacte, en supposant qu'il existe $s > 0$ tel que,*

$$\Phi(\mathcal{M}(B^\alpha, \cdot)) = \Phi(f),$$

où $f(t) = \mathcal{M}(B^\alpha, t) + \mathbf{1}_{\{t < s\}}$, il vient $\tau^\alpha = \tau(B^\alpha) = \tau(B^\alpha \cup \{x\})$ pour tout x à distance inférieure à s d'un quelconque ensemble maximal s -séparé de B^α . Puisque ces x ne forment pas un ensemble compact, \mathcal{S}^α n'est pas compact non plus, ce qui contredit le corollaire 2.34.

Remarque 2.31 (Sur la condition (H6)). *Puisque qu'un ensemble non compact ne peut pas être arbitrairement approché par un ensemble fini en distance de Hausdorff, alors pour que la convergence de B_n^α ait lieu avec $(\mu_n)_{n \geq 1}$ une suite de mesures à support fini, les plus petits ensembles de μ doivent être compacts, ce qui est vérifié lorsque $\tau(B) < \infty$ implique que B est précompact.*

L'hypothèse de stabilité de \mathcal{F} est naturelle puisque la limite en distance de Hausdorff des minimiseurs doit être dans \mathcal{F} pour être un minimiseur pour la mesure limite. Il est cependant possible d'alléger un peu cette hypothèse.

Remarque 2.32 (Hypothèse de stabilité de \mathcal{F}). *Comme la taille minimale $\min\{\tau(A); A \in \mathcal{F}, \mu_n(A) \geq 1 - \alpha\}$ est majorée si $\min\{\tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha\} < \infty$, alors on peut supposer à la place de la stabilité de la classe \mathcal{F} que toutes les classes*

$$\mathcal{F}^M := \mathcal{F} \cap \{A; \tau(A) \leq M\}$$

pour $M < \infty$ sont stables; ce qui est une notion plus faible puisque $\tau(\lim B_n) \leq \liminf \tau(B_n)$ pour toute suite de fermés $(B_n)_{n \geq 1}$ comme le montre le lemme 2.41 et l'hypothèse (H3).

Nous avons utilisé le contraste de Hausdorff sur l'espace des ensembles de fermés de (E, d) . Nous pouvons nous demander s'il n'était pas possible d'utiliser la distance de Hausdorff plutôt que le contraste. C'est le sujet de la remarque suivante.

Remarque 2.33 (Convergence contraire). *La convergence de $\text{Haus}(\mathcal{S}^\alpha | \mathcal{S}_n^\alpha)$ n'est pas vraie en général. Choisissons, par exemple $\tau(B) = \int_0^1 \mathcal{M}(B, t) dt$ et $d\mu_n = f_n d\lambda$ sur \mathbb{R} , où $f_n(x) = 1 + x/n$ pour $-1/2 < x < 1/2$ et $f(x) = 0$ sinon. Alors, le plus petit ensemble $[-1/2; 1/2 - \alpha]$ est optimal (i.e. appartient à \mathcal{S}^α) pour μ la loi uniforme sur $[0, 1]$ et ne sera jamais inclus dans un épaississement arbitrairement petit de $[-1/2 + \alpha; 1/2] \in \mathcal{S}_n^\alpha$.*

Lorsque \mathcal{S}^α est un singleton, ce théorème montre alors que tous les B_n^α convergent vers le *plus petit* borélien de masse $1 - \alpha$ de la mesure limite μ . Lorsque \mathcal{S}^α n'est pas un singleton, ce théorème exprime alors le fait que chacun des B_n^α contenus dans \mathcal{S}_n^α approche un élément de \mathcal{S}^α , et ceci uniformément. En revanche, ce que précise la remarque 2.33, certains éléments de \mathcal{S}^α ne seront pas approchés par les B_n^α .

La démonstration de ce théorème 2.28 permet de rassurer quant à la non unicité du *plus petit* borélien en montrant le corollaire suivant.

Corollaire 2.34. *Soient (E, d) un espace polonais et μ une mesure de probabilité sur $(E, \mathcal{B}(E))$ telle que $\tau^\alpha < \infty$. Soit $0 < \alpha < 1$. On suppose (H2), (H3), (H4), (H5) et (H6). Alors $\arg \min\{\tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha\}$ est compact pour la topologie de la distance de Hausdorff.*

Nous pouvons profiter des outils développés dans ce paragraphe pour montrer la continuité du minimiseur pour une mesure donnée, par rapport au niveau α .

2.3.4 Continuité du minimiseur

La technique de démonstration consistant à appliquer les deux lemmes 2.41 et 2.42 à la suite, permet d'établir une sorte de continuité du minimiseur par rapport à α . Comme pour la consistance du minimiseur, nous nous plaçons de l'espace des ensembles de fermés de (E, d) munis du contraste de Hausdorff.

Proposition 2.35. *Soient (E, d) un espace polonais et μ une mesure de probabilité sur $(E, \mathcal{B}(E))$. Soient $0 < \alpha < 1$ et $(\alpha_n)_{n \geq 1}$ convergeant vers α . On suppose (H2), (H3), (H4), (H5) et (H6). Alors,*

$$\text{Haus}(\mathcal{S}^{\alpha_n} | \mathcal{S}^\alpha) \rightarrow 0.$$

Nous pouvons alors facilement faire la remarque suivante dans le cas où le minimiseur est unique, grâce au contraste de Hausdorff.

Remarque 2.36. *Si \mathcal{S}^α est un singleton, pour tout $0 < \alpha < 1$, alors la proposition établit la continuité de l'application qui associe à α l'ensemble de plus petite taille parmi les ensembles de \mathcal{F} de mesure supérieure à $1 - \alpha$.*

2.3.5 Cas de la mesure empirique

La proposition 2.23 et le théorème 2.28 permettent de monter la consistance du minimiseur pour la mesure empirique μ_n . Rappelons les notations suivantes.

$$\mathcal{S}_n^\alpha = \arg \min \{ \tau(A); A \in \mathcal{F}, \mu_n(A) \geq 1 - \alpha \},$$

$$\mathcal{S}^\alpha = \arg \min \{ \tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha \}.$$

Nous obtenons alors le résultat suivant.

Théoreme 2.37 (Consistance : cas indépendant identiquement distribué). *Soient (E, d) un espace polonais et μ une mesure de probabilité sur $(E, \mathbb{B}(E))$ telle que $\tau^\alpha < \infty$, pour tout $0 < \alpha < 1$. Pour un échantillon $(X_n)_{n \geq 1}$ de variables aléatoires de loi commune μ , on note*

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

la mesure empirique. On suppose (H2), (H3), (H4), (H5) et (H6). Alors, presque sûrement,

$$\text{Haus}(\mathcal{S}_n^\alpha | \mathcal{S}^\alpha) \rightarrow 0.$$

De même, le corollaire 2.24 permet de montrer la convergence des minimiseurs dans des cas où les variables aléatoires sont dépendantes.

Théoreme 2.38 (Consistance pour une chaîne de Markov). *Soient (E, d) un espace polonais et μ une mesure de probabilité sur $(E, \mathbb{B}(E))$ telle que $\tau^\alpha < \infty$, pour tout $0 < \alpha < 1$. Pour $(X_n)_{n \geq 1}$ une suite de variables aléatoires issue d'une chaîne de Markov ergodique de mesure invariante μ , on note*

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

la mesure empirique. On suppose (H2), (H3), (H4), (H5) et (H6). Alors, presque sûrement,

$$\text{Haus}(\mathcal{S}_n^\alpha | \mathcal{S}^\alpha) \rightarrow 0.$$

2.4 Applications statistiques

L'application statistique la plus directe de ces résultats est la localisation de masse de mesure. L'objectif est de trouver l'ensemble de plus petite taille, ayant une probabilité pour une mesure μ supérieure à $1 - \alpha$, à partir d'une suite de mesures approchant μ .

La suite de mesures peut être une suite de mesures empiriques $\mu_n = \sum_{1 \leq i \leq n} \delta_{X_i}$ issue d'une suite de variables aléatoires $\{X_n\}_{n \geq 1}$ indépendantes et identiquement distribuées, en utilisant le théorème 2.37. Mais d'autres suites de mesures, telles que la suite des mesures d'occupation d'une chaîne de Markov (théorème 2.38), peuvent être τ -tendue.

Nous présentons alors des applications de ces théorèmes pour des choix de τ et de \mathcal{F} particuliers.

2.4.1 Localisation de masse

2.4.1.a Présentation

Dans le cas simple où \mathcal{F} est l'ensemble des boules d'un espace métrique (E, d) dans lequel toutes les boules de même rayon sont isométriques (elles ont alors la même taille, quel que soit τ), le théorème 2.28 assure que la suite $(B_n)_{n \geq 1}$ des plus petites boules de mesure empirique μ_n supérieure à $1 - \alpha$, converge en distance de Hausdorff vers la (ou une) plus petite boule de mesure μ supérieure à $1 - \alpha$.

Sur \mathbb{R}^d , l'ensemble des convexes est stable (voir définition 2.1). En posant

$$B \mapsto \tau(B) := \int_0^\varepsilon \mathcal{M}(B, t) t^d dt,$$

pour un $\varepsilon > 0$ fixé, le théorème 1.13 affirme que $\mathcal{M}(B, t) t^d$ converge vers la mesure de Lebesgue de $\lambda(B)$ (à une constante multiplicative c près) pour les ensembles Jordan-mesurables, et donc notamment les convexes. Trouver le convexe de mesure μ supérieure à $1 - \alpha$ de plus petite taille pour τ est alors une notion proche de celle qui consiste à trouver le convexe de mesure supérieure à $1 - \alpha$ de volume minimal. Le théorème 2.28 assure cependant, dans le cas de la minimisation sur τ , la convergence des minimiseurs associés à une suite de mesures $(\mu_n)_{n \geq 1}$ τ -tendue, qui converge vers μ , telle que la mesure empirique.

2.4.1.b Simulations

Le calcul de τ est en général trop complexe pour être réalisé sur une classe \mathcal{F} trop grande, pour trouver un minimiseur. Afin de pallier à ce problème, nous avons considéré dans ces simulations un ensemble \mathcal{F} sur lequel τ est relativement simple

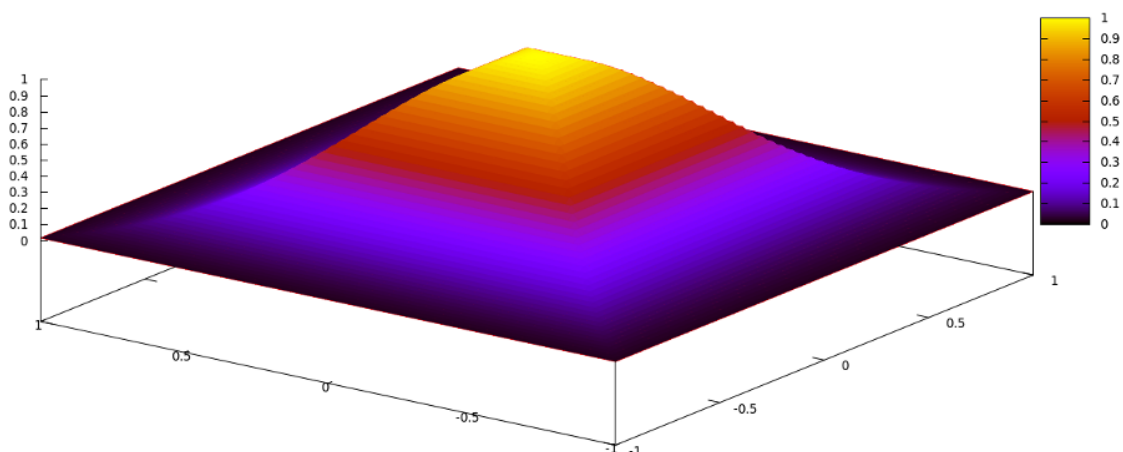


FIGURE 2.1 – Mesure à lignes de niveau carrées

à calculer : les rectangles dont les bords sont parallèles aux axes dans \mathbb{R}^2 . Pour obtenir des calculs simples, nous avons aussi considéré la distance non euclidienne issue de la norme uniforme sur \mathbb{R}^2 , pour laquelle le packing est connu.

L'algorithme pour effectuer les simulations fonctionne de la manière suivante.

- 1- On tire un échantillon de taille n .
- 2- On tire $nrect$ rectangles de manière aléatoire.
- 3- On calcule le nombre d'éléments de l'échantillon dans chacun des rectangles.
- 4- On calcule le τ de chacun de ces rectangles.
- 5- On cherche le rectangle ayant le plus petit τ parmi ceux contenant au moins une proportion $1 - \alpha$ des éléments de l'échantillon.

Du fait de l'étape 2, le rectangle trouvé n'est pas forcément optimal, cependant, lorsque le nombre d'éléments de l'échantillon devient trop grand, il est trop coûteux d'énumérer tous les rectangles, et cet aléatoire offre alors une alternative réalisable.

Les simulations ont permis de mettre en évidence l'importance du choix de la classe \mathcal{F} sur laquelle on minimise. Dans un premier temps, nous avons tiré un échantillon suivant la loi dont la densité

$$(x, y) \mapsto 4e^{-\max(x^2, y^2)}$$

est représentée sur la figure 2.1, dont les lignes de niveau sont carrées pour que les rectangles puissent correspondre facilement.

L'échantillon est de taille 500. Nous montrons ici le rectangle de plus petite taille retenu par l'algorithme, pour la fonction de taille $\tau(B) = \int_0^{1/2} \mathcal{M}(B, t) \cdot t^2 dt$, pour α variant de 0.2 à 0.8. Les résultats sont présentés en figure 2.2.

Dans ce cas simple et adapté aux rectangles, l'algorithme semble trouver des valeurs cohérentes. Le cas $\alpha = 0.2$ semble le plus perturbé, ce qui provient vraisemblablement l'aléa de l'algorithme, ainsi que du fait que les bords des rectangles sont dépendants de valeurs plus rares, relativement éloignées du centre de masse en $(0, 0)$.

Dans un second temps, nous avons fait varier la taille n de l'échantillon, de 100, à 500, puis à 1000 et choisi une mesure gaussienne standard ; l'objectif étant d'illustrer la convergence du minimiseur et de la fonction de taille τ . Les résultats sont présentés en figure 2.3.

La stabilité du rectangle optimal vient illustrer le théorème 2.28 de la consistance en distance de Hausdorff du minimiseur pour la mesure empirique. De même, les valeurs de τ varient autour de 22, ce qui illustre le théorème 2.25, bien que le nombre de valeurs pour n (100, 500 et 1000) soit faible.

2.4.2 Quantification robuste

2.4.2.a Présentation

De nombreux domaines d'applications cherchent à regrouper en classes les différents points d'un ensemble. On parle alors de quantification. La méthode des k -means est une réponse à ce problème, et pour un nombre k de classes donné, regroupe les données de sorte que la somme de la distance (au carré) moyenne des points d'une classe à leur barycentre soit minimale (une présentation plus précise est faite en section 2.1). Pour certaines applications, le fait de devoir choisir un nombre de classes k à l'avance est un inconvénient. Un autre aspect qui peut être un inconvénient est le manque de stabilité des barycentres des classes par rapport aux données parfois observé. L'objet de ce paragraphe est de présenter une méthode basée sur les résultats de ce chapitre, qui vise à quantifier une mesure de manière robuste, sans se fixer de nombre de classe à l'avance, pouvant imposer ou non une forme spécifique aux classes.

Pour une mesure μ sur un espace métrique (E, d) totalement borné de dimension de packing supérieure (voir définition 1.15) d , nous voulons ici « localiser » les ensembles connexes « bien disjoints » qui contiennent une masse importante. Formellement, pour une classe \mathcal{F} d'ensembles de E , qui seront les ensembles connexes que l'on admettra, et pour $\varepsilon > 0$, qui quantifie la « disjonction » de ces ensembles connexes, nous voulons alors minimiser

$$B \mapsto \tau(B) := \int_0^\varepsilon \mathcal{M}(B, t) t^d dt$$

sur l'ensemble des $B \in \mathcal{F}_\varepsilon$ tels que $\mu(B) \geq 1 - \alpha$ où

$$\mathcal{F}_\varepsilon := \{\cup_{F \in \mathcal{G}} F; \mathcal{G} \subset \mathcal{F}, \forall F, G \in \mathcal{G} \inf_{x \in F, y \in G} d(x, y) \geq \varepsilon\},$$

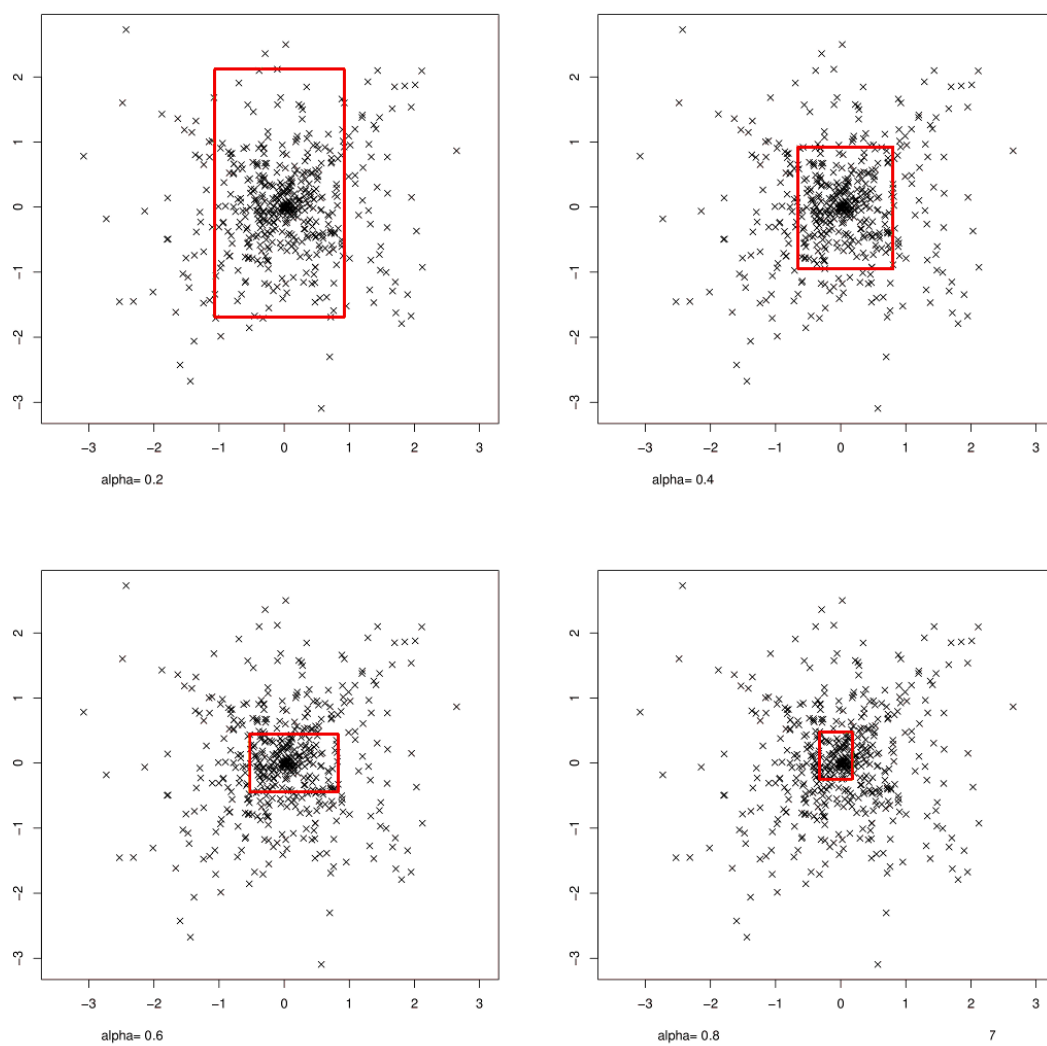


FIGURE 2.2 – Évolution en fonction de α

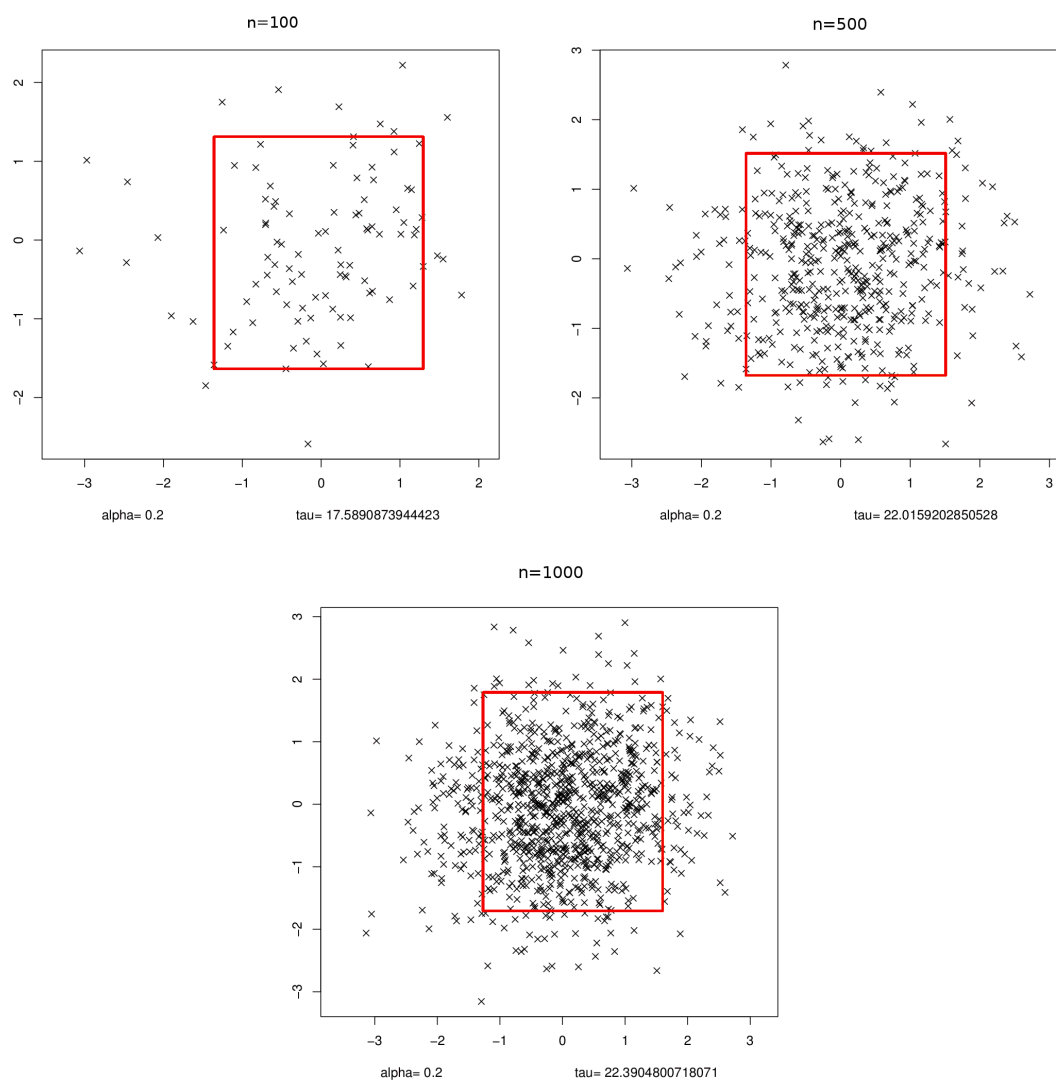


FIGURE 2.3 – Évolution avec le taille de l'échantillon

comme défini dans l'exemple 2.8 (il suffit alors, pour que \mathcal{F}_ε soit une classe stable, que \mathcal{F} vérifie les hypothèses du lemme A.2 ; c'est, par exemple, le cas de l'ensemble des convexes). Les composantes connexes d'un minimiseur sont alors les classes de la quantification de μ . Le théorème 2.28 permet alors d'affirmer qu'une suite de minimiseurs de τ sur les éléments de \mathcal{F}_ε tels que $\mu_n(B) \geq 1 - \alpha$ converge en distance de Hausdorff vers l'union B^α des zones de forte masse de μ (lorsque B^α est unique). Notamment, le nombre de composantes connexes (les classes) de B_n est constant à partir d'un certain rang et correspond à celui de B^α .

