



# Non-parametric estimation of convex bodies and convex polytopes

Victor-Emmanuel Brunel

## ► To cite this version:

Victor-Emmanuel Brunel. Non-parametric estimation of convex bodies and convex polytopes. Statistics [math.ST]. Université Pierre et Marie Curie - Paris VI; University of Haifa, 2014. English. NNT : . tel-01066977

**HAL Id: tel-01066977**

**<https://theses.hal.science/tel-01066977>**

Submitted on 23 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Non parametric estimation of convex bodies and convex polytopes

---

*By:*

Victor-Emmanuel Brunel

*Under the supervision of:*

Prof. Alexander Goldenshluger, University of Haifa

and Prof. Alexandre Tsybakov, UPMC

A Thesis Submitted in Partial Fulfillment of the Requirements for the  
Degree of "Doctor in Mathematics" at UPMC  
and for the Degree of "Doctor of Philosophy" at the University of Haifa  
*defended on July 4th, 2014.*

JURY:

Prof. Lucien BIRGE	Examinator
Prof. Xiahong CHEN	Examinator
Prof. Antonio CUEVAS	Examinator
Prof. Arnak DALALYAN	Examinator
Prof. Alexander GOLDENSHLUGER	Advisor
Prof. Adityanand GUNTUBOYINA	Referee
Prof. Oleg LEPSKI	Referee
Prof. Alexandre TSYBAKOV	Advisor

Date of submission: May 12, 2014

Copyright © 2014 by Victor-Emmanuel Brunel.

All rights reserved.

# Acknowledgements

Is it possible to write a PhD dissertation, and skip the acknowledgements page ? What would a PhD dissertation be without that page ? Well, let me put that question this way. What would my PhD have been without Sacha Tsybakov and Sasha Goldenshluger as my advisors, who were always available to discuss mathematical questions, who were always able to show me how most of the problems could be reformulated in a very easy way, and thanks to whom I have now very rich scientific perspectives ? Without the Caesarea Rothschild Institute, at the University of Haifa, where I was kindly welcome during the first year of my PhD ? Without Arnak Dalalyan, Pierre Bellec (one of my scientific brothers) and Nicolas Chopin, my colleagues who are always, no matter the time, keen on talking about mathematics, and spent many coffee (or tea) breaks, writing on the white board or just discussing interesting questions ? Without these people whom I sincerely admire and whom I am always pleased to meet, during scientific conferences, seminars, or any other occasions, Pierre A., Michaël C., Céline D., Magalie F., Eric G., Mohammed H., Olga K., Vladimir K., Guillaume L., Oleg L., Karim L., Katia M., Sasha R., Philippe R., Joseph S. ? Without the great ENSAE team, with whom I have been working as an "assistant de statistique" during the last three years, Elise C., Romain A., Catherine L., Ronan L. S., Alexandre D. C., Alexandre G., Xavier M., Julien C., Malika Z., Kambiz M., Julien S., the kindest people in the Service des stages, Luciana B., Marie-Françoise J., Françoise M., and also Philippe C., Olivier L., Judith R., if mentioning only these colleagues ?

The following theorem provides an answer to that question :

**Theorem.** *Let the previous question hold. Then, the answer is the following : It would not have been a challenging, enriching, interesting PhD, spending four years in such a great, scientific, and family environment.*

As a corollary, I decided to write an acknowledgements page. So here it is, a warm THANK YOU to all the people quoted above. I would also like to thank Aditya Guntuboyina and Oleg Lepski for writing such encouraging reports on my dissertation in a very short term, and the

kind and benevolent members of the jury of my PhD defense, as well as all of those who attended the defense and made it such a beautiful day.

# Abstract (English)

In this thesis, we are interested in statistical inference on convex bodies in the Euclidean space  $\mathbb{R}^d$ . Two models are investigated. The first one consists of the observation of  $n$  independent random points, with common uniform distribution on an unknown convex body. The second one