La remarque 2.32 précise une condition plus faible sur la classe \mathcal{F} qui suffit à la convergence du minimiseur B_n : supposer à la place de la stabilité de la classe \mathcal{F} que toutes les classes

$$\mathcal{F}^M := \mathcal{F} \cap \{A; \tau(A) \leq M\}$$

pour $M < \infty$ sont stables.

Il devient alors possible de définir une façon robuste de classification non supervisée, en associant à chaque point de l'espace la partie de connexe la plus proche du minimiseur. La robustesse devient alors une conséquence de la localisation de masse : les points loin des zones denses en masse (donc rares), ne modifient donc ni le minimiseur, ni la classification.

2.4.2.b Simulations

Nous reprenons ici l'algorithme précédent, en y ajoutant une étape où il peut choisir éventuellement l'union de deux ou trois rectangles distancés d'au moins $\varepsilon > 0$, lorsque ceux-ci forment un ensemble de plus petite taille, \mathcal{F} est alors l'ensemble des unions d'un, deux, ou trois composantes connexes. Afin de faciliter les calculs, la classe \mathcal{F} est l'ensemble des unions d'un, deux ou trois carrés isométriques distancés d'au moins $\varepsilon > 0$. L'objectif est de vérifier l'intuition qui ferait détecter à cet algorithme deux modes d'une distribution (et plus généralement le nombre de modes d'une distribution).

Intuitivement, nous nous attendons à ce que lorsque les modes sont bien distincts, un α proche de 0 suffirait à les découvrir. En revanche, lorsque ceux-ci sont proches, un α proche de 1 semble indispensable pour détecter tous les modes, mais demande alors un n relativement grand afin d'éviter les zones de points regroupés « par hasard », ne correspondant pas à une réelle zone dense de la loi.

Le choix des rectangles (ou des carrés) a été fait ici pour répondre à une limite technique (le temps de calcul). Il est cependant possible de modifier cet algorithme pour chercher à minimiser sur la classe de l'union des connexes ou des convexes distancés d'au moins ε , qui pourrait alors ne pas souffrir du problème de choix inadapté de classe sur laquelle minimiser (comme dans nos simulations : un carré pour les zones denses de gaussiennes).

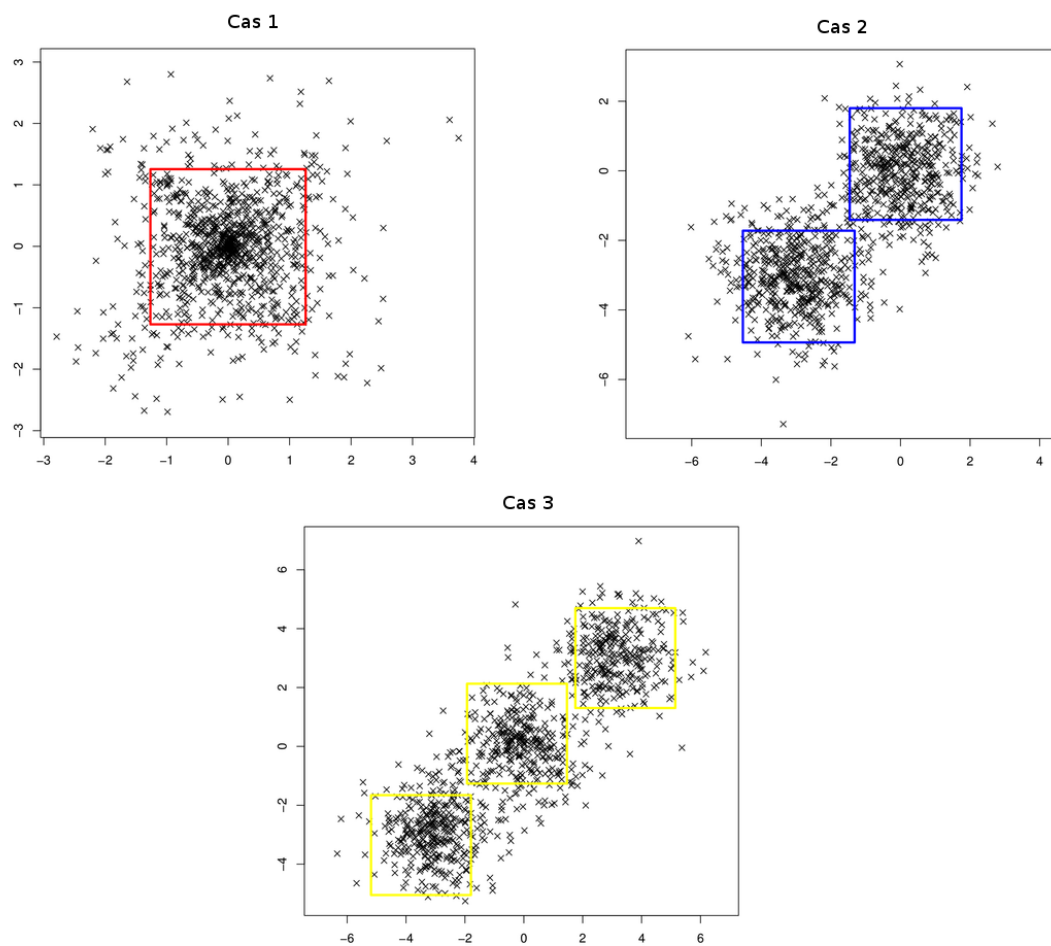


FIGURE 2.4 – Comparaison du minimiseur pour un écart $e = 3$, $\alpha = 0.2$, pour les cas 1, 2 et 3

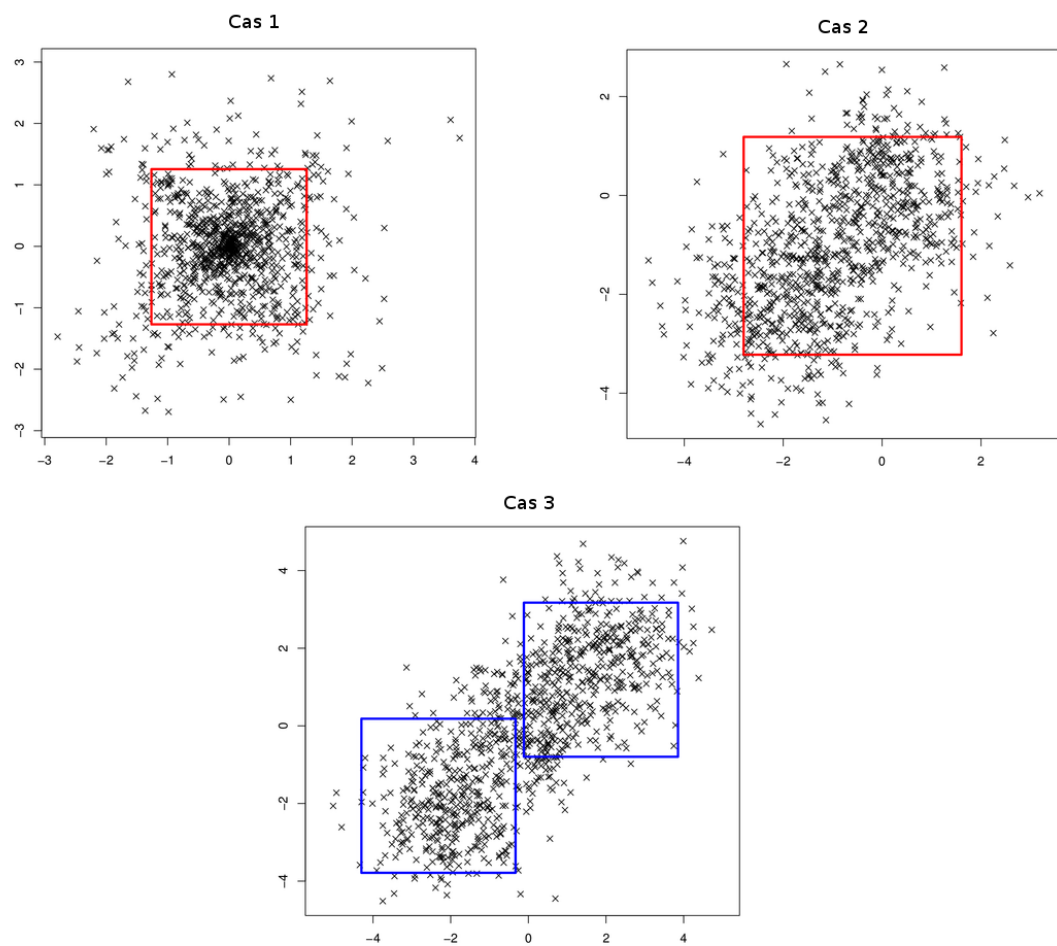


FIGURE 2.5 – Comparaison du minimiseur pour un écart $e = 2$, $\alpha = 0.2$, pour les cas 1, 2 et 3

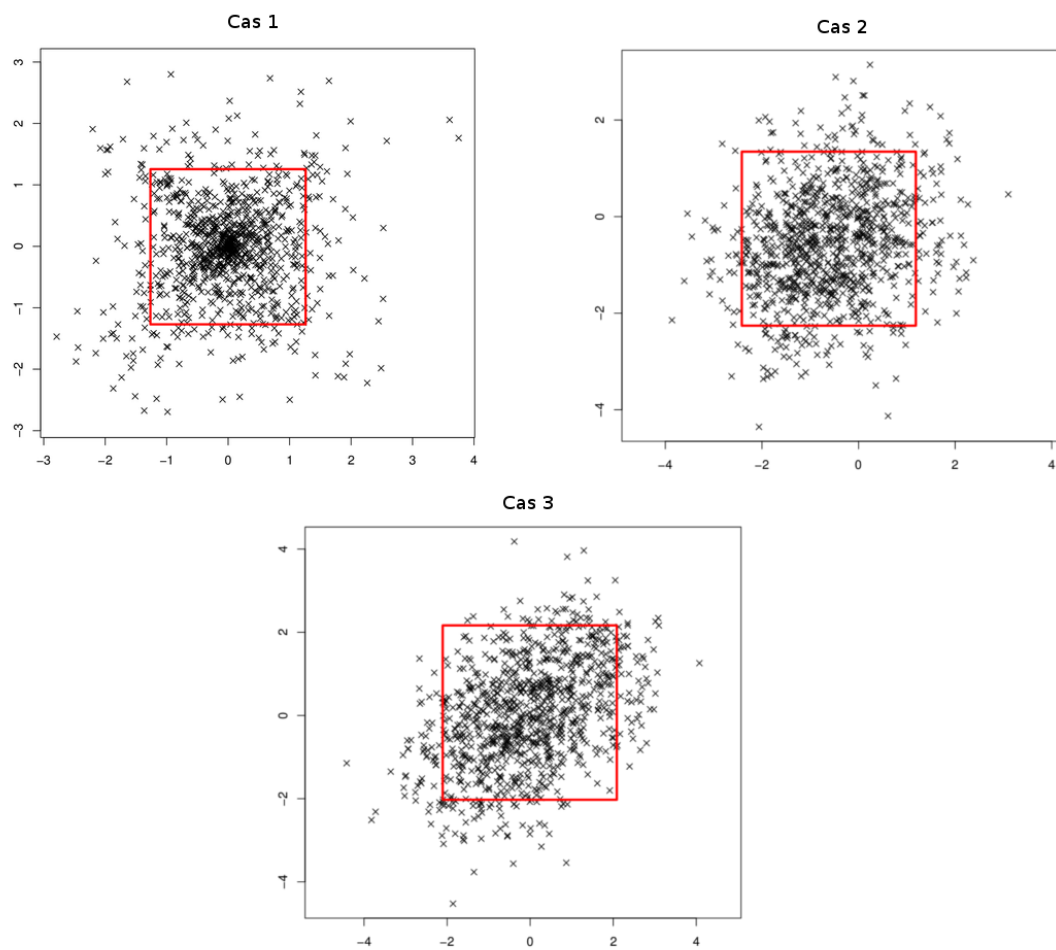


FIGURE 2.6 – Comparaison du minimiseur pour un écart $e = 1$, $\alpha = 0.2$, pour les cas 1, 2 et 3

2.4.2.c Tableau des résultats et interprétations

Nous avons tiré un échantillon de taille 1000 suivant trois cas. Le cas 1 correspond à la loi de densité décrite en figure 2.1. Le cas 2 est un mélange de deux gaussiennes centrées en $(0, 0)$ et en $(-e, -e)$ pour e (appelé écart) prenant les valeurs 1, 2, et 3. Le cas 3 correspond à un mélange de trois gaussiennes centrées en $(0, 0)$, (e, e) et en $(-e, -e)$, pour les mêmes valeurs 1, 2, et 3 de e .

Dans chaque cas, nous avons calculé le minimiseur parmi l'union d'un, deux, ou trois carrés isométriques, pour la taille τ définie par

$$\tau(B) = \int_0^\varepsilon \mathcal{M}(B, t) t^2 dt.$$

Chacune de ces simulations a été répétée cent fois.

Les tableaux suivants montrent alors pour chaque valeur de l'écart e entre les gaussiennes, le nombre de carrés du minimiseur (1, 2, ou 3) choisi le plus souvent, suivi entre parenthèses, de la fréquence de ce choix sur les cents simulations.

Écart = 3					
α	cas 1	$\alpha \setminus \text{cas}$	2	3	
0.1	1 (1)	0.1	2 (0.78)	3 (0.72)	
0.2	1 (1)	0.2	2 (1)	3 (1)	
0.3	1 (1)	0.3	2 (1)	3 (1)	
0.5	1 (1)	0.5	2 (1)	3 (1)	

Écart = 2			Écart = 1		
$\alpha \setminus \text{cas}$	2	3	$\alpha \setminus \text{cas}$	2	3
0.1	1 (1)	1 (0.94)	0.1	1 (1)	1 (1)
0.2	1 (1)	2 (0.95)	0.2	1 (1)	1 (1)
0.3	1 (0.72)	2 (0.90)	0.3	1 (1)	1 (1)
0.5	2 (0.95)	3 (0.61)	0.5	1 (1)	1 (0.98)

Lorsque l'écart entre les gaussiennes mélangées est faible, l'ensemble de niveau est connexe, et le minimiseur de τ l'est aussi. On trouve alors toujours (dans nos simulations) un minimiseur connexe (un seul carré). Lorsque l'écart devient plus grand (pour $e = 3$), le minimiseur n'est plus connexe, et est atteint avec un carré sur chaque gaussienne du mélange (avec une fréquence de 0.78 et 0.72 pour un mélange de deux et trois gaussiennes respectivement). Pour un écart de 2, le nombre de composantes connexes augmente avec α , et la fréquence de réussite diminue.

2.5 Démonstrations

Le lemme suivant est fondamental dans la démonstration de l'existence du *plus petit* borélien et reste utilisé dans la plupart des autres démonstrations de ce chapitre.

Lemme 2.39. *Soit (E, d) un espace polonais. Soit $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité convergeant faiblement vers μ_∞ . Alors,*

$$\text{Supp } \mu_\infty \subset \lim_n \text{Supp } \mu_n.$$

De plus, pour tout $t > 0$

$$\mathcal{M}(\text{Supp } \mu_\infty, t) \leq \mathcal{M}(\lim_n \text{Supp } \mu_n, t) \leq \liminf_n \mathcal{M}(\text{Supp } \mu_n, t).$$

Démonstration du lemme 2.39. Soient $t > 0$ et $x \in \text{Supp } \mu_\infty$. Alors, pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que $\mu_\infty(B(x, \varepsilon/2)) > \eta$. Alors, d'après le théorème du portemanteau, pour n suffisamment grand, $\mu_n(B(x, \varepsilon/2)) > \eta$ et ainsi, $x \in (\text{Supp } \mu_n)^\varepsilon$ pour tout n suffisamment grand, ce qui démontre le premier point et la première inégalité.

Pour la seconde inégalité, choisissons $m \leq \mathcal{M}(\lim_n \text{Supp } \mu_n, t)$. Soit $\{x_i\}_{1 \leq i \leq m}$ un sous-ensemble t -séparé de $\lim_n \text{Supp } \mu_n$. Alors, il existe $\delta > 0$ tel que

$$\inf_{i \neq j} d(x_i, x_j) > t + \delta.$$

D'après la définition de $\lim_n \text{Supp } \mu_n$ et le fait que m soit fini, il existe $k \geq 1$ tel que $(x_i)_{1 \leq i \leq m} \subset \cap_{n \geq k} (\text{Supp } \mu_n)^{\delta/2}$

À présent le choix $x_i^n \in B(x_i, \delta/2) \cap \text{Supp } \mu_n$ pour $1 \leq i \leq m$. $\{x_i^n\}_{1 \leq i \leq m}$ donne un ensemble t -séparé inclus dans $\text{Supp } \mu_n$ puisque pour $i \neq j$

$$\begin{aligned} d(x_i^n, x_j^n) &\geq d(x_i, x_j) - d(x_i, x_i^n) - d(x_j, x_j^n) \\ &> t + \delta - \delta/2 - \delta/2 = t. \end{aligned}$$

Ainsi, $m \leq \liminf_n \mathcal{M}(\text{Supp } \mu_n, t)$, pour tout $m \leq \mathcal{M}(\lim_n \text{Supp } \mu_n, t)$, ce qui achève la démonstration. □

Démonstration du théorème 2.15. Soit $(B_n)_{n \geq 1}$ une suite minimisante, qui vérifie $\mu(B_n) \geq 1 - \alpha$ pour tout $n \geq 1$. Soit $\mu_n = \mu(\cdot | B_n)$. Dans un premier temps, on montre que $(\mu_n)_{n \geq 1}$ est une suite tendue. Soit K un ensemble compact tel que

$\mu(K) \geq 1 - \varepsilon(1 - \alpha)$, alors

$$\begin{aligned}\mu_n(K) &= \mu(K \cap B_n) / \mu(B_n) \\ &= (1 - \mu(K^c \cup B_n^c)) / \mu(B_n) \\ &\geq 1 - \mu(K^c) / \mu(B_n) \\ &\geq 1 - \varepsilon.\end{aligned}$$

Donc, quitte à extraire une sous-suite, $(\mu_n)_{n \geq 1}$ converge faiblement vers μ_∞ . Soit $B = \lim_n \text{Supp } \mu_n$. B^ε est un ensemble ouvert et contient $\text{Supp } \mu_\infty$ d'après le lemme 2.39. Donc, en utilisant le théorème du portemanteau,

$$\begin{aligned}1 = \mu_\infty(B^\varepsilon) &\longleftarrow \mu_n(B^\varepsilon) = \mu(B^\varepsilon \cap B_n) / \mu(B_n) \\ &\leq \mu(B^\varepsilon) / (1 - \alpha).\end{aligned}$$

En faisant tendre ε vers 0 et en utilisant $\bigcap_{\varepsilon > 0} B^\varepsilon = B$, il s'ensuit $\mu(B) \geq 1 - \alpha$.

La démonstration s'achève en utilisant (H3), et à nouveau le lemme 2.39. \square

Démonstration de la proposition 2.17. Clairement, τ^α est décroissante, donc, il suffit de montrer que $\tau^\alpha \leq \lim_{\varepsilon \rightarrow 0^+} \tau^{\alpha+\varepsilon}$. Soient $B_\alpha \in \arg \min\{\tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha\}$ et $\mu^\alpha = \mu(\cdot | B_\alpha)$. Pour toute suite $(\alpha_n)_{n \geq 1} \subset (0; 1)$ décroissante et qui converge vers α , $(\mu^{\alpha_n})_{n \geq 1}$ est tendue. En effet, pour un ensemble compact K tel que $\mu(K) \geq 1 - \varepsilon(1 - \alpha_1)$

$$\begin{aligned}\mu^{\alpha_n}(K) &= \mu(K \cup B_{\alpha_n}) / \mu(B_{\alpha_n}) \\ &\geq (1 - \mu(K^c) - \mu(B_{\alpha_n}^c)) / \mu(B_{\alpha_n}) \\ &= 1 - \mu(K^c) / \mu(B_{\alpha_n}) \\ &\geq 1 - \varepsilon.\end{aligned}$$

Ainsi, quitte à extraire une sous-suite, μ^{α_n} converge vers une mesure de probabilité μ_∞ . Soit $B_\infty = \lim_n \text{Supp } \mu_n$. B_∞^ε est un ensemble ouvert contenant $\text{Supp } \mu_\infty$, donc, en utilisant le théorème du portemanteau,

$$\begin{aligned}1 = \mu_\infty(B_\infty^\varepsilon) &\longleftarrow \mu^{\alpha_n}(B_\infty^\varepsilon) \\ &= \mu(B_\infty^\varepsilon \cap B_{\alpha_n}) / \mu(B_{\alpha_n}) \\ &\leq \mu(B_\infty^\varepsilon) / (1 - \alpha_n) \longrightarrow \mu(B_\infty^\varepsilon) / (1 - \alpha).\end{aligned}$$

En faisant tendre ε vers 0, ceci montre $\mu(B_\infty) \geq 1 - \alpha$. Le lemme 2.39 et l'hypothèse (H3) permettent de conclure

$$\tau^\alpha \leq \tau(B_\infty) \leq \lim_n \inf \tau(B_{\alpha_n}) = \lim_n \tau(B_{\alpha_n}).$$

\square

Pour pouvoir montrer la continuité de la fonction taille (proposition 2.20), on établit un lemme qui énonce que pour une mesure à support connexe, l'épaississement de tout borélien capture de la masse.

Lemme 2.40. *Soit μ une mesure de probabilité à support connexe alors pour tout borélien A tel que $\mu(A) < 1$ et tout $\varepsilon > 0$, le ε -voisinage de A satisfait*

$$\mu(A^\varepsilon) > \mu(A).$$

Démonstration du lemme 2.40. Dans le but d'établir une contradiction, on suppose $\mu(A^\varepsilon) = \mu(A)$ pour un certain $\varepsilon > 0$. Alors, $\mu(A^{\varepsilon/3}) = \mu(A)$ et $\mu(((A^\varepsilon)^c)^{\varepsilon/3}) \geq \mu((A^\varepsilon)^c) = 1 - \mu(A)$ de sorte que $\mu(A^{\varepsilon/3} \cup ((A^\varepsilon)^c)^{\varepsilon/3}) = 1$. Puisque $A^{\varepsilon/3} \cap ((A^\varepsilon)^c)^{\varepsilon/3} = \emptyset$ et $A^{\varepsilon/3}$ et $((A^\varepsilon)^c)^{\varepsilon/3}$ sont fermés, alors le support de μ est non connexe, ce qui contredit l'hypothèse. \square

On peut alors montrer la continuité de la fonction de taille.

Démonstration de la proposition 2.20. D'après la proposition 2.17 il suffit de montrer que $\alpha \mapsto \tau^\alpha$ est continue à gauche. De plus, puisque $\alpha \mapsto \tau^\alpha$ est décroissante, il suffit de montrer $\lim_{\varepsilon \rightarrow 0^+} \tau^{\alpha-\varepsilon} \leq \tau^\alpha$.

Soient $\alpha > 0$ et $B \in \mathcal{F}$ tels que $\tau^\alpha = \tau(B)$. Si $\mu(B) > 1 - \alpha$ alors le résultat est trivial. On suppose alors $\mu(B) = 1 - \alpha$. D'après le lemme 2.40, pour tout $0 < \delta < \alpha$ il existe $\varepsilon > 0$ tel que $\mu(B^\delta) \geq 1 - \alpha + \varepsilon$. On définit une suite décroissante $(\delta_n)_{n \geq 1}$ convergente vers 0, telle que $\mu(B^{\delta_n}) > 1 - \alpha$ pour tout $n \geq 1$. Puisque μ est \mathcal{F} -régulière, il existe $(K_n)_{n \geq 1} \subset \mathcal{F}$ telle que pour tout $n \geq 1$

$$\begin{aligned} \tau(K_1) &< \infty, \text{ car } \tau^{\alpha-\varepsilon} < \infty, \\ K_n &\subset B^{\delta_n} \cap K_{n-1}, \\ 1 - \alpha &< \mu(K_n) =: 1 - \alpha_n. \end{aligned}$$

Clairement, $\alpha_n \rightarrow \alpha$. Alors, puisque $\mathcal{M}(K_n, t) \leq \mathcal{M}(B, t - 2\delta_n)$, il s'ensuit - d'après (H1) et (H2)

$$\lim_{\varepsilon \rightarrow 0^+} \tau^{\alpha-\varepsilon} \longleftarrow \tau^{\alpha_n} \leq \tau(K_n) \longrightarrow \tau(\cap_k K_k) = \tau^\alpha.$$

\square

Démonstration de la proposition 2.23. Par la finitude de τ^α , pour tout $\varepsilon > 0$, il existe $B \in \mathcal{F}$ tel que $\tau(B) < \infty$. Pour un tel B , la loi des grands nombres assure que, presque sûrement, $\lim_n \mu_n(B) = \mu(B)$. Soit $\delta > 0$. Alors, pour tout $C \in \mathcal{F}$ tel que $B \subset C$ et $\mu(B) \leq \liminf_n \mu_n(C)$, pour $A = B \in \mathcal{F}$, les trois conditions

suivantes sont trivialement vérifiées

$$\begin{aligned}\mu(B) &\leq \liminf_n \mu_n(A), \\ B &\subset A \subset B^\delta \cap C, \\ \tau(A) &< \infty.\end{aligned}$$

□

Démonstration du théorème 2.25. Soit K un compact tel que $\mu_n(K) \geq 1 - \varepsilon(1 - \alpha)$ pour tout $n \geq 1$ et $\mu(K) \geq 1 - \varepsilon(1 - \alpha)$. Alors,

$$\begin{aligned}\mu_n^\alpha(K) &= \mu_n(K \cap B_n^\alpha) / \mu_n(B_n^\alpha) \\ &\geq 1 - \varepsilon,\end{aligned}$$

montrant la tension de la suite $(\mu_n^\alpha)_{n \geq 1}$. On note μ_∞^α une mesure limite et $B_\infty^\alpha = \lim_k \text{Supp } \mu_{n_k}^\alpha$, pour une sous-suite convergente de μ_n^α (que l'on renomme $(\mu_n^\alpha)_{n \geq 1}$). Soit $\varepsilon = \inf\{\eta > 0; \mu_n(B) \leq \mu(B^\eta) + \eta, \forall B \in \mathbb{B}(E)\}$ la distance de Prokhorov entre μ et μ_n . Alors, en utilisant à la fois le théorème du portemanteau et le fait que la distance de Prokhorov métrise la topologie de la convergence faible (voir théorème A.5), pour tout $\eta > 0$ il existe n_η tel que pour tout $n > n_\eta$, puisque $\text{Supp } \mu_\infty^\alpha \subset B_\infty^\alpha$ d'après le lemme 2.39,

$$\begin{aligned}1 &= \mu_\infty^\alpha(B_\infty^\alpha) \\ &\leq \mu_n^\alpha((B_\infty^\alpha)^\eta) + \eta \\ &= \mu_n((B_\infty^\alpha)^\eta \cap B_n^\alpha) / \mu_n(B_n^\alpha) + \eta \\ &\leq \mu_n((B_\infty^\alpha)^\eta) / (1 - \alpha) + \eta \\ &\leq (\mu((B_\infty^\alpha)^{\eta+\varepsilon}) + \varepsilon) / (1 - \alpha) + \eta \longrightarrow \mu(B_\infty^\alpha) / (1 - \alpha) \text{ lorsque } \varepsilon, \eta \rightarrow 0.\end{aligned}$$

Ainsi, $\mu(B_\infty^\alpha) \geq 1 - \alpha$ et donc, par définition de τ^α ,

$$\tau^\alpha \leq \tau(B_\infty^\alpha).$$

La seconde inégalité vient en application directe de (H3) et du lemme 2.39.

On renomme une sous-suite de $(\mu_n)_{n \geq 1}$, de sorte que $\limsup_{n \rightarrow \infty} \tau(B_n^\alpha) = \lim_{n \rightarrow \infty} \tau(B_n^\alpha)$. On choisit alors $\varepsilon > 0$, et $B_\varepsilon \in \arg \min\{\tau(A); A \in \mathcal{F}, \mu(A) \geq 1 - \alpha + \varepsilon\}$. Sans perte de généralité, il peut être supposé que $\tau(B_\varepsilon) < \infty$. Puisque la suite $(\mu_n)_{n \geq 1}$ converge vers μ , le théorème du portemanteau énonce que pour tout $\delta > 0$,

$$\liminf_n \mu_n(B_\varepsilon^\delta) \geq \mu(B_\varepsilon).$$

Soit $\delta_1 > 0$. Alors, puisque $E \in \mathcal{F}$, la τ -tension implique qu'il existe $A_1 \in \mathcal{F}$ tel que

$$\begin{aligned}\tau(A_1) &< \infty, \\ B_\varepsilon &\subset A_1 \subset B_\varepsilon^{\delta_1}, \\ 1 - \alpha + \varepsilon &\leq \mu(B_\varepsilon) \leq \liminf_n \mu_n(A_1).\end{aligned}$$

Soit $(\delta_k)_{k \geq 1}$ une suite décroissante convergeant vers 0. En utilisant de manière récursive la τ -tension de la suite $(\mu_n)_{n \geq 1}$, avec (en reprenant les notations de la définition de τ -tension) les choix $B = B_\varepsilon$, $\delta = \delta_k$ et $C = A_{k-1}$, on définit une suite décroissante $(A_k)_{k \geq 1}$ d'éléments de \mathcal{F} , telle que pour tout $k \geq 1$,

$$\begin{aligned}B_\varepsilon &\subset A_k \subset B_\varepsilon^{\delta_k}, \\ 1 - \alpha + \varepsilon &\leq \mu(B_\varepsilon) \leq \liminf_n \mu_n(A_k).\end{aligned}$$

Ensuite, pour tout $k \geq 1$, il existe $n_k \geq 1$ tel que $\mu_n(A_k) \geq 1 - \alpha$ pour tout $n > n_k$. Ainsi, puisque $A_k \in \mathcal{F}$, $\tau(B_n^\alpha) \leq \tau(A_k)$ pour tout $n \geq n_k$. Par construction et puisque B_ε est fermé, $\cap_k A_k = B_\varepsilon$. En utilisant (H1) et (H2), il vient

$$\tau(B_n^\alpha) \leq \tau(A_k) \longrightarrow_k \tau(\cap_k A_k) = \tau(B_\varepsilon),$$

ce qui montre que pour tout $\varepsilon > 0$,

$$\limsup_n \tau(B_n^\alpha) \leq \tau^{\alpha-\varepsilon}.$$

L'égalité des cinq termes lorsque μ est \mathcal{F} -régulière et à support connexe est alors une application directe de la proposition 2.20. \square

La démonstration du théorème 2.28 se base sur les deux lemmes suivants.

Lemme 2.41. *Soient (E, d) un espace polonais et $(B_n)_{n \geq 1}$ une suite de sous-ensembles de E . Alors, pour tout $\varepsilon > 0$ tel que $\mathcal{M}(\lim_k B_k, \varepsilon) < \infty$, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n > n_0$,*

$$\lim_k B_k \subset B_n^{2\varepsilon}.$$

Démonstration de lemme 2.41. Soient $\varepsilon > 0$ et $m = \mathcal{M}(\lim_k B_k, \varepsilon)$. Soit $\{x_i\}_{1 \leq i \leq m}$ un sous-ensemble ε -séparé maximal de $\lim_n B_n$. Alors, par définition de $\lim_k B_k$ et puisque m est fini, pour tout $\delta > 0$, il existe k_δ tel que pour tout $n \geq k_\delta$,

$$\{x_i\}_{1 \leq i \leq m} \subset B_n^\delta.$$

La maximalité de $\{x_i\}_{1 \leq i \leq m}$ assure

$$\lim_k B_k \subset \bigcup_{1 \leq i \leq m} \overline{B(x_i, \varepsilon)} \subset B_n^{\varepsilon+\delta}.$$

La démonstration s'achève en choisissant $\delta \leq \varepsilon$. \square

Lemme 2.42. Soit (E, d) un espace métrique. Soit $(B_n)_{n \geq 1}$ une suite d'ensembles fermés telle que $\text{Haus}(B|B_n) \rightarrow 0$, pour un certain fermé B . On suppose que pour tout $s > 0$, il existe $0 < t < s$ tel que

$$\liminf \mathcal{M}(B_n, t) \leq \mathcal{M}(B, t) < \infty. \quad (2.5)$$

Alors, le long d'une sous-suite (renommée) de $(B_n)_{n \geq 1}$,

$$d_H(B_n, B) \rightarrow 0.$$

Démonstration de lemme 2.42. Il suffit de montrer

$$\forall \varepsilon > 0, \forall n_0 \in \mathbb{N}, \exists n \geq n_0, B_n \subset (B)^\varepsilon.$$

Dans le but montrer une contradiction, supposons

$$\exists \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, B_n \not\subset (B)^\varepsilon. \quad (2.6)$$

Alors,

$$\exists \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n > n_0, \exists x_n \in B_n, \forall y \in B, d(x_n, y) > \varepsilon.$$

Soient $t < \varepsilon$ et $m = \mathcal{M}(B, t)$. Puisque $\mathcal{M}(B, \cdot)$ est continue à droite, il existe $\delta > 0$ tel que

$$m = \mathcal{M}(B, t) = \mathcal{M}(B, t + \delta).$$

Soit $\{z_i\}_{i \leq m}$ un sous-ensemble $(t + \delta)$ -séparé optimal de B . Alors, puisque pour tout $\eta > 0$ et n suffisamment grand,

$$B \subset (B_n)^\eta,$$

on peut construire $\{y_i\}_{i \leq m} \subset B_n$ tel que pour tout $1 \leq i \leq m$, $d(y_i, z_i) \leq \eta$. Ainsi, pour tout $i \neq j$,

$$d(y_i, y_j) \geq d(z_i, z_j) - d(z_i, y_i) - d(z_j, y_j) > t + \delta - 2\eta.$$

On choisit alors $\eta < \delta/2 \wedge (\varepsilon - t)$ tel que $\{y_i\}_{i \leq m}$ est un sous-ensemble t -séparé de B_n . Alors, $\{x_n\} \cup \{y_i\}_{i \leq m}$ est aussi un sous-ensemble t -séparé de B_n puisque $x_n \in B_n$ et

$$d(x_n, y_i) \geq d(x_n, z_i) - d(z_i, y_i) > \varepsilon - \eta > t,$$

montrant ainsi $\mathcal{M}(B, t) + 1 \leq \mathcal{M}(B_n, t)$, pour tout n suffisamment grand. Ceci contredit (2.5) et montre ainsi que (2.6) était faux. \square

Ces deux lemmes appliqués l'un à la suite de l'autre aboutissent à la convergence en distance de Hausdorff de la suite d'ensembles $(B_n)_{n \geq 1}$. C'est la stratégie de la démonstration du théorème 2.28 de consistance en distance de Hausdorff.

Démonstration du théorème 2.28. Soit $B_n^\alpha \in \mathcal{S}_n^\alpha$. On pose $\mu_n^\alpha = \mu_{n|B_n^\alpha}$. Le théorème 2.25 ainsi que le lemme 2.39 et la continuité (H4) de $\cdot \mapsto \tau \cdot$ montrent que $(\mu_n^\alpha)_{n \geq 1}$ est relativement compacte pour la topologie de la convergence faible et que $B_\infty^\alpha := \lim_k B_{n_k} \in \mathcal{S}^\alpha$ où $(n_k)_{k \geq 1}$ est une sous-suite le long de laquelle μ_n^α converge.

Le théorème 2.25 ainsi que (H6) assure $\mathcal{M}(B_\infty^\alpha, \varepsilon) < \infty$ pour tout $\varepsilon > 0$. Le lemme 2.41 montre ainsi le long d'une sous-suite,

$$\text{Haus}(B_\infty^\alpha | B_n^\alpha) \rightarrow 0.$$

Alors, le lemme 2.39 énonce

$$\forall t > 0, \mathcal{M}(B_\infty^\alpha, t) \leq \liminf \mathcal{M}(B_n^\alpha, t),$$

alors que le théorème 2.25 ainsi que la continuité (H4) de $\cdot \mapsto \tau \cdot$ et (H3) montrent

$$\Phi(\mathcal{M}(B_\infty^\alpha, \cdot)) = \Phi(\liminf \mathcal{M}(B_n^\alpha, \cdot)).$$

Alors, puisque Φ est strictement croissante (H5), pour tout $s > 0$, il existe $t < s$ tel que

$$\mathcal{M}(B_\infty^\alpha, t) = \mathcal{M}(\lim_n B_n^\alpha, t) = \liminf \mathcal{M}(B_n^\alpha, t).$$

Le lemme 2.42 s'applique alors et donne

$$d_H(B_\infty^\alpha, B_n^\alpha) \rightarrow 0, \tag{2.7}$$

pour une sous-suite de $(B_n)_{n \geq 1}$. Puisque (2.7) est vraie pour toute sous-suite de n'importe quelle sous-suite de la suite originale $(B_n^\alpha)_{n \geq 1}$, en prenant $B_n^\alpha \in \mathcal{S}_n^\alpha$ pour que $1/n + d_H(B_n^\alpha, B_\infty^\alpha) \geq \text{Haus}(\mathcal{S}_n^\alpha | \mathcal{S}^\alpha)$ pour tout $B_\infty^\alpha \in \mathcal{S}^\alpha$, il s'ensuit

$$\text{Haus}(\mathcal{S}_n^\alpha | \mathcal{S}^\alpha) \rightarrow 0.$$

□

Démonstration du corollaire 2.34. On pose $(\mu_n)_{n \geq 1} = (\mu)_{n \geq 1}$ dans le théorème 2.28 et le résultat est alors une directe application de (2.7) de sa démonstration. □

Démonstration de la proposition 2.35. On choisit une suite $(\alpha_n)_{n \geq 1}$ qui converge vers α . Par continuité de la fonction de taille, pour n assez grand, $\tau^{\alpha_n} < \infty$ et donc, (H6) implique la compacité de tout élément de \mathcal{S}^{α_n} . On définit $\mu_n = \mu(\cdot | B_n)$ pour tout $B_n \in \mathcal{S}^{\alpha_n}$. La suite $(\mu_n)_{n \geq 1}$ est tendue. En effet, soit $K \in \mathcal{F}$ tel que $\mu(K) \geq 1 - \varepsilon(1 - \max_{n \geq 1} \alpha_n)$, alors

$$\begin{aligned} \mu_n(K) &= \mu(K \cap B_n) / \mu(B_n) \\ &= (1 - \mu(K^c \cup B_n^c)) / \mu(B_n) \\ &\geq 1 - \mu(K^c) / \mu(B_n) \\ &\geq 1 - \varepsilon. \end{aligned}$$

Soit $B_\infty = \lim_k B_{n_k}$ où $(n_k)_{k \geq 1}$ est une suite le long de laquelle $(\mu_n)_{n \geq 1}$ est convergente (qui existe par tension de $(\mu_n)_{n \geq 1}$). On pose $\mu_\infty = \lim_k (\mu_{n_k})_{k \geq 1}$. En utilisant des arguments similaires à ceux du théorème 2.15, on peut montrer que $\mu(B_\infty) \geq 1 - \alpha$. Ainsi, en utilisant (H3) et le lemme 2.39,

$$\tau^\alpha \leq \tau(B_\infty) \leq \Phi(\liminf \mathcal{M}(B_n, \cdot)) \leq \liminf \tau(B_n) = \lim \tau^{\alpha_n} = \tau^\alpha, \quad (2.8)$$

où la dernière inégalité est due à la condition de continuité (H4) de $\cdot \mapsto \tau \cdot$. L'application du lemme 2.41 à une sous-suite convergente de $(\mu_n)_{n \geq 1}$ montre que le long de cette sous-suite,

$$\text{Haus}(B_\infty | B_n) \rightarrow 0.$$

Cependant, la seconde inégalité - qui est en fait une égalité - de (2.8) avec la stricte croissance de Φ (H5) montre que pour tout $s > 0$ il existe $t < s$ tel que

$$\mathcal{M}(B_\infty, t) = \mathcal{M}(\lim_n B_n, t) = \liminf \mathcal{M}(B_n, t).$$

Le lemme 2.42 montre ainsi

$$d_H(B_\infty, B_n) \rightarrow 0. \quad (2.9)$$

On remarque alors que le corollaire 2.34 établit que \mathcal{S}^{α_n} est compact. Puisque (2.9) est vérifiée pour une sous-suite de n'importe quelle sous-suite de la suite originale $(B_n)_{n \geq 1}$, en prenant

$$B_n \in \arg \max \left\{ \min_{B \in \mathcal{S}^\alpha} d_H(B, A) \mid A \in \mathcal{S}^{\alpha_n} \right\},$$

de sorte que $d_H(B_n, B_\infty) \geq \text{Haus}(\mathcal{S}^{\alpha_n} | \mathcal{S}^\alpha)$ pour tout $B_\infty \in \mathcal{S}^\alpha$, il vient

$$\text{Haus}(\mathcal{S}^{\alpha_n} | \mathcal{S}^\alpha) \rightarrow 0.$$

□

2.6 Conclusion

Cette nouvelle notion de localisation présente de bonnes propriétés que sont la continuité de la fonction de taille et du *plus petit* borélien par rapport à α ainsi que la consistance en distance de Hausdorff des supports des mesures localisées. Elle trouve aussi sa justification dans le fait qu'elle se base sur la notion de packing, qui intervient dans différentes notions génériques de taille, comme présenté à l'introduction de ce chapitre.

Cette méthode permet aussi de pouvoir imposer une forme au support de la mesure localisée, à travers un choix de \mathcal{F} , permettant de faire éventuellement usage d'un a priori sur ces formes. En revanche, les calculs des simulations, pour des mesures à support fini, ont pu être réalisés uniquement grâce à un choix restrictif de \mathcal{F} , qui pourrait devenir moins restrictif en cas d'amélioration de l'algorithme. Aucun algorithme véritablement efficace n'a pour l'instant été développé. Cet inconvénient s'étend au calcul des minimiseurs pour des mesures μ standards, quelque soit le choix de τ . Il n'existe pour l'instant pas de méthode pour trouver, par exemple, le minimiseur de

$$B \mapsto \tau(B) = \int (\mathcal{M}(B, t) - 1) t^d dt,$$

sur les $B \in \mathcal{F} \cap \{\mu(B) \geq 1 - \alpha\}$ pour μ une gaussienne standard sur \mathbb{R}^d .

Chapitre 3

Mesure empirique et distance de Wasserstein

Ce chapitre regroupe des résultats nouveaux obtenus à partir d'un théorème original. Ce travail, en collaboration avec Emmanuel Boissard, a donné lieu à un article à paraître dans les annales de l'Institut Henri Poincaré. Les démonstrations et formulations présentées ici diffèrent légèrement de l'article, et une partie nouvelle est présentée ici (la section 3.4).

Une première partie de ce chapitre est consacrée à la présentation du problème. Puis, vient le théorème principal de ce chapitre, qui donne lieu à différents résultats sur les trois parties suivantes. Une partie est ensuite consacrée aux applications statistiques, puis viennent les démonstrations et une conclusion.

3.1 Présentation

Nous nous intéressons à la vitesse de convergence en distance de Wasserstein de la mesure empirique vers sa mesure limite. Le cadre est le suivant. (E, d, μ) est un espace polonais mesuré et $(X_i)_{i \geq 1}$ est une suite de variables aléatoires à valeurs dans E indépendantes de loi commune μ . On note

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

la mesure empirique associée à $(X_i)_{i \geq 1}$. [Varadarajan, 1958] a montré que presque sûrement (p.s.) μ_n converge faiblement vers μ .

Pour $1 \leq p < +\infty$, la distance de Wasserstein d'ordre p est définie sur l'ensemble $\mathcal{P}_p(E)$ des mesures de probabilité sur (E, d) admettant un moment d'ordre p fini, par

$$W_p^p(\mu, \nu) = \inf_{\pi \in \mathcal{P}(\mu, \nu)} \int d^p(x, y) \pi(dx, dy)$$

où l'infimum porte sur l'ensemble $\mathcal{P}(\mu, \nu)$ des mesures de probabilité avec première marginale μ , et seconde marginale ν . Ceci définit une distance sur \mathcal{P}_p , et la convergence en cette distance est équivalente à la convergence faible et du moment d'ordre p .

Ces distances jouent un rôle important dans différents domaines des mathématiques. Elles interviennent en géométrie notamment à travers leur rôle central dans la théorie de Sturm-Lott-Villani, elles fournissent aux statistiques un outil pour quantifier la proximité de deux mesures et elles apparaissent encore en traitement d'images.

Ce chapitre présente une majoration de l'espérance de la distance de Wasserstein entre la mesure empirique et la vraie mesure

$$E(W_p(\mu, \mu_n)).$$

Ce travail avait pour but original d'étendre les résultats de [Bolley et al., 2007] dans le cas de mesures de probabilité sur \mathbb{R}^d .

De tels résultats sont recherchés notamment pour des applications numériques et statistiques. [Graf et Luschgy, 2000] s'intéresse, par exemple, à l'approximation en distance de Wasserstein, d'une mesure de probabilité donnée par une mesure de probabilité de support fini.

De nombreux travaux s'intéressent aux taux de convergence selon certaines métriques associées à la convergence faible. En particulier, l'article [Dudley, 1969] qui s'intéresse au cas des métriques BL^* et de Prokhorov, est à l'origine d'une technique de démonstration qui est reprise dans ce chapitre.

D'autres travaux ont traité de la convergence de la mesure empirique en distance de Wasserstein, dont les suivants.

L'article [Horowitz et Karandikar, 1994] a établi une majoration de la convergence de $\mathbb{E}[W_2^2(\mu_n, \mu)]$ vers 0 pour un mesure μ quelconque sur \mathbb{R}^d sous une condition de moment. Les articles [Ajtai et al., 1984] et [Talagrand, 1992] ont étudié le problème de matching équivalent pour la mesure uniforme sur le cube unité en dimension $d \geq 2$. Les articles [Dobric et Yukich, 1995] et [Barthe et Bordenave, 2011] ont approfondi l'étude du problème de matching. Ces articles montrent des résultats précis pour les mesures sur \mathbb{R}^d , améliorant ainsi des résultats de [Horowitz et Karandikar, 1994] et de [Bolley et al., 2007]. Dans le cas d'une mesure sur \mathbb{R} , [Del Barrio et al., 1999] a obtenu un théorème central limite pour $W_1(\mu_n, \mu)$ sous la condition

$$\int_{-\infty}^{+\infty} \sqrt{F(t)(1-F(t))} dt < +\infty,$$

où F est la fonction de répartition de μ .

L'approximation d'une mesure par sa version empirique est aussi une réponse au problème de quantification, c'est-à-dire de l'approximation d'une mesure par une mesure à support fini. Une référence sur ce sujet est le livre [Graf et Luschgy, 2000].

3.2 Première majoration

Le cadre est toujours le suivant. (E, d, μ) est un espace métrique mesuré et $(X_i)_{i \geq 1}$ est une suite de variables aléatoires indépendantes de loi commune μ . X est aussi une variable aléatoire de loi μ . On note μ_n la mesure empirique associée à la suite $(X_i)_{i \geq 1}$. L'objectif est de majorer $EW_p^p(\mu, \mu_n)$ pour $p \geq 1$. Dans un premier temps, E sera supposé compact. Pour énoncer le résultat principal, on introduit la définition suivante.

Définition 3.1 (t-recouvrement par des boules). *Un ensemble $\{B_i\}_{i \in I}$ de boules fermées de E est appelé un **t-recouvrement** de B si le rayon de tout B_i n'excède pas t et*

$$B \subset \bigcup_{i \in I} B_i.$$

*Pour un ensemble B , on appelle alors **nombre de recouvrement pour t** , noté $\mathcal{N}(B, t)$, le cardinal du plus petit t-recouvrement par des boules de B .*

On peut alors énoncer le résultat principal de ce chapitre, qui trouve son intérêt dans son application à des espaces métriques mesurés particuliers.

Théorème 3.2. *Soit (E, d, μ) un espace métrique mesuré de diamètre D . Alors pour tout $0 < t < D^p$,*

$$EW_p^p(\mu, \mu_n) \leq 2^p \left[2^{p+1}t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{\mathcal{N}(E, u^{1/p}) - 1} du \right].$$

L'inégalité de Jensen (théorème A.6) permet de déduire directement de ce résultat, la majoration suivante de l'espérance de la distance de Wasserstein.

$$EW_p(\mu, \mu_n) \leq 2 \left[2^{p+1}t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{\mathcal{N}(E, u^{1/p}) - 1} du \right]^{1/p}.$$

La démonstration de ce théorème est inspirée de [Dudley, 1969] et utilise des arguments de chaînage. Elle se rapproche aussi de [Barthe et Bordenave, 2011], qui traite le cas \mathbb{R}^d en découpant l'espace en cubes de plus en plus petits, ce qui ne pouvait se faire dans un espace métrique quelconque.

Afin de bien comprendre les implications de ce théorème, différentes applications de celui-ci sont traitées.

3.3 Applications

3.3.1 Espaces de dimension finie

La première application du théorème 3.2 traite du cas de la dimension d finie, au sens où l'on suppose qu'il existe $d > 0$ tel que pour tout $x \in E$, tout $r > 0$ et tout $u \leq r$,

$$\mathcal{N}(B(x, r), u) \leq (r/u)^d. \quad (3.1)$$

Le résultat s'énonce alors ainsi.

Théorème 3.3. *Soit (E, d_E, μ) un espace métrique mesuré. On suppose qu'il existe $d > 0$ tel que pour tout $x \in E$ et tout $r > 0$, $\mathcal{N}(B(x, r), u) \leq (r/u)^d$. On suppose aussi que pour $p < d/2$ et $\delta > dp/(d-p) \vee 2p$, $Ed_E(x_0, X)^\delta < \infty$. Alors, il existe $C > 0$ dépendant seulement de p , d , et $Ed_E(x_0, X)^\delta$, tel que pour tout $n \geq 1$,*

$$EW_p^p(\mu, \mu_n) \leq Cn^{-p/d}.$$

Ce résultat est optimal pour les dimensions supérieures ou égales à trois (dans le cas d'un espace euclidien) dans le sens où la vitesse de convergence en $n^{-p/d}$ est celle de $E(W_p^p(\mu, \mu_n))$, comme le montre [Barthe et Bordenave, 2011]. L'apport de ce résultat est celui du cas non euclidien, et du cas des dimensions non entières au sens du recouvrement par des boules. Il est intéressant de remarquer que l'on retrouve aussi le taux de la borne inférieure du quantifieur optimal. Pour les dimensions d supérieures à 3, la mesure à support fini de cardinal n la plus proche d'une mesure à support infini définie sur un espace de dimension d est à une distance de Wasserstein de l'ordre de $n^{1/d}$. Ce résultat permet donc d'affirmer qu'il est asymptotiquement équivalent de choisir la mesure empirique d'un tirage de n variables aléatoires, comme approximation d'une mesure μ , ou de choisir la mesure la plus proche supportée sur n points.

3.3.2 Mesures gaussiennes sur les espaces de Banach

La seconde application du théorème 3.2 concerne les mesures gaussiennes sur un espace de Banach de dimension infinie.

Théorème 3.4. *Soient $(E, \|\cdot\|)$ un espace de Banach, E^* son dual et μ une mesure gaussienne sur E . Pour $p \geq 1$, on pose*

$$\varphi(u) = \mu(B(0, u)) \text{ et } \sigma^2 = \sup_{\xi \in E^*, \|\xi\| \leq 1} \int \xi^2 d\mu.$$

Alors, pour tout $\zeta < 1/2$, il existe deux constantes $C_{\zeta,1}$ et $C_{\zeta,2} > 0$ qui dépendent seulement de $\varphi^{-1}(2^{-\zeta})$, ζ , p et σ telles que pour tout $n > 1$

$$EW_p^p(\mu, \mu_n) \leq C_1 \frac{E\|X\|^p}{\sqrt{n}} + C_{\zeta,1}(\varphi^{-1}(n^{-\zeta}))^p + C_{\zeta,2}(\log n^\zeta)^{p/2} n^{-1/2-\zeta}.$$

Autrement dit,

$$EW_p^p(\mu, \mu_n) = O(\varphi^{-1}(n^{-\zeta})^p).$$

Les constantes du théorème sont explicites dans la démonstration, et sont de taille raisonnable. Il est possible d'améliorer le résultat pour avoir la condition $\zeta < 1$ (au lieu de $\zeta < 1/2$), en modifiant le lemme 1.47, mais le gain en vitesse est négligeable pour la plupart des cas de dimension infinie pour lesquels $\varphi^{-1}(n^{-\zeta})$ est plus lent que $n^{-\zeta}$, pour tout $\zeta < 1/2$.

Ce résultat étend celui de [Dereich *et al.*, 2003] qui énonce la même vitesse de convergence (sous une hypothèse supplémentaire sur φ), pour la mesure empirique dont les poids en chaque Dirac sont modifiés pour obtenir une distance de Wasserstein à μ la plus petite possible.

La connaissance de φ pour certaines mesures gaussiennes nous permet alors de préciser cette majoration.

Exemple 3.5 (Mesure de Wiener sur $\mathcal{C}(0, 1)$). *Sur $(E, d) = (\mathcal{C}(0, 1), \|\cdot\|_\infty)$, l'espace des fonctions continues sur $[0; 1]$ muni de la norme uniforme, la mesure de Wiener μ vérifie*

$$\mu(B(0, u)) \sim C \exp\left(-\frac{\pi^2}{8u^2}\right),$$

pour une certaine constante $C > 0$. Le théorème 3.4 donne donc

$$EW_p(\mu_n, \mu) = O\left(\frac{1}{\sqrt{\log n}}\right).$$

L'ordre est optimal puisque [Dereich *et al.*, 2003] fournit une minoration du même ordre.

Exemple 3.6 (Mesure de Wiener sur \mathbf{L}^2). *Sur $(E, d) = (\mathbf{L}^2([0; 1]), \|\cdot\|_\infty)$, l'espace des fonctions \mathbf{L}^2 de $[0; 1]$, la mesure de Wiener μ vérifie*

$$\mu(B(0, u)) \sim \frac{4u}{\sqrt{\pi}} \exp(-\pi/8u^2).$$

Le théorème 3.4 donne donc à nouveau

$$EW_p(\mu_n, \mu) = O\left(\frac{1}{\sqrt{\log n}}\right).$$

L'ordre est encore optimal puisque [Dereich *et al.*, 2003] fournit une minoration du même ordre.

Exemple 3.7 (Brownien fractionnaire). *Sur $(E, d) = (\mathcal{C}([0; 1]), \|\cdot\|_\infty)$, l'espace des fonctions continues sur $[0; 1]$ muni de la norme uniforme, la mesure μ du mouvement brownien fractionnaire d'exposant de Hurst h vérifie (voir [Graf et al., 2003])*

$$-\log(\mu(B(0, u))) \sim c_h u^{-1/h}.$$

Le théorème 3.4 donne donc à nouveau

$$EW_p(\mu_n, \mu) = O\left(\frac{1}{(\log n)^h}\right).$$

[Graf et al., 2003] montre que cet ordre de convergence est optimal.

D'autres exemples de fonctions φ connues se trouvent facilement dans la littérature.

3.4 Retour sur la localisation de masse

Il est possible d'établir un lien avec la localisation de masse du chapitre 2. L'espérance de la puissance p de la distance de Wasserstein entre une mesure et sa version empirique peut être majorée par une quantité qui est un τ tel que dans le chapitre 2. La proposition suivante établit ce lien.

Théorème 3.8. *Soit (E, d, μ) un espace métrique mesuré. On note \mathcal{F} l'ensemble des fermés de E . On pose*

$$\tau_n^\alpha = \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{\mathcal{N}(K, u^{1/p}) - 1} du; \mu(K) \geq 1 - \alpha, t > 0, K \in \mathcal{F} \right\}. \quad (3.2)$$

Alors, pour tout $x_0 \in E$ et $n \geq 1$,

$$EW_p^p(\mu, \mu_n) \leq C_1 \frac{Ed(x_0, X)^p}{\sqrt{n}} + C_2 \int_0^{1/2} \frac{\tau_n^\alpha}{\sqrt{\alpha}} d\alpha.$$

où $C_1 = 2^{p-1}(1 + \sqrt{2})$ et $C_2 = 2^{3p+1} + 2^p$.

Ce résultat présente l'avantage de s'appliquer à tous les espaces métriques et à toutes les mesures telles qu'il existe $\varepsilon > 0$ tel que $Ed(x_0, X)^{2p+\varepsilon} < \infty$ pour un quelconque $x_0 \in E$.

Remarque 3.9. *L'intégrale $\int_0^{1/2} \frac{\tau_n^\alpha}{\sqrt{\alpha}} d\alpha$ est, en effet, finie dans ce cas. Pour s'en convaincre, il suffit de prendre $K = B(x_0, r_\alpha)$ dans l'équation 3.2 qui définit τ_n^α ,*

de sorte que $\mu(K) \geq 1 - \alpha$ et $t = r_\alpha^p$, pour voir ainsi que $\tau_n^\alpha \leq r_\alpha^p$. Or, l'inégalité de Markov montre

$$\mu(B(x_0, r_\alpha)) \geq 1 - Ed(x_0, X)^{2p+\varepsilon} / r_\alpha^{2p+\varepsilon},$$

et donc que l'on peut choisir

$$r_\alpha^p = [Ed(x_0, X)^{2p+\varepsilon}]^{1/2-\eta} \alpha^{-1/2+\eta},$$

où $\eta = \frac{\varepsilon}{4p+2\varepsilon} > 0$. Et ainsi,

$$\int_0^{1/2} \frac{\tau_n^\alpha}{\sqrt{\alpha}} d\alpha \leq [Ed(x_0, X)^{2p+\varepsilon}]^{1/2-\eta} \int_0^{1/2} \alpha^{-1+\eta} d\alpha < \infty.$$

La démonstration du théorème 3.4 sur le cas gaussien est faite dans cette thèse en appliquant ce théorème 3.8. Cela montre que ce résultat est utilisable, ce qui n'est pas évident au premier abord. Cela montre aussi qu'il est précis, puisque nous avons vu (en 3.3) que les vitesses de convergence issues du théorème 3.4 sont optimales dans différents cas.

Il est aussi possible de retrouver le cas de la dimension finie (théorème 3.3) à partir de ce résultat, ce qui n'est pas étonnant puisque sa démonstration imite celle du cas de la dimension finie. On obtient alors des conditions légèrement affaiblies sur le moment ainsi que de meilleures constantes.

Remarque 3.10. τ_n est majorée par (et proche de, voir proposition 1.4) la fonction de taille

$$K \mapsto \hat{\tau}_n(K) = \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{\mathcal{M}(K, u^{1/p}) - 1} du; t > 0 \right\},$$

où l'on a remplacé le recouvrement par le packing. Cette fonction de taille vérifie les hypothèses (H1), (H2) et (H3) du chapitre 2. Une démonstration de cette remarque est faite en section 3.8.

3.5 Déviations

La majoration de l'espérance de la distance de Wasserstein entre les mesures μ et μ_n permet d'obtenir des inégalités de déviation par simple application du théorème 1.53, dès qu'une inégalité de transport est vérifiée par une mesure. Les inégalités de transport considérées ici sont celles présentées au chapitre 1 au paragraphe 1.4.4.

3.5.1 Cas des espaces de dimension finie

On considère ici le cas des espaces de dimension d finie, dans le sens où l'inégalité (3.1) est vérifiée.

Théoreme 3.11 (Déviations sur un espace de dimension finie). *Soit (E, d) un espace vérifiant (3.1) pour un $d > 0$. Soit μ une mesure sur E vérifiant une inégalité de transport $T_p(K)$ pour un certain $K > 0$. Soit $\delta > dp/(d - p) \vee 2p$. On suppose $p < d/2$. Alors, il existe $C > 0$ dépendant seulement de p, d , et $Ed(x_0, X)^\delta$, tel que pour tout $\lambda > 0$ et tout $n \geq 1$,*

$$P\left(W_p(\mu, \mu_n) \leq (C + \lambda)n^{-1/d}\right) \geq 1 - \exp(-n^{1-2/d}\lambda^2/K).$$

3.5.2 Cas des mesures gaussiennes

Le théorème 1.54 assure que les mesures gaussiennes définies sur un espace de Banach vérifient une inégalité $T_p(2\sigma^2)$. Le théorème 3.4 montre alors le théorème suivant, pour lequel nous reprenons les notations

$$\varphi(u) = \mu(B(0, u)) \text{ et } \sigma^2 = \sup_{\xi \in E^*, \|\xi\| \leq 1} \int \xi^2 d\mu,$$

pour $p \geq 1$.

Théoreme 3.12 (Déviation pour les mesures gaussiennes). *Soient $(B, \|\cdot\|)$ un espace de Banach, et μ une mesure gaussienne sur B . Alors, pour tout $\zeta < 1/2$, il existe C dépendant seulement de $p, \sigma, Ed(x_0, X)^p, \zeta$ et $\varphi^{-1}(2^{-\zeta})$, tel que pour tout $\lambda > 0$, et tout $n \geq 1$,*

$$P\left(W_p(\mu, \mu_n) \leq (C + \lambda)\varphi^{-1}(n^{-\zeta})\right) \geq 1 - \exp(-n\lambda^2\varphi^{-1}(n^{-\zeta})^2/2\sigma^2).$$

La constante C n'est pas précisée, pour simplifier la présentation du théorème ; mais il est possible de la calculer à partir du théorème 3.4.

3.5.3 Cas général

Il est possible d'utiliser les mêmes arguments pour le cas général d'un espace polonais, à l'aide du théorème 3.8, ce qui donne alors le théorème suivant.

Théoreme 3.13 (Déviation dans un espace polonais). *Sur un espace (E, d, μ) polonais mesuré, tel que μ vérifie une inégalité de transport $T_p(K)$ pour un certain $K > 0$, on définit τ_n^α comme dans (3.2) et on pose*

$$t_n = C_1 \frac{Ed(x_0, X)^p}{\sqrt{n}} + C_2 \int_0^{1/2} \frac{\tau_n^\alpha}{\sqrt{\alpha}} d\alpha, \quad (3.3)$$

où $C_1 = 2^{p-1}(1 + \sqrt{2})$ et $C_2 = 2^{3p+1} + 2^p$.

Alors, pour tout $\lambda > 0$,

$$P(W_p(\mu, \mu_n) \geq (1 + \lambda)t_n) \leq e^{-n(\lambda t_n)^2/K}.$$

3.6 Cas dépendant

Les résultats obtenus jusqu'ici peuvent s'étendre, en utilisant les mêmes techniques, à certains cas où la suite $(X_i)_{i \geq 1}$ n'est pas indépendante. Nous allons nous intéresser ici aux chaînes de Markov vérifiant une certaine propriété de contraction.

Pour une chaîne de Markov de transition P et de mesure invariante μ , nous dirons que la chaîne vérifie une inégalité de décroissance de la variance si, pour toute fonction f bornée mesurable,

$$\text{Var}_\mu P^n f \leq C \lambda^n \text{Var}_\mu f. \quad (3.4)$$

Nous ferons cette hypothèse sur les chaînes que nous considérerons, bien que ces résultats (ou des résultats similaires) puissent être obtenus à partir d'hypothèses plus faibles.

Le premier résultat que nous énonçons concerne le cas où l'espace métrique sur lequel vit la chaîne de Markov est borné.

Théorème 3.14 (Cas à support borné). *Soit μ la mesure de probabilité invariante d'une chaîne de Markov $(X_i)_{i \geq 1}$ de transition P sur un espace (E, d) métrique borné par $D < \infty$. On note ν la loi de X_1 , que l'on suppose absolument continue par rapport à μ . Supposons que P vérifie l'inégalité (3.4), pour $C > 0$ et $\lambda > 0$. Alors, la mesure d'occupation*

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

vérifie, pour tout $t > 0$, et tout $r > 1$

$$E_\nu W_p^p(\mu, \mu_n) \leq 2^p \left[2^{p+1}t + \frac{C_r}{\sqrt{n}} \int_t^\infty (\mathcal{N}(E, u^{1/p}) - 1)^{1/2+1/r} du \right],$$

$$\text{où } C_r = \frac{\sqrt{2C}}{\sqrt{1-\lambda}} \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)}^{1/2}.$$

De là, en reprenant la démonstration du théorème 3.8, il est possible d'obtenir un théorème dans un cas général, sous condition de finitude de moment.

Théoreme 3.15 (Cas général). *Soit μ la mesure de probabilité invariante d'une chaîne de Markov $(X_i)_{i \geq 1}$ de transition P sur un espace (E, d) métrique. On note ν la loi de X_1 , que l'on suppose absolument continue par rapport à μ . Supposons que P vérifie l'inégalité (3.4), pour $C > 0$ et $\lambda > 0$. Alors, la mesure d'occupation*

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

vérifie, pour tout $t > 0$, et tout $r > 1$

$$E_\nu W_p^p(\mu, \mu_n) \leq 2^{p-1} C_r \left(\frac{Ed(x_0, X)^p}{\sqrt{n}(2^{\frac{2r}{r-1}} - 1)} + (2^{2p+2} + 2) \int_0^{1/2} \frac{\tau_n^{\alpha, r}}{\alpha^{1/2+1/r}} d\alpha \right),$$

$$\text{où } C_r = \frac{\sqrt{2C}}{\sqrt{1-\lambda}} \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)}^{1/2} \text{ et}$$

$$\tau_n^{\alpha, r} = \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \left(\mathcal{N}(K, u^{1/p}) - 1 \right)^{1/2+1/r} du; \mu(K) \geq 1 - \alpha, t > 0 \right\}.$$

Une application de ces deux précédents résultats aux cas des espaces de dimension finie, permet d'obtenir les deux résultats suivants.

Théoreme 3.16 (Cas à support borné de dimension finie). *Soit μ la mesure de probabilité invariante d'une chaîne de Markov $(X_i)_{i \geq 1}$ de transition P sur un espace (E, d) métrique borné par D , de dimension $d < \infty$ (dans le sens de (3.1)). On note ν la loi de X_1 , que l'on suppose absolument continue par rapport à μ . Supposons que P vérifie l'inégalité (3.4), pour $C > 0$ et $\lambda > 0$. Alors, la mesure d'occupation*

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

vérifie, pour une certaine constante $C_r > 0$, qui dépend de $r > 0, p \geq 1$, du diamètre D et de $\left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)}$,

$$E_\nu W_p^p(\mu, \mu_n) \leq C_r n^{-(p/d(1+1/2r))}.$$

Théoreme 3.17 (Cas en dimension finie et moment fini). *Soit μ la mesure de probabilité invariante d'une chaîne de Markov $(X_i)_{i \geq 1}$ de transition P sur un espace (E, d) métrique de dimension $d < \infty$ (dans le sens de (3.1)). On note ν la loi de X_1 , que l'on suppose absolument continue par rapport à μ . Supposons que P vérifie l'inégalité (3.4), pour $C > 0$ et $\lambda > 0$. Pour X de loi μ , on suppose*

$$M_q = Ed(x_0, X)^q < \infty,$$

pour un certain $x_0 \in E$ et $q > p/(1/2 - 1/r)$. Alors, la mesure d'occupation

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

vérifie, pour une certaine constante $C_r > 0$, qui dépend de $q, M_q, r > 0, p \geq 1$ et de $\|\frac{d\nu}{d\mu}\|_{L^r(\mu)}$,

$$E_\nu W_p^p(\mu, \mu_n) \leq C_r n^{-(p/d(1+1/2r))}.$$

Il est possible d'étendre ces résultats à d'autres cas de dépendance en suivant la même technique de démonstration. La clé est d'avoir un contrôle de la corrélation entre les $(X_i)_{i \geq 1}$; il faut pouvoir majorer, pour tout $i, j \geq 1$,

$$\frac{\text{Cov}(\mathbf{1}_A(X_i), \mathbf{1}_A(X_j))}{\sqrt{\text{Var}(\mathbf{1}_A(X_i))\text{Var}(\mathbf{1}_A(X_j))}}.$$

3.7 Applications statistiques

Ces résultats ont des applications statistiques directes simples que nous exposons ici.

3.7.1 Test statistique

Dans le cas d'un espace (E, d) de dimension finie, au sens de 3.1, sur lequel est tiré un échantillon de loi μ inconnue, nous voulons tester l'hypothèse

$$(H_0) : \mu = \mu_0,$$

pour une loi $\mu_0 \in \mathcal{P}_p(E)$ fixée, contre l'hypothèse alternative

$$(H_1) : \mu \neq \mu_0.$$

Le théorème 3.11 affirme alors que si μ_0 vérifie une inégalité de transport $T_p(K)$ pour un certain $K > 0$ et qu'il existe $\delta > dp/(d-p) \vee 2p$ tel que $p < d/2$ et $Ed(x_0, X)^\delta < \infty$, alors, sous (H_0) , il existe $C > 0$ (que l'on peut calculer à partir de la démonstration du théorème), tel que pour tout $\lambda > 0$ et tout $n \geq 1$,

$$P\left(W_p(\mu_0, \mu_n) \leq (C + \lambda)n^{-1/d}\right) \geq 1 - \exp(-n^{1-2/d}\lambda^2/K).$$

En posant alors, pour $0 < \alpha < 1$,

$$\lambda_n = \sqrt{-n^{-1+2/d}K \log \alpha},$$

le test qui consiste à rejeter (H_0) lorsque

$$W_p(\mu_0, \mu_n) \geq (C + \lambda_n)n^{-1/d},$$

est un test de risque inférieur à α .

De la même façon, on peut construire un test statistique dans le cadre général d'un espace polonais pour une mesure vérifiant une inégalité de transport $T_p(K)$.

Pour $\lambda_n = \sqrt{-\frac{K}{n} \log \alpha}$ et t_n défini par (3.3), on rejettera alors l'hypothèse (H_0) lorsque

$$W_p(\mu_0, \mu_n) \geq t_n + \lambda_n,$$

avec un risque inférieur à α .

Ces tests non paramétriques présentent l'avantage d'être exacts (i.e. non asymptotiques).

3.7.2 Régions de confiance

Sur un espace polonais (E, d) , on cherche à estimer une mesure μ que l'on sait vérifier une inégalité de transport $T_p(K)$. On observe un échantillon i.i.d. de loi μ et on estime μ par la mesure empirique μ_n . Pour $0 < \alpha < 1$, $\lambda_n = \sqrt{-\frac{K}{n} \log \alpha}$, et t_n défini par (3.3), on sait alors qu'avec probabilité au moins $1 - \alpha$,

$$W_p(\mu, \mu_n) \leq t_n + \lambda_n.$$

Autrement dit, la boule de centre μ_n et de rayon $t_n + \lambda_n$ dans l'espace de Wasserstein est une région de confiance de niveau $1 - \alpha$ pour la mesure μ .

3.8 Démonstrations

3.8.1 Première majoration

La démonstration du théorème 3.2 commence par celle des deux lemmes suivants.

Lemme 3.18. *Soient (E, d) un espace métrique de diamètre D tel que $n(k) := \mathcal{N}(E, D\varepsilon^k)$ soit fini pour un certain $k \in \mathbb{N}$ et $0 < \varepsilon < 1$. On pose $B_1^0 = E$, alors, il existe un ensemble de $k + 1$ partitions $\{B_i^j\}_{1 \leq i \leq n(j)} : 0 \leq j \leq k$ tel que pour tout $1 \leq j \leq k$ et $1 \leq i \leq n(j)$,*

- (1) *il existe $1 \leq l \leq n(j - 1)$ tel que $B_i^j \subset B_l^{j-1}$*
- (2) *et il existe x_i^j tel que pour tout $x \in B_i^j$, $d(x, x_i^j) \leq D\varepsilon^j / (1 - \varepsilon)$.*

Démonstration du lemme 3.18. Puisque $n(k)$ est fini, alors pour tout $1 \leq j \leq k$, il existe un ensemble $\{x_i^j\}_{1 \leq i \leq n(j)}$ tel que

$$E \subset \bigcup_{i=1}^{n(j)} B(x_i^j, D\varepsilon^j).$$

On définit pour tout $0 \leq j \leq k$,

$$V_0^j = \emptyset, \text{ et pour tout } 1 \leq i \leq n(j), V_i^j = \bigcap_{1 \leq l \leq n(j)} \{x \mid d(x, x_i^j) \leq d(x, x_l^j)\} \setminus V_{i-1}^j.$$

Il est à noter que pour tout $0 \leq j \leq k$ et tout $1 \leq i \leq n(j)$, $V_i^j \subset B(x_i^j, D\varepsilon^j)$.

On pose ensuite, pour tout $1 \leq i \leq n(k)$,

$$B_i^k = V_i^k,$$

et pour tout $0 \leq j \leq k-1$ et tout $1 \leq i \leq n(j)$,

$$B_i^j = \bigcup_{x_l^{j+1} \in V_i^j} B_l^{j+1}.$$

On vérifie alors que les $\{B_i^j\}_{1 \leq i \leq n(j)} : 0 \leq j \leq k$ définis ci-dessus satisfont les deux conditions (1) et (2). La première est évidente d'après la définition. Pour la seconde, on remarque d'abord que pour tout $x \in B_i^k$

$$d(x, x_i^k) \leq D\varepsilon^k.$$

Ensuite, pour tout $0 \leq j \leq k-1$, pour tout $x \in B_i^j$ il existe l tel que $x \in B_l^{j+1} \subset B_i^j$ de sorte que

$$\begin{aligned} d(x, x_i^j) &\leq d(x, x_l^{j+1}) + d(x_l^{j+1}, x_i^j) \\ &\leq \sum_{i=j+1}^k D\varepsilon^i + D\varepsilon^j \\ &= D \sum_{i=j}^k \varepsilon^i \\ &\leq D \frac{\varepsilon^j}{1 - \varepsilon} \end{aligned}$$

où la seconde inégalité s'établit par induction sur j et puisque $x_l^{j+1} \in V_i^j$. □

Le prochain lemme est central dans la démonstration du théorème 3.2.

Lemme 3.19. Soient μ et ν deux mesures de probabilité sur un espace métrique (E, d) de diamètre D . Soit $(B_i^j)_{i,j}$ une suite de partitions telle que dans le lemme 3.18. Alors,

$$W_p^p(\mu, \nu) \leq 2^p D^p \left[2^{1-kp} + \sum_{l=1}^k (2^{-(l-1)p} - 2^{-lp}) \sum_{j=1}^{n(l)} |\mu(B_j^l) - \nu(B_j^l)| \right].$$

Démonstration du lemme 3.19. En utilisant le lemme 3.18 avec $\varepsilon = 1/2$ et la notation $\mu \wedge \nu(B) = \mu(B) \wedge \nu(B)$, on définit la mesure suivante :

$$\mu\nu^k = \sum_{j=0}^{k-1} \sum_{i=1}^{n(j)} [\mu \wedge \nu(B_i^j) - \sum_{x_i^{j+1} \in B_i^j} \mu \wedge \nu(B_i^{j+1})] \delta_{x_i^j} + \sum_{i=1}^{n(k)} \mu \wedge \nu(B_i^k) \delta_{x_i^k}.$$

Soit

$$\mu|B = \begin{cases} \frac{\mu(\cdot \cap B)}{\mu(B)} & \text{si } \mu(B) > 0 \\ 0 & \text{sinon.} \end{cases}$$

On définit alors le couplage (X, Z) comme suit.

$$Z \sim \mu\nu^k,$$

puis pour tout $0 \leq l \leq k-1$ et tout $1 \leq i \leq n(l)$,

$$\mathcal{L}(X|Z = x_i^l) = \sum_{x_j^{l+1} \in B_i^l} \frac{\mu(B_j^{l+1}) - \mu \wedge \nu(B_j^{l+1})}{\mu(B_i^l) - \sum_{x_j^{l+1} \in B_i^l} \mu \wedge \nu(B_j^{l+1})} \mathcal{L}(X|Z = x_j^{l+1}),$$

et pour tout $1 \leq i \leq n(k)$,

$$\mathcal{L}(X|Z = x_i^k) = \mu|B_i^k.$$

On vérifie par induction - sur le premier B_j^l tel que $\mu(B_j^l) \leq \nu(B_j^l)$ - que $\mathcal{L}(X) = \mu$.

Puisque (X, Z) a pour marginales μ et $\mu\nu^k$, la définition de la distance de Wasserstein permet d'obtenir la majoration $W_p^p(\mu, \mu\nu^k) \leq Ed(X, Z)^p$. Puis,

puisque $X \in B_i^l$ presque sûrement si $Z = x_i^l$, il vient

$$\begin{aligned}
Ed(X, Z)^p &= \sum_{0 \leq l \leq k, 1 \leq j \leq n(l)} P(Z = x_j^l) E(d(X, Z)^p | Z = x_j^l) \\
&\leq 2^p D^p \left[\sum_{l=0}^{k-1} \sum_{j=1}^{n(l)} [\mu \wedge \nu(B_j^l) - \sum_{x_i^{l+1} \in B_j^l} \mu \wedge \nu(B_i^{l+1})] 2^{-lp} + \sum_{j=1}^{n(k)} \mu \wedge \nu(B_j^k) 2^{-kp} \right] \\
&= 2^p D^p \left[1 - \sum_{l=1}^k \sum_{j=1}^{n(l)} \mu \wedge \nu(B_j^l) (2^{-(l-1)p} - 2^{-lp}) \right] \\
&\leq 2^p D^p \left[2^{-kp} + \sum_{l=1}^k \sum_{j=1}^{n(l)} [\mu(B_j^l) - \mu \wedge \nu(B_j^l)] (2^{-(l-1)p} - 2^{-lp}) \right].
\end{aligned}$$

Comme il en va de même pour $W(\mu\nu^k, \nu)$, il s'ensuit

$$W_p^p(\mu, \nu) \leq 2^p D^p \left[2^{1-kp} + \sum_{l=1}^k \sum_{j=1}^{n(l)} |\mu(B_j^l) - \nu(B_j^l)| (2^{-(l-1)p} - 2^{-lp}) \right].$$

□

On peut maintenant démontrer le théorème 3.2.

Démonstration du théorème 3.2. On applique le lemme 3.19 à $W_p^p(\mu, \mu_n)$, il vient alors, en prenant l'espérance,

$$\begin{aligned}
EW_p^p(\mu, \mu_n) &\leq 2^p D^p \left[2^{1-kp} + \sum_{l=1}^k (2^{-(l-1)p} - 2^{-lp}) \sum_{j=1}^{n(l)} E|\mu(B_j^l) - \mu_n(B_j^l)| \right] \\
&\leq 2^p D^p \left[2^{1-kp} + \sum_{l=1}^k (2^{-(l-1)p} - 2^{-lp}) \sum_{j=1}^{n(l)} \sqrt{E[\mu(B_j^l) - \mu_n(B_j^l)]^2} \right] \\
&\leq 2^p D^p \left[2^{1-kp} + \sum_{l=1}^k (2^{-(l-1)p} - 2^{-lp}) \sum_{j=1}^{n(l)} \frac{\sqrt{\mu(B_j^l)(1 - \mu(B_j^l))}}{\sqrt{n}} \right] \\
&\leq 2^p D^p \left[2^{1-kp} + \sum_{l=1}^k (2^{-(l-1)p} - 2^{-lp}) \frac{\sqrt{n(l) - 1}}{\sqrt{n}} \right] \\
&\leq 2^p \left[D^p 2^{1-kp} + \frac{1}{\sqrt{n}} \int_{D^p 2^{-kp}}^{\infty} \sqrt{\mathcal{N}(K, u^{1/p}) - 1} du \right].
\end{aligned}$$

La démonstration se conclut en choisissant $t \in]2^{-p} D^p 2^{-kp}, D^p 2^{-kp}]$.

□

3.8.2 Espaces de dimension finie

Démonstration du théorème 3.3. Pour deux mesures μ et ν , on note $\Gamma(\mu, \nu)$ l'ensemble des mesures de première marginale μ et de seconde marginale ν . Soit $\{K_i\}_{i \geq 1}$ une partition de E . On choisit les plans de transport suivants

$$\begin{aligned}\pi_{i,n}^0 &\in \arg \min \left\{ \int d_E^p(x, y) d\pi(x, y) \mid \pi \in \Gamma(\mu_n|K_i, \mu|K_i) \right\}, \\ \hat{\pi}_{i,n}^0 &\in \arg \min \left\{ \int d_E^p(x, y) d\pi(x, y) \mid \pi \in \Gamma(\mu_n|K_i, \mu) \right\}, \\ \pi_n^1 &= (Id, Id)_{\#} \mu^n, \\ \hat{\pi}_i^1 &\in \arg \min \left\{ \int d_E^p(x, y) d\pi(x, y) \mid \pi \in \Gamma(\mu, \mu|K_i) \right\},\end{aligned}$$

ainsi que les mesures suivantes

$$\begin{aligned}\mu^n &= \sum_{i \geq 1} \frac{\mu_n(K_i) \wedge \mu(K_i)}{\left(\sum_{j \geq 1} \mu_n(K_j) \wedge \mu(K_j) \right)} \mu|K_i, \\ \hat{\mu}^n &= \left(\sum_{j \geq 1} \mu_n(K_j) \wedge \mu(K_j) \right) \mu^n + \left(1 - \sum_{j \geq 1} \mu_n(K_j) \wedge \mu(K_j) \right) \mu.\end{aligned}$$

On définit alors les plans de transport suivants

$$\begin{aligned}\Gamma(\mu_n, \hat{\mu}^n) \ni \Pi_n^0 &= \sum_{i \geq 1} \mu_n(K_i) \wedge \mu(K_i) \pi_{i,n}^0 \\ &\quad + \sum_{i \geq 1} (\mu_n(K_i) - \mu(K_i))^+ \hat{\pi}_{i,n}^0, \\ \Gamma(\hat{\mu}^n, \mu) \ni \Pi_n^1 &= \sum_{i \geq 1} \mu_n(K_i) \wedge \mu(K_i) \pi_n^1 \\ &\quad + \sum_{i \geq 1} (\mu(K_i) - \mu_n(K_i))^+ \hat{\pi}_i^1.\end{aligned}$$

Ainsi, en utilisant l'inégalité $2^{1-p}(a+b)^p \leq a^p + b^p$ due à la convexité de $x \mapsto x^p$ et l'inégalité triangulaire de la distance de Wasserstein $W_p(\mu_n, \mu) \leq W_p(\mu_n, \hat{\mu}^n) + W_p(\hat{\mu}^n, \mu)$ - avec la convention $\mu_n(K_i) W_p^p(\mu_n|K_i, \cdot) = 0$ si $\mu_n(K_i) = 0$,

$$\begin{aligned}2^{1-p} W_p^p(\mu_n, \mu) &\leq \int d_E^p(x, y) d\Pi_n^0(x, y) + \int d_E^p(x, y) d\Pi_n^1(x, y) \\ &\leq \sum_{i \geq 1} \mu_n(K_i) \wedge \mu(K_i) W_p^p(\mu_n|K_i, \mu|K_i)\end{aligned}\tag{3.5}$$

$$+ \sum_{i \geq 1} (\mu_n(K_i) - \mu(K_i))^+ W_p^p(\mu_n|K_i, \mu)\tag{3.6}$$

$$+ \sum_{i \geq 1} (\mu(K_i) - \mu_n(K_i))^+ W_p^p(\mu, \mu|K_i).\tag{3.7}$$

On peut alors appliquer le théorème 3.2 à la mesure $\mu(\cdot|K_i)$ et la mesure empirique $\mu_n(\cdot|K_i)$ pour un échantillon de taille $n\mu_n(K_i)$. Il vient ainsi pour tout $(t_i)_{i \geq 1} \in \mathbb{R}_+^{\mathbb{N}}$,

$$\begin{aligned}
E(3.5) &\leq E \sum_{i \geq 1} \mu_n(K_i) W_p^p(\mu_n|K_i, \mu|K_i) \\
&\leq E 2^p \sum_{i \geq 1} \mu_n(K_i) \left[2^{p+1} t_i + \frac{1}{\sqrt{n\mu_n(K_i)}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(K_i, u^{1/p}) - 1} du \right] \\
&\leq 2^{2p+1} \sum_{i \geq 1} \mu(K_i) t_i + 2^p \sum_{i \geq 1} \frac{E \sqrt{\mu_n(K_i)}}{\sqrt{n}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(K_i, u^{1/p}) - 1} du \\
&\leq 2^{2p+1} \sum_{i \geq 1} \mu(K_i) t_i + 2^p \sum_{i \geq 1} \frac{\sqrt{\mu(K_i)}}{\sqrt{n}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(K_i, u^{1/p}) - 1} du.
\end{aligned}$$

Soit $r > 1$. On pose $K_i = B(x_0, r^i) \setminus B(x_0, r^{i-1})$ pour $i \geq 2$ et $K_1 = B(x_0, r)$, pour $x_0 \in E$ quelconque, de sorte que, pour $d > 2p$, $\delta > 0$ et $i \geq 2$, en utilisant l'inégalité de Markov et l'hypothèse de recouvrement (3.1),

$$\mu(K_i) \leq \frac{Ed(x_0, X)^\delta}{r^{\delta(i-1)}} \text{ et } \int_{t_i}^{\infty} \sqrt{\mathcal{N}(K_i, u^{1/p}) - 1} du \leq r^{id/2} t_i^{1-d/2p} / (d/2p - 1).$$

Ainsi,

$$\begin{aligned}
E(3.5) &\leq 2^{2p+1} \sum_{i \geq 1} \mu(K_i) t_i + 2^p \sum_{i \geq 1} \frac{\sqrt{\mu(K_i)}}{\sqrt{n}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(K_i, u^{1/p}) - 1} du \\
&\leq 2^{2p+1} \sum_{i \geq 2} \frac{Ed(x_0, X)^\delta}{r^{\delta(i-1)}} t_i + 2^p \frac{(Ed(x_0, X)^\delta)^{1/2}}{r^{\delta(i-1)/2} \sqrt{n}} r^{id/2} t_i^{1-d/2p} / (d/2p - 1) \\
&\quad + 2^{2p+1} t_1 + \frac{2^p}{\sqrt{n}} r^{d/2} t_1^{1-d/2p} / (d/2p - 1).
\end{aligned}$$

On pose

$$\begin{aligned}
t_i &= \left(2^{2(p+1)} Ed(x_0, X)^\delta r^\delta n r^{-i(\delta+d)} \right)^{-p/d} \text{ pour } i \geq 2, \text{ et} \\
t_1 &= \left(2^{2(p+1)} n r^{-d} \right)^{-p/d}.
\end{aligned}$$

Ainsi, pour $\delta > dp/(d-p)$,

$$\begin{aligned}
E(3.5) &\leq C_1 \sum_{i \geq 2} r^{i(-\delta+\delta p/d+p)} n^{-p/d} + C_2 r^p n^{-p/d} \\
&\leq C_3 n^{-p/d},
\end{aligned}$$

où

$$\begin{aligned} C_1 &= \left(Ed(x_0, X)^\delta r^\delta \right)^{1-p/d} 2^{2p+1-2p(p+1)/d} \left(\frac{d}{d-2p} \right) \\ C_2 &= 2^{2p+1-2p(p+1)/d} \left(\frac{d}{d-2p} \right) \\ C_3 &= C_1 r^{(-\delta+\delta p/d+p)} / (r^{(-\delta+\delta p/d+p)} - 1) + C_2 r^p. \end{aligned}$$

Pour établir une majoration de (3.6) et (3.7), on remarque que

$$W_p^p(\mu_n|K_i, \mu) \vee W_p^p(\mu|K_i, \mu) \leq Ed(x_0, X)^p + r^{ip},$$

de sorte que

$$\begin{aligned} E(3.6 + 3.7) &\leq \sum_{i \geq 1} E|\mu_n(K_i) - \mu(K_i)|(Ed(x_0, X)^p + r^{ip}) \\ &= \frac{1}{n} \sum_{i \geq 1} E|n\mu_n(K_i) - n\mu(K_i)|(Ed(x_0, X)^p + r^{ip}) \\ &\leq \frac{1}{n} \sum_{i \geq 1} \sqrt{E(n\mu_n(K_i) - n\mu(K_i))^2} (Ed(x_0, X)^p + r^{ip}) \end{aligned} \quad (3.8)$$

$$\leq \frac{1}{\sqrt{n}} \sum_{i \geq 1} \sqrt{\mu(K_i)} (Ed(x_0, X)^p + r^{ip}) \quad (3.9)$$

$$\begin{aligned} &\leq \frac{1}{\sqrt{n}} \sum_{i \geq 2} \frac{\sqrt{Ed(x_0, X)^\delta}}{r^{(i-1)\delta/2}} (Ed(x_0, X)^p + r^{ip}) \\ &\quad + \frac{1}{\sqrt{n}} (Ed(x_0, X)^p + r^p), \end{aligned} \quad (3.10)$$

où (3.8) est une application de l'inégalité de Jensen à la fonction $x \mapsto x^2$, (3.9) est une majoration de la variance d'une somme de variable de Bernoulli, et enfin (3.10) s'obtient par application de l'inégalité de Markov à $\mu(K_i)$.

On a donc montré

$$E(3.6 + 3.7) \leq \frac{C_4}{\sqrt{n}},$$

où - puisque $\delta > 2p$,

$$C_4 = \sqrt{Ed(x_0, X)^\delta} \left(Ed(x_0, X)^p \frac{r^{-\delta/2}}{1 - r^{-\delta/2}} + \frac{r^{2p-\delta/2}}{1 - r^{p-\delta/2}} \right) + Ed(x_0, X)^p + r^p.$$

□

3.8.3 Mesures gaussiennes

Démonstration du théorème 3.4. Pour un ensemble K d'un espace vectoriel et un réel positif r , on note $rK = \{rx; x \in K\}$. On rappelle que pour tout $t > 0$,

$$\mathcal{N}(rK, t) = \mathcal{N}(K, t/r).$$

On note B la boule unité de E et \mathcal{K} la boule unité de l'espace de Hilbert à noyau auto reproduisant. La démonstration est une application du théorème 3.8 et cherche donc à majorer

$$\tau_n^\alpha = \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{\mathcal{N}(K, u^{1/p}) - 1} du; \mu(K) \geq 1 - \alpha, t > 0 \right\}.$$

Étape 1 : majoration de l'entropie métrique

Dans le cas infini dimensionnel, qui est celui pour lequel le résultat est intéressant, $\int_t^\infty \sqrt{\mathcal{N}(K, u^{1/p}) - 1} du$ est en général du même ordre que $D^p \sqrt{\mathcal{N}(K, t^{1/p})}$ quand t tend vers 0, où D est le diamètre de K . La majoration suivante, pour un borélien K tel que $\mu(K) \geq 1 - \alpha$, reste donc pertinente

$$\tau_n^\alpha \leq t + n^{-1/2} D^p \sqrt{\mathcal{N}(K, t^{1/p})}, \quad (3.11)$$

pour tout $t > 0$. Pour $r > 0$ et $\lambda > 0$, on choisit

$$K = r\mathcal{K} + \lambda B.$$

Ainsi, en utilisant le lemme 1.47 de Kuelbs et Li, on obtient pour $u > \lambda$

$$\begin{aligned} \mathcal{N}(K, u) &\leq \mathcal{N}(r\mathcal{K}, u - \lambda) \\ &\leq \left[\varphi \left(\psi^{-1} \left(\frac{u - \lambda}{r} \right) \right) \right]^{-2}, \end{aligned}$$

où, $\psi(t) = \frac{\sqrt{2}t}{\sqrt{-\log \varphi(t)}}$. Par ailleurs, le théorème 1.49 de Borell donne

$$\mu(K) \geq \Phi[\Phi^{-1}(\varphi(\lambda)) + r],$$

et on pose alors

$$\Phi[\Phi^{-1}(\varphi(\lambda)) + r] = 1 - \alpha. \quad (3.12)$$

Ainsi, pour une suite $(z_n)_{n \geq 1} \subset (0, 1)$, en posant $t^{1/p} = r\psi(\varphi^{-1}(z_n)) + \lambda > \lambda$, la majoration (3.11) devient

$$\frac{\tau_n^\alpha}{\sqrt{\alpha}} \leq \frac{[r\psi(\varphi^{-1}(z_n)) + \lambda]^p + D^p n^{-1/2} z_n^{-1}}{(1 - \Phi[\Phi^{-1}(\varphi(\lambda)) + r])^{1/2}}.$$

On choisit alors $r = 2\lambda/\psi(\varphi^{-1}(z_n))$, (λ est alors solution de $\Phi[\Phi^{-1}(\varphi(\lambda)) + 2\lambda/\psi(\varphi^{-1}(z_n))] = 1 - \alpha$) de sorte que

$$\tau_n^\alpha \leq [3\lambda]^p + D^p n^{-1/2} z_n^{-1}. \quad (3.13)$$

Étape 2 : majoration de λ

Afin de majorer le λ choisi, on remarque que, par croissance de $\Phi^{-1}(\varphi(\lambda)) + 2\lambda/\psi(\varphi^{-1}(z_n))$ en λ , s'il existe λ_0 tel que

$$\Phi^{-1}(1 - \alpha) \leq \Phi^{-1}(\varphi(\lambda_0)) + 2\lambda_0/\psi(\varphi^{-1}(z_n)), \quad (3.14)$$

alors $\lambda \leq \lambda_0$.

On pose alors $\lambda_1 = \varphi^{-1}(z_n)$ de sorte que $2\lambda_1/\psi(\varphi^{-1}(z_n)) = \sqrt{-2 \log \varphi(\lambda_1)}$. Ainsi, en posant $\lambda_0 = \lambda_1 + \Phi^{-1}(1 - \alpha)\psi(\varphi^{-1}(z_n))/2$,

$$\begin{aligned} \Phi^{-1}(1 - \alpha) &= -\sqrt{-2 \log \varphi(\lambda_1)} + 2\lambda_1/\psi(\varphi^{-1}(z_n)) + \Phi^{-1}(1 - \alpha) \\ &= -\sqrt{-2 \log \varphi(\lambda_1)} + 2\lambda_0/\psi(\varphi^{-1}(z_n)) \\ &\leq -\sqrt{-2 \log \varphi(\lambda_0)} + 2\lambda_0/\psi(\varphi^{-1}(z_n)) \end{aligned} \quad (3.15)$$

$$\leq \Phi^{-1}(\varphi(\lambda_0)) + 2\lambda_0/\psi(\varphi^{-1}(z_n)), \quad (3.16)$$

où l'inégalité (3.15) est due à la croissance de $-\sqrt{-\log \varphi}$ et (3.16) est due au lemme A.1. L'inégalité (3.14) est bien vérifiée, et donc

$$\lambda \leq \lambda_0 = \varphi^{-1}(z_n) + \Phi^{-1}(1 - \alpha)\psi(\varphi^{-1}(z_n))/2.$$

Étape 3 : finalisation

La majoration (3.13) devient alors

$$\tau_n^\alpha \leq 3^p [\varphi^{-1}(z_n) + \Phi^{-1}(1 - \alpha)\psi(\varphi^{-1}(z_n))/2]^p + D^p n^{-1/2} z_n^{-1}.$$

Or, le premier terme du membre de droite vérifie

$$\begin{aligned} 3^p [\varphi^{-1}(z_n) + \Phi^{-1}(1 - \alpha)\psi(\varphi^{-1}(z_n))/2]^p &= 3^p \left[\varphi^{-1}(z_n) + \Phi^{-1}(1 - \alpha) \frac{\varphi^{-1}(z_n)}{\sqrt{-2 \log z_n}} \right]^p \\ &= 3^p \varphi^{-1}(z_n)^p \left[1 + \Phi^{-1}(1 - \alpha) \frac{1}{\sqrt{-2 \log z_n}} \right]^p \end{aligned}$$

et le diamètre D de K vérifie

$$\begin{aligned} D &= 2r\sigma + 2\lambda \\ &= 2\lambda \left(\frac{2\sigma}{\psi(\varphi^{-1}(z_n))} + 1 \right) \\ &\leq 2 \left[\varphi^{-1}(z_n) + \Phi^{-1}(1 - \alpha)\psi(\varphi^{-1}(z_n))/2 \right] \left(\frac{2\sigma}{\psi(\varphi^{-1}(z_n))} + 1 \right) \\ &= 2 \left[\sqrt{2}\sigma + \varphi^{-1}(z_n)/\sqrt{-\log(z_n)} \right] \sqrt{-\log(z_n)} + \Phi^{-1}(1 - \alpha) [2\sigma + \psi(\varphi^{-1}(z_n))]. \end{aligned}$$

Ainsi, puisque $\int_0^{1/2} \alpha^{-1/2} d\alpha = \sqrt{2}$, en posant $C_p = \int_0^{1/2} \alpha^{-1/2} (\Phi^{-1}(1-\alpha))^p d\alpha$, il vient en utilisant la convexité de $x \mapsto x^p$

$$\begin{aligned} \int_0^{1/2} \frac{\tau_n^\alpha}{\sqrt{\alpha}} d\alpha &\leq 2^{p-1} 3^p \left[\sqrt{2} + \frac{C_p}{(-2 \log z_n)^{p/2}} \right] \varphi^{-1}(z_n)^p \\ &\quad + 2^{2p-1} \sqrt{2} \left[\sqrt{2} \sigma + \varphi^{-1}(z_n) / \sqrt{-\log(z_n)} \right]^p (-\log z_n)^{p/2} n^{-1/2} z_n^{-1} \\ &\quad + 2^{p-1} C_p \left[2\sigma + \psi(\varphi^{-1}(z_n)) \right]^p n^{-1/2} z_n^{-1}. \end{aligned}$$

En posant $z_n = n^{-\zeta}$ pour un certain $\zeta < 1/2$, le théorème 3.8 donne alors pour $n > 1$,

$$\begin{aligned} EW_p^p(\mu, \mu_n) &\leq C_1 \frac{Ed(0, X)^p}{\sqrt{n}} + C_2 \int_0^{1/2} \frac{\tau_n^\alpha}{\sqrt{\alpha}} d\alpha \\ &\leq C_1 \frac{E\|X\|^p}{\sqrt{n}} + C_{\zeta,1} (\varphi^{-1}(n^{-\zeta}))^p + C_{\zeta,2} (\log n^\zeta)^{p/2} n^{-1/2-\zeta}, \end{aligned}$$

où

$$\begin{aligned} C_1 &= 2^{p-1} (1 + \sqrt{2}), \\ C_2 &= 2^{3p+1} + 2^p, \\ C_{\zeta,1} &= C_2 2^{p-1} \left[3^p \sqrt{2} + C_p \frac{2^{p/2}}{(\zeta \log 2)^{p/2}} \right], \\ C_{\zeta,2} &= C_2 2^{2p-1} \left[\sqrt{2} \left[\sqrt{2} \sigma + \varphi^{-1}(2^{-\zeta}) (\zeta \log 2)^{-1/2} \right]^p \right], \\ &\quad + C_2 2^{p-1} \left[C_p (\zeta \log 2)^{-p/2} \left[2\sigma + \psi(\varphi^{-1}(2^{-\zeta})) \right]^p \right]. \end{aligned}$$

□

3.8.4 Localisation de masse

Démonstration du théorème 3.8. De même que dans la proposition 3.3, pour une partition $(K_i)_{i \geq 1}$ de E ,

$$2^{1-p} W_p^p(\mu_n, \mu) \leq \sum_{i \geq 1} \mu_n(K_i) \wedge \mu(K_i) W_p^p(\mu_n|K_i, \mu|K_i) \quad (3.17)$$

$$+ \sum_{i \geq 1} (\mu_n(K_i) - \mu(K_i))^+ W_p^p(\mu_n|K_i, \mu) \quad (3.18)$$

$$+ \sum_{i \geq 1} (\mu(K_i) - \mu_n(K_i))^+ W_p^p(\mu|K_i, \mu). \quad (3.19)$$

et,

$$\begin{aligned}
E(3.17) &\leq E \sum_{i \geq 1} \mu_n(K_i) W_p^p(\mu_n|K_i, \mu|K_i) \\
&\leq E 2^p \sum_{i \geq 1} \mu_n(K_i) \left[2^{p+1} t_i + \frac{1}{\sqrt{n \mu_n(K_i)}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(K_i, u^{1/p}) - 1} du \right] \\
&\leq 2^{2p+1} \sum_{i \geq 1} \mu(K_i) t_i + 2^p \sum_{i \geq 1} \frac{E \sqrt{\mu_n(K_i)}}{\sqrt{n}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(K_i, u^{1/p}) - 1} du \\
&\leq 2^{2p+1} \sum_{i \geq 1} \mu(K_i) t_i + 2^p \sum_{i \geq 1} \frac{\sqrt{\mu(K_i)}}{\sqrt{n}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(K_i, u^{1/p}) - 1} du \\
&\leq 2^{2p+1} \sum_{i \geq 1} \mu(K_i) t_i + 2^p \sum_{i \geq 1} \frac{\sqrt{\mu(K_i)}}{\sqrt{n}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1} du. \quad (3.20)
\end{aligned}$$

Pour majorer (3.18) et (3.19), on remarque que - pour x_0 tel que pour tout $i \geq 1$, $\sup_{x \in K_i} d(x, x_0) \leq \text{diam}(\cup_{j \leq i} K_j)$,

$$\begin{aligned}
W_p^p(\mu_n|K_i, \mu) \vee W_p^p(\mu|K_i, \mu) &\leq Ed(x_0, X)^p + \text{diam}(\cup_{j \leq i} K_j)^p \\
&\leq Ed(x_0, X)^p + t_i + \int_{t_i}^{\infty} \sqrt{\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1} du.
\end{aligned}$$

de sorte que, si $\mu(K_i) \leq 2^{-i}$,

$$\begin{aligned}
E(3.18 + 3.19) &\leq \sum_{i \geq 1} E |\mu_n(K_i) - \mu(K_i)| (Ed(x_0, X)^p + t_i + \int_{t_i}^{\infty} \sqrt{\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1} du) \\
&\leq \frac{1}{\sqrt{n}} \sum_{i \geq 1} \sqrt{\mu(K_i)} (Ed(x_0, X)^p + t_i + \int_{t_i}^{\infty} \sqrt{\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1} du) \\
&\leq \frac{1 + \sqrt{2}}{\sqrt{n}} Ed(x_0, X)^p \\
&\quad + \sum_{i \geq 1} 2^{-i/2} (t_i + \frac{1}{\sqrt{n}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1} du).
\end{aligned}$$

Ainsi,

$$\begin{aligned}
2^{1-p}EW(\mu, \mu_n) &\leq \frac{1 + \sqrt{2}}{\sqrt{n}} Ed(x_0, X)^p \\
&\quad + (2^{2p+1} + 1) \sum_{i \geq 1} 2^{-i/2} \left(t_i + \frac{1}{\sqrt{n}} \int_{t_i}^{\infty} \sqrt{\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1} du \right) \\
&\leq \frac{1 + \sqrt{2}}{\sqrt{n}} Ed(x_0, X)^p + (2^{2p+2} + 2) \int_0^{1/2} \frac{\tau_n^\alpha}{\sqrt{\alpha}} d\alpha.
\end{aligned}$$

□

Remarque 3.20. Une analyse de la démonstration permet d'identifier une perte dans la majoration de ce théorème : il s'agit essentiellement de la majoration

$$\mathcal{N}(K_k, t) \leq \mathcal{N}\left(\bigcup_{i \leq k} K_i, t\right),$$

faite en (3.20).

Démonstration de la remarque 3.10. Il est clair que

$$K \mapsto t + \frac{1}{\sqrt{n}} \int_t^{\infty} \sqrt{\mathcal{M}(K, u^{1/p}) - 1} du$$

vérifie les hypothèses (H1), (H2) et (H3), pour tout $t > 0$. On veut alors montrer que l'infimum sur $t > 0$ de cette fonction est une fonction qui vérifie encore ces hypothèses.

L'hypothèse de croissance (H1) reste clairement vraie en prenant l'infimum sur $t > 0$.

Pour montrer l'hypothèse (H2), on choisit une suite $(f_k)_{k \geq 1}$ décroissante de fonctions décroissantes (continues à droite) dans $\mathbb{R}^+ \rightarrow \mathbb{N}^* \cup \{+\infty\}$. Pour chaque $k \geq 1$, l'infimum en $t > 0$ est réalisé pour $t_k \in \mathbb{R}^+$. La suite $(t_k)_{k \geq 1}$ est alors décroissante (du fait de la décroissance de $(f_k)_{k \geq 1}$) et converge donc vers une limite $t_\infty \in \mathbb{R}^+$. Un rapide calcul à partir de l'optimalité de t_k montre que $f_k(t_k) \leq n + 1$ (on aurait sinon $(t_k + \varepsilon) + \frac{1}{\sqrt{n}} \int_{t_k + \varepsilon}^{\infty} \sqrt{f_k - 1} < t_k + \frac{1}{\sqrt{n}} \int_{t_k}^{\infty} \sqrt{f_k - 1}$ pour un $\varepsilon > 0$ suffisamment petit, contredisant l'optimalité de t_k) et donc que

$$f_k \mathbf{1}_{[t_k; +\infty[} \leq f_1 \wedge (n + 1),$$

ce qui permet d'appliquer le théorème de convergence dominée de Lebesgue de sorte que

$$\begin{aligned}
&\inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^{\infty} \sqrt{f_k - 1}; t > 0 \right\} \\
&= t_k + \frac{1}{\sqrt{n}} \int_{t_k}^{\infty} \sqrt{f_k - 1} \longrightarrow t_\infty + \frac{1}{\sqrt{n}} \int_{t_\infty}^{\infty} \sqrt{f_\infty - 1},
\end{aligned}$$

où $f_\infty = \lim_k f_k$. Par ailleurs, l'optimalité de t_k et le fait que $t_\infty \leq t_k$ montre que

$$n + 1 \leq \lim_{s \rightarrow t_\infty^+} f_k(s) \rightarrow \lim_{s \rightarrow t_\infty^+} f_\infty(s),$$

(puisque l'on aurait sinon $(t_k - \varepsilon) + \frac{1}{\sqrt{n}} \int_{t_k - \varepsilon}^\infty \sqrt{f_k - 1} < t_k + \frac{1}{\sqrt{n}} \int_{t_k}^\infty \sqrt{f_k - 1}$ pour un $\varepsilon > 0$ suffisamment petit) ce qui permet de montrer que t_∞ réalise le minimum sur $t > 0$ de

$$\inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{f_\infty - 1}; t > 0 \right\},$$

et montre donc (H2). Pour montrer (H3), on choisit une suite $(f_k)_{k \geq 1}$ de fonctions décroissantes continues à droite dans $\mathbb{R}^+ \rightarrow \mathbb{N}^*$. On veut alors montrer que

$$\begin{aligned} & \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{\liminf_k f_k - 1}; t > 0 \right\} \\ & \leq \liminf_k \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{f_k - 1}; t > 0 \right\}. \end{aligned} \quad (3.21)$$

Si le membre de (3.21) est égal à $+\infty$, l'inégalité est triviale. On suppose alors le contraire. Il vient donc que pour une sous-suite $(k_j)_{j \geq 1}$ qui réalise la limite inférieure du membre de droite,

$$T = \sup\{t > 0; f_{k_j}(t) > 1, j \geq 1\} < \infty.$$

Par ailleurs, si l'on note t_k le réel qui minimise $\inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{f_k - 1}; t > 0 \right\}$, il existe une sous suite de $(k_j)_{j \geq 1}$ (que l'on note à nouveau $(k_j)_{j \geq 1}$) telle que $t_{k_j} \rightarrow t_\infty$ pour un certain $t_\infty \in \mathbb{R}$, puisque $(t_{k_j})_{j \geq 1} \subset [0; T]$ pour tout $j \geq 1$. Il vient alors, pour tout $\varepsilon > 0$,

$$\begin{aligned} & \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{\liminf_k f_k - 1}; t > 0 \right\} \\ & \leq \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{\liminf_j f_{k_j} - 1}; t > 0 \right\} \\ & \leq \inf \left\{ \liminf_j \left(t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{f_{k_j} - 1} \right); t > 0 \right\} \end{aligned} \quad (3.22)$$

$$\leq \liminf_j t_{k_j} + \varepsilon + \frac{1}{\sqrt{n}} \int_{t_{k_j}}^\infty \sqrt{f_{k_j} - 1} + \int_{t_\infty + \varepsilon}^{t_{k_j} \vee (t_\infty + \varepsilon)} \sqrt{f_{k_j} - 1} \quad (3.23)$$

$$= \liminf_k \inf \left\{ t + \frac{1}{\sqrt{n}} \int_t^\infty \sqrt{f_k - 1}; t > 0 \right\} + \varepsilon, \quad (3.24)$$

où (3.22) est une application du lemme de Fatou, (3.23) vient du fait que $t_{k_j} \rightarrow t_\infty$ et donc que $t_{k_j} \vee (t_\infty + \varepsilon) = t_\infty + \varepsilon$ pour j suffisamment grand, et enfin, (3.24) exprime juste le fait que t_k est un minimiseur et que $(k_j)_{j \geq 1}$ est une sous-suite qui réalise la limite inférieure. \square

3.8.5 Cas dépendant

Démonstration du théorème 3.14. La démonstration du théorème consiste à appliquer le lemme 3.19. Il nous faut alors majorer $E|\mu_n(A) - \mu(A)|$ pour tout borélien A . Pour cela, nous utilisons l'hypothèse (3.4) faite sur la chaîne de Markov. Notons $f_A = \mathbf{1}_A - \mu(A)$. Alors,

$$\begin{aligned} E_\nu |\mu_n(A) - \mu(A)| &= \frac{1}{n} E_\nu \left| \sum_{i=1}^n f_A(X_i) \right| \\ &\leq \frac{1}{n} \sqrt{\sum_{i=1}^n \sum_{j=1}^n E_\nu f_A(X_i) f_A(X_j)} \end{aligned}$$

Pour $j \geq i$ et en utilisant l'inégalité de Hölder deux fois avec $1/r + 1/2 + 1/s = 1$ et le fait que $f \mapsto Pf$ est une contraction de $L^{2s/(s+2)}(\mu)$, il vient

$$\begin{aligned} E_\nu f_A(X_i) f_A(X_j) &= E_\mu \frac{d\nu}{d\mu}(X_0) E(f_A(X_i) f_A(X_j) | X_0) \\ &\leq \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)} \left(E_\mu E(f_A(X_i) f_A(X_j) | X_0)^{2s/(s+2)} \right)^{1/2+1/s} \\ &\leq \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)} \left(E_\mu E(f_A(X_0) f_A(X_{j-1}) | X_0)^{2s/(s+2)} \right)^{1/2+1/s} \\ &\leq \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)} \|P^{j-i} f_A\|_{L^2(\mu)} \|f_A\|_{L^s(\mu)}. \end{aligned}$$

L'hypothèse (3.4) donne alors, puisque f_A est centrée sous E_μ ,

$$E_\nu f_A(X_i) f_A(X_j) \leq C \lambda^{j-i} \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)} \|f_A\|_{L^2(\mu)} \|f_A\|_{L^s(\mu)}.$$

Enfin, puisque $s \geq 2$ et que pour $t \geq 2$, $\|f_A\|_{L^t(\mu)}^t \leq \mu(A)(1 - \mu(A))$, il vient

$$E_\nu f_A(X_i) f_A(X_j) \leq C \lambda^{j-i} \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)} [\mu(A)(1 - \mu(A))]^{1-1/r},$$

et donc,

$$E_\nu |\mu_n(A) - \mu(A)| \leq \frac{\sqrt{2C}}{\sqrt{n}\sqrt{1-\lambda}} \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)}^{1/2} [\mu(A)(1 - \mu(A))]^{1/2-1/2r}.$$

Le lemme 3.19 donne donc, en posant $C_r = \frac{\sqrt{2C}}{\sqrt{1-\lambda}} \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)}^{1/2}$,

$$\begin{aligned}
E_\nu W_p^p(\mu, \mu_n) &\leq 2^p D^p \left[2^{1-kp} + \sum_{l=1}^k \sum_{j=1}^{n(l)} E_\nu |\mu(B_j^l) - \mu_n(B_j^l)| (2^{-(l-1)p} - 2^{-lp}) \right] \\
&\leq 2^p D^p \left[2^{1-kp} + \frac{C_r}{\sqrt{n}} \sum_{l=1}^k (2^{-(l-1)p} - 2^{-lp}) \sum_{j=1}^{n(l)} [\mu(B_j^l)(1 - \mu(B_j^l))]^{1/2-1/2r} \right] \\
&\leq 2^p D^p \left[2^{1-kp} + \frac{C_r}{\sqrt{n}} \sum_{l=1}^k (2^{-(l-1)p} - 2^{-lp}) (n(l) - 1)^{1/2+1/r} \right] \\
&\leq 2^p \left[D^p 2^{1-kp} + \frac{C_r}{\sqrt{n}} \int_{D^{p2-kp}}^\infty (\mathcal{N}(E, u^{1/p}) - 1)^{1/2+1/r} du \right],
\end{aligned}$$

où la troisième inégalité est due à l'inégalité de Hölder. \square

Démonstration du théorème 3.15. La démonstration de ce théorème reprend celle du cas indépendant (théorème 3.8). Pour une partition $(K_i)_{i \geq 1}$ de E ,

$$2^{1-p} W_p^p(\mu_n, \mu) \leq \sum_{i \geq 1} \mu_n(K_i) \wedge \mu(K_i) W_p^p(\mu_n|K_i, \mu|K_i) \quad (3.25)$$

$$+ \sum_{i \geq 1} (\mu_n(K_i) - \mu(K_i))^+ W_p^p(\mu_n|K_i, \mu) \quad (3.26)$$

$$+ \sum_{i \geq 1} (\mu(K_i) - \mu_n(K_i))^+ W_p^p(\mu|K_i, \mu). \quad (3.27)$$

Cherchons dans un premier temps à majorer (3.25). Pour cela, nous allons utiliser le lemme 3.19, et nous avons donc besoin de majorer $E_\nu |\mu_n(A) - \mu(A)|$. En reprenant les arguments de la démonstration du théorème 3.14, avec $f_A = \mathbf{1}_A - \mathbf{1}_{K_i} \frac{\mu(A)}{\mu(K_i)}$ pour $A \subset K_i$ et en remarquant qu'alors, pour $t \geq 2$,

$$\|f_A\|_{L^t(\mu)}^t \leq \mu(K_i) \left[\frac{\mu(A)}{\mu(K_i)} \left(1 - \frac{\mu(A)}{\mu(K_i)} \right) \right],$$

il vient, pour $A \subset K_i$

$$E_\nu |\mu_n(A) - \mu(A) \frac{\mu_n(K_i)}{\mu(K_i)}| \leq \frac{\sqrt{2C}}{\sqrt{n}\sqrt{1-\lambda}} \left\| \frac{d\nu}{d\mu} \right\|_{L^r(\mu)}^{1/2} \left[\mu(K_i) \frac{\mu(A)}{\mu(K_i)} \left(1 - \frac{\mu(A)}{\mu(K_i)} \right) \right]^{1/2-1/2r}.$$

En appliquant le lemme 3.19 à $\mu|K_i$ et $\mu_n|K_i$, nous obtenons alors, pour une partition $(B_{j,i}^l)_{1 \leq j \leq n(l)}$ de K_i , pour $i \geq 1$,

$$\begin{aligned}
E_\nu(3.25) &\leq \sum_{i \geq 1} 2^p D^p \left[\mu(K_i) 2^{1-k_i p} + \sum_{l=1}^{k_i} (2^{-(l-1)p} - 2^{-lp}) \sum_{j=1}^{n_i(l)} E_\nu \left| \frac{\mu_n(K_i)}{\mu(K_i)} \mu(B_{j,i}^l) - \mu_n(B_{j,i}^l) \right| \right] \\
&\leq \sum_{i \geq 1} 2^p D^p \left[\mu(K_i) 2^{1-k_i p} + \sum_{l=1}^{k_i} (2^{-(l-1)p} - 2^{-lp}) \frac{C_r}{\sqrt{n}} \mu(K_i)^{1/2-1/2r} (n_i(l) - 1)^{1/2+1/r} \right] \\
&\leq \sum_{i \geq 1} 2^p \left[\mu(K_i) D^p 2^{1-k_i p} + \frac{C_r}{\sqrt{n}} \mu(K_i)^{1/2-1/2r} \int_{D^p 2^{-k_i p}}^{\infty} (\mathcal{N}(K_i, u^{1/p}) - 1)^{1/2+1/r} du \right] \\
&\leq \sum_{i \geq 1} 2^p \left[\mu(K_i) 2^{p+1} t_i + \frac{C_r}{\sqrt{n}} \mu(K_i)^{1/2-1/2r} \int_{t_i}^{\infty} (\mathcal{N}(K_i, u^{1/p}) - 1)^{1/2+1/r} du \right],
\end{aligned}$$

pour $t_i > 0$ pour tout $i \geq 1$.

Pour majorer (3.26) et (3.27), on remarque que - pour x_0 tel que pour tout $i \geq 1$, $\sup_{x \in K_i} d(x, x_0) \leq \text{diam}(\cup_{j \leq i} K_j)$,

$$\begin{aligned}
W_p^p(\mu_n|K_i, \mu) \vee W_p^p(\mu|K_i, \mu) &\leq Ed(x_0, X)^p + \text{diam}(\cup_{j \leq i} K_j)^p \\
&\leq Ed(x_0, X)^p + t_i + \int_{t_i}^{\infty} (\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1)^{1/2+1/r} du.
\end{aligned}$$

de sorte que, si $\mu(K_i) \leq 2^{-i}$, et pour $t_i > 0$ pour tout $i \geq 1$,

$$\begin{aligned}
E_\nu(3.26 + 3.27) &\leq \sum_{i \geq 1} E_\nu |\mu_n(K_i) - \mu(K_i)| \left(Ed(x_0, X)^p + t_i + \int_{t_i}^{\infty} (\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1)^{1/2+1/r} du \right) \\
&\leq \frac{C_r}{\sqrt{n}} \sum_{i \geq 1} \mu(K_i)^{1/2-1/2r} \left(Ed(x_0, X)^p + t_i + \int_{t_i}^{\infty} (\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1)^{1/2+1/r} du \right) \\
&\leq \frac{C_r}{\sqrt{n}(2^{\frac{2r}{r-1}} - 1)} Ed(x_0, X)^p \\
&\quad + C_r \sum_{i \geq 1} 2^{-i(1/2-1/2r)} \left(t_i + \frac{1}{\sqrt{n}} \int_{t_i}^{\infty} (\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1)^{1/2+1/r} du \right).
\end{aligned}$$

Ainsi,

$$\begin{aligned}
2^{1-p}EW(\mu, \mu_n) &\leq \frac{C_r}{\sqrt{n}(2^{\frac{2r}{r-1}} - 1)} Ed(x_0, X)^p \\
&\quad + (2^{2p+1} + 1)C_r \sum_{i \geq 1} 2^{-i(1/2-1/2r)} \\
&\quad \times \left(t_i + \frac{1}{\sqrt{n}} \int_{t_i}^{\infty} \left(\mathcal{N}(\cup_{j \leq i} K_j, u^{1/p}) - 1 \right)^{1/2+1/r} du \right) \\
&\leq C_r \left(\frac{Ed(x_0, X)^p}{\sqrt{n}(2^{\frac{2r}{r-1}} - 1)} + (2^{2p+2} + 2) \int_0^{1/2} \frac{\tau_n^{\alpha, r}}{\alpha^{1/2+1/r}} d\alpha \right).
\end{aligned}$$

□

3.9 Conclusion

Les résultats obtenus dans ce chapitre fournissent une majoration précise de la distance de Wasserstein entre une mesure et sa version empirique. Les vitesses de convergence de cette distance, dans des cas tels que les espaces de dimension finie ou dans le cas des mesures gaussiennes sur des espaces de Banach, coïncident souvent avec la quantification optimale, ce qui montre qu'il n'est, dans ces cas, pas nécessaire de chercher les quantifieurs optimaux pour obtenir une vitesse optimale ; la mesure empirique suffit.

Nous avons pu voir que ces majorations permettent de formuler des tests non paramétriques de niveau contrôlé de manière non asymptotique, dans un cadre relativement général.

Lorsque la dimension de l'espace est finie mais n'est pas entière, il n'est plus clair que la vitesse de convergence est optimale, on peut alors se demander si la vitesse de convergence optimale sera une puissance de la dimension au sens de Hausdorff, du recouvrement par des boules ouvertes, ou si cela définira une nouvelle notion de dimension.

Chapitre 4

Barycentre dans l'espace de Wasserstein

Ce dernier chapitre rassemble des résultats originaux sur la régularité des barycentres dans l'espace de Wasserstein. Un travail en collaboration avec Emmanuel Boissard et Jean-Michel Loubes a abouti à un article à l'origine de ce chapitre. Cependant, certains résultats issus des travaux plus personnels à Emmanuel et Jean-Michel ne sont pas abordés, et un cadre plus général des résultats de l'article, étendant le cadre de \mathbb{R}^d à celui des espaces géodésiques, est présenté ici.

Une première partie de chapitre vient introduire le problème. Les résultats originaux viennent ensuite former la deuxième partie. Puis, les applications viennent illustrer les résultats dans une troisième partie, et enfin les démonstrations concluent ce chapitre.

4.1 Introduction

Ce chapitre étudie les barycentres dans les espaces de Wasserstein tels que présentés dans le chapitre 1. L'origine de l'intérêt pour cet étude est l'article [Agueh et Carlier, 2010], dont le résultat principal est un problème dual à celui que minimise le barycentre, analogue à la dualité de Kantorovitch. Il y est notamment montré, qu'un barycentre minimise aussi un problème dit multi-marginal. Ce problème consiste à chercher une mesure à $J > 0$ marginales fixées qui réalise le minimum

$$\inf \left\{ \int \sum_{1 \leq i \leq J} \lambda_i |x_i - T(x)|^2 d\gamma(x_1, \dots, x_J); \gamma \in \Gamma(\mu_1, \dots, \mu_J) \right\},$$

où $T : \mathbb{R}^d \mapsto \mathbb{R}$ est l'application barycentre associée aux poids $(\lambda_i)_{1 \leq i \leq J}$. [Agueh et Carlier, 2010] montre alors que le barycentre sur \mathbb{R}^d est la mesure image de l'application barycentre T . L'introduction du problème multi-marginal remonte à [Gangbo et Świąch, 1998], qui démontre l'existence d'un transport entre les marginales pour la mesure minimisant le problème multi-marginal.

Ce chapitre montre l'existence du barycentre, ainsi que la continuité de celui-ci par rapport aux mesures auxquelles il est associé.

4.2 Propriétés du barycentre

Le cadre générique dans lequel nous nous plaçons, est celui d'un espace géodésique (E, d) localement compact et les notations sont celles du chapitre 1. Nous rappelons la définition du barycentre.

Définition 4.1 (p -barycentre). *Soient (E, d) un espace métrique et $p > 0$. Soient $(\lambda_i)_{1 \leq i \leq J}$ des poids et $(x_i)_{1 \leq i \leq J} \in E^J$. Le p -barycentre de $(x_i)_{1 \leq i \leq J}$ pour les poids $(\lambda_i)_{1 \leq i \leq J}$ est défini, lorsqu'il existe, comme l'ensemble des minimiseurs de*

$$\sum_{1 \leq i \leq J} \lambda_i d(x, x_i)^p.$$

Tout élément de cet ensemble est aussi appelé un barycentre.

4.2.1 Existence

Pour tout $p > 1$, le théorème 1.27 nous assure, qu'étant donnés des poids $(\lambda_i)_{1 \leq i \leq J}$, il existe toujours un p -barycentre associé à un J -uplet d'éléments de E . Cependant, l'espace de Wasserstein associé à cet espace n'est pas localement compact (voir exemple 1.35), et il n'est donc pas possible de déduire du théorème 1.27 qu'un p -barycentre dans l'espace de Wasserstein existe.

Nous le déduisons de l'équivalence avec le problème multi-marginal.

Théorème 4.2 (Barycentre et problème multi-marginal). *Soit (E, d) un espace géodésique complet séparable. Soient $p \geq 1$ et $J \in \mathbb{N}^*$. Étant donné $(\mu_i)_{1 \leq i \leq J} \in \mathcal{P}_p(E)^J$ et des poids $(\lambda_i)_{1 \leq i \leq J}$, il existe une mesure $\gamma \in \Gamma(\mu_1, \dots, \mu_J)$ qui minimise*

$$\hat{\gamma} \mapsto \int \inf_{x \in E} \sum_{1 \leq i \leq J} \lambda_i d(x_i, x)^p d\hat{\gamma}(x_1, \dots, x_J).$$

De plus, s'il existe une application barycentre $T : E^J \rightarrow E$ mesurable, alors la mesure $\nu = T_{\#}\gamma$ est un barycentre de $(\mu_i)_{1 \leq i \leq J}$ et si cette application est unique, tout barycentre ν est de la forme $\nu = T_{\#}\gamma$.

Remarque 4.3. *La mesurabilité de $(x_1, \dots, x_J) \mapsto \inf_{x \in E} \sum_{1 \leq i \leq J} \lambda_i d(x_i, x)^p$ provient de la séparabilité : l'infimum est le même que s'il portait sur un sous-ensemble dénombrable dense.*

En particulier, le théorème 1.27 affirme qu'il existe une application $T : E^J \rightarrow E$ qui associe un barycentre à un J -uplet de E , que nous appelons application **barycentre** et nous obtenons alors le résultat suivant.

Théorème 4.4 (Existence du barycentre). *Soient (E, d) un espace géodésique complet séparable et $p > 1$. Étant donnés des poids $(\lambda_i)_{1 \leq i \leq J}$, s'il existe une application barycentre sur E mesurable, alors pour tout $(\mu_j)_{1 \leq j \leq J} \in \mathcal{P}_p(E)$, il existe un barycentre associé.*

Ce résultat se trouve dans [Agueh et Carlier, 2010] pour le cas des mesures de \mathbb{R}^d qui donnent une mesure nulle aux ensembles de dimension de Hausdorff inférieure à $d - 1$. Cependant, leur démonstration, qui utilise le problème multi-marginal, se base seulement sur l'existence du minimiseur du problème multi-marginal. Cette existence est alors présentée comme corollaire d'un résultat de [Gangbo et Świąch, 1998]), mais c'est aussi une conséquence de la tension des mesures à marginales fixées (lemme 4.5), ce qui permet de s'abstenir de la condition sur la mesure nulle sur les ensembles de dimension de Hausdorff inférieure à $d - 1$ et de pouvoir se placer dans le cadre plus général des espaces géodésiques.

Nous utilisons la notation $\Gamma(\mu_1, \dots, \mu_J)$ pour l'ensemble des mesures ayant pour i -ème marginale μ_i , pour $1 \leq i \leq J$ et de même, $\Gamma(C_1, \dots, C_J)$ désigne l'ensemble des mesures ayant leur i -ème marginale dans $C_i \subset \mathcal{P}_p(E)$ pour tout $1 \leq i \leq J$.

La démonstration du théorème 4.2 se base sur l'existence de la solution du problème multi-marginal, qui est une conséquence directe du lemme suivant.

Lemme 4.5 (Compacité des mesures à marginales dans un compact). *Soit (E, d) un espace polonais (i.e. métrique, séparable et complet). Soient C_1, \dots, C_J des ensembles compacts de $\mathcal{P}_p(E)$ muni de la distance de Wasserstein. Alors $\Gamma(C_1, \dots, C_J)$ est compact pour la distance de Wasserstein sur $\mathcal{P}_p(E^J)$.*

Remarque 4.6. *La topologie supposée sur $\mathcal{P}_p(E)$ est celle induite par la distance de Wasserstein.*

Remarque 4.7. *Dans les résultats ci-dessus, la compacité locale de l'espace E n'est demandée que pour l'existence du barycentre. Il est tout à fait possible de remplacer cette condition par l'existence du barycentre.*

L'unicité du barycentre n'est en revanche pas assurée, y compris lorsque les barycentres de l'espace sous-jacent sont uniques.

Exemple 4.8 (Non unicité du barycentre). *Le barycentre, pour les poids $(t, 1 - t)$, de deux mesures μ et ν sur \mathbb{R}^d , de la forme*

$$\frac{1}{n} \sum \delta_{x_i},$$

est atteint par la mesure image du plan de transport entre μ et ν par l'application

$$T : (x, y) \mapsto tx + (1 - t)y.$$

Une démonstration de ce résultat est faite dans [Villani, 2009].

Si l'on se place sur \mathbb{R}^2 , et que l'on choisit

$$\mu = \frac{1}{2}(\delta_{(-1,-1)} + \delta_{(1,1)}) \text{ et } \nu = \frac{1}{2}(\delta_{(1,-1)} + \delta_{(-1,1)}),$$

alors, le transport optimal entre μ et ν n'est pas unique. Les deux plans de transport suivants sont optimaux,

$$\pi_1 = \frac{1}{2}(\delta_{((-1,-1),(1,-1))} + \delta_{((1,1),(-1,1))}) \text{ et } \pi_2 = \frac{1}{2}(\delta_{((-1,-1),(-1,1))} + \delta_{((1,1),(1,-1))}).$$

Ainsi, les deux mesures suivantes sont un barycentre de μ et ν ,

$$T_{\#}\pi_1 \text{ et } T_{\#}\pi_2.$$

Ceci est une conséquence du fait que la distance entre chaque point d'une mesure aux points de l'autre mesure est la même. Le barycentre dans l'espace de Wasserstein ne sera donc pas toujours unique dès que l'on pourra construire de tels quatre points, ce qui restreint bien les cas d'unicité de tous les barycentres.

En revanche, pour le cas du barycentre entre deux mesures, il est clair que si ces deux mesures sont reliées par un unique chemin de longueur minimale, alors leur barycentre est unique. La proposition 3.5 de [Agueh et Carlier, 2010] fournit un critère d'unicité du barycentre pour des mesures sur \mathbb{R}^d : étant donnés des poids et J mesures, il existe un unique barycentre associé dès lors qu'une des mesures s'annule sur tous les ensembles de mesures de Hausdorff inférieure à $d - 1$.

Il est intéressant de remarquer que l'on peut retrouver les barycentres de l'espace sous-jacent E dans l'espace de Wasserstein. Si chacune des mesures dont on prend le barycentre est une masse de Dirac en un point, la mesure barycentre sera la masse de Dirac en le barycentre de ces points. Il est facile de s'en convaincre à l'aide du théorème 4.2 : il n'existe qu'une seule mesure $\gamma \in \Gamma(\delta_{x_1}, \dots, \delta_{x_J})$ et le minimum de

$$\hat{\gamma} \mapsto \int \inf_{x \in E} \sum_{1 \leq i \leq J} \lambda_i d(x_i, x)^p d\hat{\gamma}(x_1, \dots, x_J).$$

est donc atteint pour cette mesure. L'infimum sur E est alors aussi atteint en $T(x_1, \dots, x_J)$ le barycentre de (x_1, \dots, x_J) et $T_{\#}\gamma = \delta_{T(x_1, \dots, x_J)}$.

4.2.2 Continuité

Nous nous intéressons ici à la continuité du barycentre. Celui-ci n'étant pas défini de manière unique, ce que nous entendons par continuité est décrit dans le théorème suivant.

Théoreme 4.9 (Continuité du barycentre). *Soient (E, d) un espace géodésique dont l'application barycentre associée aux poids $(\lambda_i)_{1 \leq i \leq J}$ est continue et $(\mu_i^n)_{1 \leq i \leq J, n \geq 1}$ une suite de J -uplets de $\mathcal{P}_p(E)$ qui converge en distance de Wasserstein vers un J -uplet $(\mu_i)_{1 \leq i \leq J}$. Pour $n \geq 1$, on note μ_B^n un barycentre de $(\mu_i^n)_{1 \leq i \leq J}$. Alors, la suite $(\mu_B^n)_{n \geq 1}$ est compacte et tout point limite est un barycentre de $(\mu_i)_{1 \leq i \leq J}$.*

Remarque 4.10. *Contrairement à ce que pourrait laisser penser l'intuition, un espace E pour lequel le barycentre est lipschitzien, n'a pas nécessairement un barycentre dans l'espace de Wasserstein associé lipschitzien. C'est en effet le cas de l'espace de Wasserstein du \mathbb{R}^d pour $d \geq 2$, où le barycentre sur l'espace de Wasserstein n'est pas unique, et ne peut donc pas être lipschitzien.*

Une première conséquence de ce résultat est la compacité du barycentre.

Corollaire 4.11 (Compacité de l'ensemble des barycentres). *Soient (E, d) un espace géodésique tel que l'application barycentre est continue pour les poids $(\lambda_j)_{1 \leq j \leq J}$ et $(\mu_i)_{1 \leq i \leq J}$ un J -uplet de $\mathcal{P}_p(E)$. Alors l'ensemble des barycentres de $(\mu_i)_{1 \leq i \leq J}$ associé à ces poids est compact pour la distance de Wasserstein.*

En effet, en prenant une suite incluse dans l'ensemble des barycentres d'un J -uplet donné, le théorème 4.9 assure qu'elle est compacte et que tout point limite est un barycentre.

En reprenant la définition de contraste de Hausdorff du chapitre 2, il est possible de reformuler le théorème. Nous rappelons cette définition ici.

Définition 4.12 (Contraste de Hausdorff). *Soient A et B deux sous-ensembles fermés d'un espace polonais (E, d) . Le **contraste de Hausdorff** entre A et B est défini par*

$$\text{Haus}(A|B) := \inf\{\varepsilon > 0 | A \subset B^\varepsilon\}.$$

Le théorème 4.9 prend alors la forme suivante.

Corollaire 4.13 (Continuité du barycentre par le contraste de Hausdorff). *Soient (E, d) un espace géodésique et $(\mu_i^n)_{1 \leq i \leq J, n \geq 1}$ une suite de J -uplets de $\mathcal{P}_p(E)$ qui converge en distance de Wasserstein vers un J -uplet $(\mu_i)_{1 \leq i \leq J}$. Pour $n \geq 1$, on note B^n l'ensemble des barycentres de $(\mu_i^n)_{1 \leq i \leq J}$ et B ceux de $(\mu_i)_{1 \leq i \leq J}$. Alors,*

$$\text{Haus}(B^n|B) \rightarrow 0.$$

4.3 Perspectives d'applications

Les résultats présentés jusqu'ici ne permettent pas de borner la distance de Wasserstein entre deux barycentres en fonction des familles de mesures dont on prend le barycentre, et il n'est pas possible d'obtenir une telle borne dans le cas général, puisque le barycentre n'est pas toujours unique, comme nous l'avons vu dans la remarque 4.8. Cependant, en utilisant le contraste de Hausdorff, il est possible d'imaginer pouvoir borner la distance de Wasserstein entre un barycentre d'une famille et le plus proche des barycentres d'une autre famille, en fonction de la distance entre les deux familles. Une première perspective d'application dans ce cas serait alors l'agrégation d'estimateurs de mesures. Une seconde application présentée ici est celle des mélanges d'images.

4.3.1 Agrégation d'estimateurs d'une mesure

Supposons que l'on se trouve sur un espace géodésique dont le barycentre est continu. L'objectif est alors d'estimer une mesure μ . Supposons que nous ayons à disposition plusieurs approches qui amènent à différentes estimations μ_1, \dots, μ_J . Chacune des ces estimations est alors supposée convergente en distance de Wasserstein vers la mesure μ de départ que l'on souhaite estimer. Il est alors naturel de choisir une mesure $\hat{\mu}$ qui minimise

$$\nu \mapsto \sum_{1 \leq j \leq J} W_2^2(\nu, \mu_j),$$

pour agréger les différentes estimations.

Bien qu'il semble assez intuitif que $\hat{\mu}$ soit un bon estimateur de μ , il n'est pas encore clair qu'il soit « meilleur » que les μ_j , ni même que le « pire » des μ_j , mais le théorème 4.9 affirme que $\hat{\nu}$ converge avec les μ_j vers μ .

Une borne du contraste de Hausdorff $\text{Haus}(B^n|B)$ où B^n est l'ensemble des barycentres de $(\mu_j^n)_{1 \leq j \leq J}$ (pour une famille de poids fixée), et $B = \{\mu\}$, permettrait d'obtenir une borne sur l'erreur commise en agrégeant les estimateurs. Un bref raisonnement montre que l'erreur commise par l'agrégat est inférieur au double de la moyenne des erreurs lorsque $p = 1$. En effet, si l'on note μ_B^n un 1-barycentre des estimateurs, pour des poids $(\lambda_j)_{1 \leq j \leq J}$,

$$W(\mu_B^n, \mu) \leq \sum_{1 \leq j \leq J} \lambda_j (W(\mu_B^n, \mu_j^n) + W(\mu_j^n, \mu)) \leq 2 \sum_{1 \leq j \leq J} \lambda_j W(\mu, \mu_j^n).$$

Le chapitre 2 et notamment le résultat de convergence des ensembles de masse localisée (théorème 2.28) permet d'améliorer encore ce résultat, de manière asymptotique. Supposons que l'on sache qu'une proportion $1 - \alpha > 1/2$ des estimateurs

converge vers la mesure μ que l'on cherche à estimer, et que les autres mesures convergent. La mesure de probabilité P_n définie sur l'espace de Wasserstein par

$$P_n = \sum_{1 \leq j \leq J} \frac{1}{j} \delta_{\mu_j^n}$$

converge alors vers une mesure limite $P = \sum_{1 \leq j \leq J} \frac{1}{j} \delta_{\mu^n}$, où une proportion $1 - \alpha$ des μ^n est égal à μ . Alors, choisir une proportion $1 - \alpha$ des estimateurs qui minimise la taille (disons)

$$B \mapsto \tau(B) = \int_0^1 \arctan \mathcal{M}(B, t) dt,$$

puis en prendre le barycentre, fournit un estimateur consistant de μ , qui est plus proche de μ que la moyenne de la proportion $1 - \alpha$ des estimateurs les plus proches de μ , à partir d'un certain rang. En effet, le théorème 2.28 affirme que l'ensemble B_n^α de plus petite taille contenant une masse supérieure à $1 - \alpha$ converge, pour la distance de Hausdorff, vers l'ensemble de plus petite taille de masse au moins $1 - \alpha$ pour P , qui est trivialement égal à $\{\mu\}$ par hypothèse. Ainsi, à partir d'un certain rang, (lorsque la distance de Hausdorff entre $\{\mu\}$ et B_n^α est inférieure à la moitié de la plus grande distance entre les autres estimateurs et leur limite), la proportion d'estimateurs choisie est la proportion des estimateurs convergents.

4.3.2 Mélange d'images

Une image en niveau de gris peut être modélisée comme une mesure de probabilité où l'intensité de gris en chaque pixel correspond à la masse de la mesure en ce pixel. Trouver le barycentre, au sens de Wasserstein, des mesures associées à un ensemble d'images constitue une méthode de mélange de ces images.

Cette technique de mélange d'image est développée dans [Rabin et al., 2012]. Afin d'obtenir des algorithmes moins coûteux en calcul, ils introduisent une notion de « sliced Wasserstein » qui consiste à remplacer la distance de Wasserstein par la moyenne des distances de Wasserstein des projections sur chaque direction de l'espace, les calculs de la distance de Wasserstein sur \mathbb{R} étant bien plus simples que ceux sur \mathbb{R}^d . Dans le but de mélanger des images, ils présentent un algorithme pour calculer le barycentre d'une famille d'images, pour cette notion de « sliced Wasserstein ». Une perspective d'application serait alors de vérifier que la continuité du barycentre telle que formulée dans le théorème 4.9 reste valable pour la notion de « sliced Wasserstein », ce qui permettrait d'affirmer qu'un mélange d'images pour cette notion dépend continûment des images que l'on cherche à mélanger.

Le théorème 4.9 permet déjà d'affirmer que le barycentre au sens de Wasserstein dépend continûment des images que l'on souhaite mélanger.

4.4 Démonstrations

Démonstration du lemme 4.5. L'espace de Wasserstein étant métrique, il suffit de montrer la compacité séquentielle de $\Gamma(C_1, \dots, C_J)$. Notons d'abord qu'il est clair que $\Gamma(C_1, \dots, C_J)$ est fermé.

Soit $(\gamma^n)_{n \geq 1} \subset \Gamma(C_1, \dots, C_J)$. Nous allons montrer qu'une sous-suite de $(\gamma^n)_{n \geq 1}$ converge en distance de Wasserstein vers une limite γ . Puisque les C_i sont compacts, alors ils sont tendus, c'est-à-dire, pour tout $\varepsilon > 0$, il existe des compacts de $B_i \subset E$ tels que, pour $1 \leq i \leq J$,

$$\forall \nu \in C_i, \nu(B_i^c) \leq \varepsilon.$$

Il vient alors,

$$\begin{aligned} \gamma^n((B_1 \times \dots \times B_J)^c) &= \gamma^n(\cup_{j=1}^J E \times \dots \times E \times B_j^c \times E \times \dots \times E) \\ &\leq \sum_{j=1}^J \gamma^n(E \times \dots \times E \times B_j^c \times E \times \dots \times E) \\ &= \sum_{j=1}^J \mu_j^n(B_j^c). \end{aligned}$$

Ainsi, la tension de $(\gamma^n)_{n \geq 1}$ est assurée et il existe donc $\gamma \in \Gamma(C_1, \dots, C_J)$ limite d'une sous-suite de $(\gamma^n)_{n \geq 1}$, que l'on renomme $(\gamma^n)_{n \geq 1}$.

En munissant $\mathcal{P}_p(E^J)$ de la distance de Wasserstein issue de la distance produit

$$d_p : (x, y) \mapsto \left(\sum_{1 \leq i \leq J} d(x_i, y_i)^p \right)^{1/p},$$

il vient, pour tout $z_0 \in E^J$,

$$\begin{aligned} \int d_p(z_0, x)^p d\gamma^n(x) &= \sum_{j=1}^J \int d(z_0, x_i)^p d\mu_i^n(x) \\ &\rightarrow \sum_{j=1}^J \int d(z_0, x_i)^p d\mu_i(x) = \int d_p(z_0, x)^p d\gamma(x). \end{aligned}$$

Puisque la convergence en distance de Wasserstein coïncide avec la convergence faible et la convergence des moments (voir théorème 1.30), nous avons montré que $(\gamma^n)_{n \geq 1}$ converge bien en distance de Wasserstein vers γ . \square

4.4.1 Existence

Démonstration du théorème 4.2. Cette démonstration est reprise de la proposition 4.2 de [Agueh et Carlier, 2010].

L'existence de la solution du problème multi-marginal est une conséquence directe du lemme 4.5.

Notons γ une solution du problème multi-marginal et $\nu = T_{\#}\gamma$. Alors,

$$W_p^p(\mu_i, \nu) \leq \int d^p(x_i, T(x)) d\gamma(x).$$

et donc,

$$\sum_{1 \leq i \leq J} \lambda_i W_p^p(\mu_i, \nu) \leq \int \sum_{1 \leq i \leq J} \lambda_i d^p(x_i, T(x)) d\gamma(x). \quad (4.1)$$

Par ailleurs, pour $\hat{\nu} \in \mathcal{P}_p(E)$, on note $\pi_i \in \Gamma(\mu_i, \hat{\nu})$ le plan de transport optimal entre $\hat{\nu}$ et μ_i . Par le théorème de désintégration, pour tout $1 \leq i \leq J$, il existe une mesure (conditionnelle) μ_i^y définie pour $\hat{\nu}$ -presque tout y , qui vérifie $\pi_i(x, y) = \mu_i^y(x) \otimes \hat{\nu}(y)$. Posons alors

$$\theta(x, y) = \mu_1^y(x_1) \otimes \dots \otimes \mu_J^y(x_J) \otimes \hat{\nu}(y),$$

et notons $\hat{\gamma}$ la loi des J premières marginales de θ . Il vient alors, par construction de θ ,

$$\begin{aligned} \sum_{1 \leq i \leq J} \lambda_i W_p^p(\mu_i, \hat{\nu}) &= \sum_{1 \leq i \leq J} \lambda_i \int d^p(x_i, y) d\theta(x, y) \\ &= \int \sum_{1 \leq i \leq J} \lambda_i d^p(x_i, y) d\theta(x, y) \\ &\geq \int \inf_{z \in E} \sum_{1 \leq i \leq J} \lambda_i d^p(x_i, z) d\theta(x, y) \\ &= \int \sum_{1 \leq i \leq J} \lambda_i d^p(x_i, T(x)) d\theta(x, y) \\ &= \int \sum_{1 \leq i \leq J} \lambda_i d^p(x_i, T(x)) d\hat{\gamma}(x) \\ &\geq \int \sum_{1 \leq i \leq J} \lambda_i d^p(x_i, T(x)) d\gamma(x) \\ &\geq \sum_{1 \leq i \leq J} \lambda_i W_p^p(\mu_i, \nu), \end{aligned} \quad (4.2)$$

où la dernière inégalité est issue de (4.1). Comme $\hat{\nu}$ est arbitraire, nous avons montré que $T_{\#}\gamma$ est un barycentre.

Par ailleurs, en prenant $\hat{\nu}$ un barycentre, (4.2) devient une égalité, de sorte que, pour θ -presque tout $(x, y) \in E^J \times E$,

$$\sum_{1 \leq i \leq J} \lambda_i d^p(x_i, y) = \inf_{z \in E} \sum_{1 \leq i \leq J} \lambda_i d^p(x_i, z) = \sum_{1 \leq i \leq J} \lambda_i d^p(x_i, T(x)),$$

et, donc, si l'application barycentre T est unique, $T(x) = y$, θ -presque sûrement et donc $T_{\#}\hat{\gamma} = \hat{\nu}$. Aussi, l'optimalité de $\hat{\nu}$, et (4.1) montrent que $\hat{\gamma}$ est solution du problème multi-marginal. \square

4.4.2 Continuité

Nous allons utiliser le lemme suivant pour démontrer la continuité du barycentre dans l'espace de Wasserstein.

Lemme 4.14. *Soient (E, d) un espace polonais, et $(\mu_n)_{n \geq 1}$ une suite de mesures qui converge faiblement vers une mesure μ et telle que pour une mesure ν*

$$W_p^p(\mu_n, \nu) \rightarrow W_p^p(\mu, \nu).$$

Alors,

$$W_p^p(\mu_n, \mu) \rightarrow 0.$$

Démonstration du lemme 4.14. Soit $\pi_n \in \Gamma(\mu_n, \nu)$ de sorte que

$$W_p^p(\mu_n, \nu) = \int d^p(x, y) d\pi_n(x, y).$$

Alors, on peut montrer que

$$\limsup_{R \rightarrow \infty} \sup_{n \geq 1} \int_{d(x, y) \geq R} d^p(x, y) d\pi_n(x, y) = 0. \quad (4.3)$$

Par le théorème de désintégration, pour tout $n \geq 1$, il existe une mesure (conditionnelle) $\nu^{x, n}$ définie pour μ_n -presque tout x , qui vérifie $\pi_n(x, y) = \nu^{x, n}(y) \otimes \mu_n(x)$. Soit $\hat{\pi}_n(x, z) \in \Gamma(\mu_n, \mu)$ telle que

$$W_p^p(\mu_n, \mu) = \int d^p(x, z) d\hat{\pi}_n(x, z).$$

On pose alors $\pi_n^\nu(x, y, z) = \nu^{x, n}(y) \otimes \hat{\pi}_n(x, z)$. En utilisant , pour $x_0 \in E$,

$$\mathbf{1}_{d(x, z) \geq R} \leq \mathbf{1}_{d(x, y) \geq R/3} + \mathbf{1}_{d(y, x_0) \geq R/3} + \mathbf{1}_{d(x_0, z) \geq R/3},$$

il vient

$$\begin{aligned} (d^p(x, z) - R^p)^+ &\leq 4^p d^p(x, y) \mathbf{1}_{d(x, y) \geq R/3} \\ &\quad + 4^p d^p(y, x_0) \mathbf{1}_{d(y, x_0) \geq R/3} + 4^p d^p(x_0, z) \mathbf{1}_{d(x_0, z) \geq R/3} \end{aligned}$$

où $(f)^+$ désigne la partie positive de f (i.e. $(f)^+ = f \vee 0$). Ainsi, pour $x_0 \in E$,

$$\begin{aligned}
W_p^p(\mu_n, \mu) &= \int d^p(x, z) d\hat{\pi}_n(x, z) \\
&= \int d^p(x, z) d\pi_n^\nu(x, y, z) \\
&= \int d^p(x, z) \wedge R^p d\pi_n^\nu(x, y, z) + \int (d^p(x, z) - R^p)^+ d\pi_n^\nu(x, y, z) \\
&\leq \int d^p(x, z) \wedge R^p d\pi_n^\nu(x, y, z) + 4^p \int_{d(x, y) \geq R/3} d^p(x, y) d\pi_n^\nu(x, y, z) \\
&\quad + 4^p \int_{d(y, x_0) \geq R/3} d^p(y, x_0) d\pi_n^\nu(x, y, z) \\
&\quad + 4^p \int_{d(x_0, z) \geq R/3} d^p(x_0, z) d\pi_n^\nu(x, y, z) \\
&= \int d^p(x, z) \wedge R^p d\hat{\pi}_n(x, y, z) \tag{4.4}
\end{aligned}$$

$$+ 4^p \int_{d(x, y) \geq R/3} d^p(x, y) d\pi_n(x, y) \tag{4.5}$$

$$+ 4^p \int_{d(y, x_0) \geq R/3} d^p(y, x_0) d\nu(y) \tag{4.6}$$

$$+ 4^p \int_{d(x_0, z) \geq R/3} d^p(x_0, z) d\mu(z). \tag{4.7}$$

Or, (4.4) tend vers 0 quand n tend vers l'infini, car μ_n converge faiblement vers μ ; (4.5) tend vers 0 uniformément en $n \geq 1$ quand R tend vers l'infini selon (4.3); et (4.6) et (4.7) tendent vers 0 quand R tend vers l'infini par tension d'une mesure sur un espace polonais; ce qui achève la démonstration du lemme. \square

Démonstration du théorème 4.9. Le théorème 4.2 affirme que tout barycentre est de la forme $T_\# \gamma$, où γ est solution du problème multi-marginal. Si l'on note γ_n la solution du problème multi-marginal telle que, pour tout $n \geq 1$,

$$\mu_B^n = T_\# \gamma_n,$$

alors le lemme 4.5 nous assure que la suite $(\gamma_n)_{n \geq 1}$ est compacte pour la topologie induite par la distance de Wasserstein. En notant γ une limite de la suite $(\gamma_n)_{n \geq 1}$, et μ_B un p -barycentre de $(\mu_i)_{1 \leq i \leq J}$, il vient par optimalité de μ_B^n ,

$$\begin{aligned}
\int \left[\sum_{1 \leq j \leq J} \lambda_j d^p(x_i, T(x)) \right] \wedge R d\gamma(x) &\leftarrow \int \left[\sum_{1 \leq j \leq J} \lambda_j d^p(x_i, T(x)) \right] \wedge R d\gamma_n(x) \\
&\leq \sum_{1 \leq j \leq J} \lambda_j W_p^p(\mu_B^n, \mu_j^n) \\
&\leq \sum_{1 \leq j \leq J} \lambda_j W_p^p(\mu_B, \mu_j^n) \rightarrow \sum_{1 \leq j \leq J} \lambda_j W_p^p(\mu_B, \mu_j),
\end{aligned}$$

et puisque

$$\sum_{1 \leq j \leq J} \lambda_j W_p^p(T_{\#}\gamma, \mu_j) \leq \lim_{R \rightarrow \infty} \int \left[\sum_{1 \leq j \leq J} \lambda_j d^p(x_i, T(x)) \right] \wedge Rd\gamma(x),$$

alors $T_{\#}\gamma$ est un p -barycentre de $(\mu_j)_{1 \leq j \leq J}$, que l'on note aussi μ_B , sans perte de généralité. Enfin, puisque, au moins le long d'une sous-suite,

$$W_p(\mu_B^n, \mu_1) \rightarrow W_p(\mu_B, \mu_1),$$

alors le lemme 4.14 assure

$$W_p(\mu_B^n, \mu_B) \rightarrow 0.$$

□

Démonstration du corollaire 4.13. Soient

$$\mu_B^n \in \arg \max \left\{ \inf_{\nu \in B} W_p(\nu, \mu); \mu \in B^n \right\}$$

pour tout $n \geq 1$. On a alors, pour tout $\mu_B \in B$,

$$\text{Haus}(B^n|B) \leq W_p(\mu_B, \mu_B^n). \quad (4.8)$$

Supposons qu'il existe une sous-suite telle que $\lim \text{Haus}(B^n|B) > 0$. Le théorème 4.9 montre alors que l'on peut en extraire une nouvelle sous-suite qui converge vers un $\mu_B \in B$, en distance de Wasserstein, ce qui contredit (4.8), et montre alors le corollaire. □

Annexe A

Rappels et notations

A.1 Rappels et lemmes

Dans cette section, nous énonçons quelques théorèmes et lemmes qui sont utilisés dans cette thèse. Les théorèmes sont énoncés sans démonstrations.

Lemme A.1. On pose $\Phi(t) = \int_{-\infty}^t e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}$. Alors, pour tout $0 < u < 1$,

$$-\sqrt{-2 \log u} \leq \Phi^{-1}(u).$$

Démonstration. C'est une conséquence de la convexité de $u \mapsto \Phi(-\sqrt{-2 \log u})$. \square

Lemme A.2. Soit \mathcal{F} un ensemble stable d'ensembles connexes d'un espace métrique (E, d) totalement compact. On suppose qu'il existe une union croissante $(K_k)_{k \geq 1}$ de boules vérifiant $\cup_{k \geq 1} K_k = E$ telle que pour tout $k \geq 1$ et tout $F \in \mathcal{F}$, $K_k \cap F$ est encore connexe, alors

$$\mathcal{F}_\varepsilon := \{\cup_{F \in \mathcal{G}} F; \mathcal{G} \subset \mathcal{F}, \forall F, G \in \mathcal{G} \inf_{x \in F, y \in G} d(x, y) \geq \varepsilon\},$$

est aussi un ensemble stable.

Démonstration. Soit une suite $(B_n)_{n \geq 1} \subset \mathcal{F}_\varepsilon$. La remarque 2.4 permet d'affirmer qu'il existe un fermé B_∞ tel qu'une sous-suite de $(B_n)_{n \geq 1}$ converge vers B_∞ pour la convergence de Hausdorff généralisée. Supposons qu'il existe deux composantes connexes $F_1, F_2 \subset B_\infty$ telles que

$$\underline{d}(F_1, F_2) := \inf\{d(x, y); x \in F_1, y \in F_2\} < \varepsilon.$$

Alors, puisque E est totalement compact, pour toute boule K telle que $K \cap F_1 \neq \emptyset$ et $K \cap F_2 \neq \emptyset$, il existe $x_1 \in F_1 \cap K$ et $x_2 \in F_2 \cap K$ tels que $\underline{d}(F_1 \cap K, F_2 \cap K) = d(x_1, x_2)$. La convergence de Hausdorff généralisée affirme par ailleurs,

$$d_H(B_n \cap K, B_\infty \cap K) \rightarrow 0.$$

En choisissant alors $K \in \{K_k\}_{k \geq 1}$ assez grand de sorte que

$$\underline{d}(F_1 \cap K, F_2 \cap K) = \eta < \varepsilon.$$

et en posant,

$$\delta = \underline{d}(K \cap F_1, K \cap B_\infty \setminus F_1) \wedge \underline{d}(K \cap F_1, K \cap B_\infty \setminus F_1) > 0,$$

il vient que pour n assez grand,

$$F_1 \cap K \subset B_n^{((\varepsilon-\eta) \wedge \delta)/3} \cap K \text{ et } F_2 \cap K \subset B_n^{((\varepsilon-\eta) \wedge \delta)/3}, \quad (\text{A.1})$$

$$B_n \cap K \subset B_\infty^{\delta/2} \cap K. \quad (\text{A.2})$$

(A.1) montre qu'il existe alors $y_1, y_2 \in B_n \cap K$ tels que $d(x_1, y_1) \leq ((\varepsilon - \eta) \wedge \delta)/3$ et $d(x_2, y_2) \leq ((\varepsilon - \eta) \wedge \delta)/3$ et donc

$$\begin{aligned} d(y_1, y_2) &\leq d(y_1, x_1) + d(x_1, x_2) + d(x_2, y_2) \\ &\leq 2(\varepsilon - \eta)/3 + \eta \\ &\leq 2\varepsilon/3 + \eta/3 < \varepsilon. \end{aligned} \quad (\text{A.3})$$

Et (A.2) montre que $B_n \cap K \cap F_1^{\delta/2} \subset F_1^{\delta/4} \cap K$ et $B_n \cap K \cap F_2^{\delta/2} \subset F_2^{\delta/4} \cap K$, puisque $K \cap B_\infty^{\delta/4} \cap F_i^{\delta/2} \subset K \cap F_i^{\delta/4}$ pour $i = 1, 2$ par définition de δ . Il s'ensuit que $y_1 \in B_n \cap K \cap F_1^{\delta/2}$ et $y_2 \in B_n \cap K \cap F_2^{\delta/2}$ sont inclus deux composantes connexes distinctes de $B_n \cap K$, ce qui contredit (A.3) et l'hypothèse que $B_n \in \mathcal{F}_\varepsilon$. \square

Théoreme A.3 (Théorème du graphe fermé). *Soient E et F deux espaces de Banach et f est une application linéaire de E dans F . Alors f est continue si et seulement si le graphe de f est une partie fermée de $E \times F$.*

Théoreme A.4 (Théorème du portemanteau). *Soit $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité sur un espace polonais (E, d) . Alors les quatre assertions suivantes sont équivalentes :*

- μ_n converge faiblement vers μ ,
- pour tout fermé F de E , $\limsup_n \mu_n(F) \leq \mu(F)$,
- pour tout ouvert O de E , $\liminf_n \mu_n(O) \geq \mu(O)$,
- pour tout borélien A de E tel que $\mu(\partial A) = 0$, $\lim_n \mu_n(A) = \mu(A)$.

Théoreme A.5 (Distance de Prokhorov). *Soit (E, d) un espace métrique. Pour μ et ν deux mesures de probabilité sur E , on pose*

$$d_P(\mu, \nu) = \inf\{\eta > 0; \mu(B) \leq \nu(B^\eta) + \eta, \forall B \in \mathbb{B}(E)\}.$$

Alors d_P est une distance sur l'espace des mesures de probabilité sur E , appelée distance de Prokhorov, et cette distance métrise la topologie de la convergence faible.

Théoreme A.6 (Inégalité de Jensen). *Soient μ une mesure de probabilité sur \mathbb{R} admettant un moment d'ordre 1, et f une fonction convexe sur \mathbb{R} . Alors,*

$$f\left(\int x d\mu(x)\right) \leq \int f(x) d\mu(x).$$

A.2 Notations

Voici la liste des notations utilisées dans la thèse.

d_H	distance de Hausdorff
\mathcal{S}^{d-1}	sphère unité de dimension $d - 1$
E	espérance mathématique,
$\mathbf{1}_A$	fonction indicatrice de A ,
∂B	bord topologique de B ,
$I(X)$	intégrale de Dudley (section 1.1.5),
$W_p(\mu, \nu)$	distance de Wasserstein entre μ et ν (définition 1.28),
B^ε	ε -épaississement de B ,
$\mathcal{P}_p(E)$	espace de Wasserstein sur E (définition 1.28)
$\mathbb{B}(E)$	ensemble des boréliens de E ,
$\mathcal{N}(B, t)$	t -recouvrement de B (définition 1.1),
$\mathcal{M}(B, t)$	t -packing de B (définition 1.3),
$\lim_n B_n$	limite d'ensembles (définition 2.1)
$\text{Haus}(A B)$	contraste de Hausdorff (définition 2.9)
$T_\# \mu$	mesure image de μ par T ,
$\Gamma(\mu_1, \dots, \mu_J)$	ensemble des mesures avec pour j -ème marginale μ_j , pour tout $1 \leq j \leq J$
$\Gamma(C_1, \dots, C_J)$	union des $\Gamma(\mu_1, \dots, \mu_J)$, pour $\mu_j \in C_j$, pour tout $1 \leq j \leq J$
$\mu \otimes \nu$	produit tensoriel des mesures μ et ν ,
$a \vee b$	maximum de a et b ,
$a \wedge b$	minimum de a et b ,
$\mu _K$	mesure μ restreinte à K .
μK	probabilité conditionnelle de μ sachant K .

Bibliographie

- [Agueh et Carlier, 2010] AGUEH, M. et CARLIER, G. (2010). Barycenters in the Wasserstein space. Preprint.
- [Ajtai et al., 1984] AJTAI, M., KOMLOS, J. et TUSNÁDY, G. (1984). On optimal matchings. Combinatorica, 4(4):259–264.
- [Barthe et Bordenave, 2011] BARTHE, F. et BORDENAVE, C. (2011). Combinatorial optimization over two random point sets. ArXiv preprint arXiv :1103.2734v1.
- [Boissard et Le Gouic, 2011] BOISSARD, E. et LE GOUIC, T. (2011). On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. ArXiv e-prints.
- [Boissard et al., 2011] BOISSARD, E., LE GOUIC, T. et LOUBES, J.-M. (2011). Distribution’s template estimate with Wasserstein metrics. ArXiv e-prints.
- [Bolley et al., 2007] BOLLEY, F., GUILLIN, A. et VILLANI, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. Probability Theory and Related Fields, 137:541–593.
- [Borell, 1975] BORELL, C. (1975). The Brunn-Minkowski inequality in Gauss space. Inventiones mathematicae, 30:207–216.
- [Burago et al., 2001] BURAGO, D., BURAGO, I. et IVANOV, S. (2001). A Course in Metric Geometry. Graduate Studies in Mathematics. American Mathematical Society.
- [Cuesta-Albertos et al., 1997] CUESTA-ALBERTOS, J. A., GORDALIZA, A. et MATRÁN, C. (1997). Trimmed k-means : An attempt to robustify quantizers. The Annals of Statistics, 25(2):pp. 553–576.
- [Del Barrio et al., 1999] DEL BARRIO, E., GINÉ, E. et MATRÁN, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. Annals of Probability, 27(2):1009–1071.
- [Dereich et al., 2003] DEREICH, S., FEHRINGER, F., MATOUSSI, A. et SCHEUTZOW, M. (2003). On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. Journal of Theoretical Probability, 16(1):249–265.

- [Dobric et Yukich, 1995] DOBRIC, V. et YUKICH, J. (1995). Exact asymptotics for transportation cost in high dimensions. J. Theoretical Prob., pages 97–118.
- [Dudley, 1969] DUDLEY, R. (1969). The speed of mean Glivenko-Cantelli convergence. The Annals of Mathematical Statistics, 40(1):40–50.
- [Dudley, 1967] DUDLEY, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. J. Functional Analysis, 1:290–330.
- [Falconer, 2003] FALCONER, K. (2003). Fractal Geometry : Mathematical Foundations and Applications. Wiley, 2 édition.
- [Gangbo et Świąch, 1998] GANGBO, W. et ŚWIĘCH, A. (1998). Optimal maps for the multidimensional Monge-Kantorovich problem. Communications on pure and applied mathematics, 51(1):23–45.
- [Gozlan et Léonard, 2007] GOZLAN, N. et LÉONARD, C. (2007). A large deviation approach to some transportation cost inequalities. Probability Theory and Related Fields, 139:235–283.
- [Graf et Luschgy, 2000] GRAF, S. et LUSCHGY, H. (2000). Foundations of quantization for probability distributions. Springer-Verlag New York, Inc. Secaucus, NJ, USA.
- [Graf et al., 2003] GRAF, S., LUSCHGY, H. et PAGÈS, G. (2003). Functional quantization and small ball probabilities for Gaussian processes. Journal of Theoretical Probability, 16(4):1047–1062.
- [Hartigan, 1987] HARTIGAN, J. A. (1987). Estimation of a convex density contour in two dimensions. Journal of the American Statistical Association, 82(397):267–270.
- [Horowitz et Karandikar, 1994] HOROWITZ, J. et KARANDIKAR, R. (1994). Mean rates of convergence of empirical measures in the Wasserstein metric. Journal of Computational and Applied Mathematics, 55(3):261–273.
- [Kolmogorov et Tihomirov, 1961] KOLMOGOROV, A. N. et TIHOMIROV, V. M. (1961). ε -Entropy and ε -Capacity of Sets in Functional Spaces. American Mathematical Society.
- [Kuelbs et Li, 1993] KUELBS, J. et LI, W. (1993). Metric entropy and the small ball problem for Gaussian measures. Journal of Functional Analysis, 116(1):133–157.
- [Ledoux et Talagrand, 1991] LEDOUX, M. et TALAGRAND, M. (1991). Probability in Banach spaces, volume 23 de Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. Springer-Verlag, Berlin. Isoperimetry and processes.
- [Nolan, 1991] NOLAN, D. (1991). The excess-mass ellipsoid. Journal of Multivariate Analysis, 39(2):348 – 371.

- [Nolan, 1992] NOLAN, D. (1992). Asymptotics for multivariate trimming. Stochastic Processes and their Applications, 42(1):157 – 169.
- [Polonik, 1997] POLONIK, W. (1997). Minimum volume sets and generalized quantile processes. Stochastic Processes and their Applications, 69(1):1 – 24.
- [Rabin et al., 2012] RABIN, J., PEYRÉ, G., DELON, J. et BERNOT, M. (2012). Wasserstein barycenter and its application to texture mixing. 6667:435–446.
- [Saint Raymond et Tricot, 1988] SAINT RAYMOND, X. et TRICOT, C. (1988). Packing regularity of sets in n -space. In Proc. Cambridge Philos. Soc, volume 103, pages 133–145. Cambridge Univ Press.
- [Suquet, 2006] SUQUET, C. (2006). Probabilités dans les espaces fonctionnels.
- [Talagrand, 1992] TALAGRAND, M. (1992). Matching random samples in many dimensions. The Annals of Applied Probability, 2(4):846–856.
- [Varadarajan, 1958] VARADARAJAN, V. (1958). On the convergence of sample probability distributions. Sankhyā : The Indian Journal of Statistics, 19(1):23–26.
- [Villani, 2009] VILLANI, C. (2009). Optimal transport. Old and new., volume 338. Grundlehren der Mathematischen Wissenschaften 338. Berlin : Springer. xxii,.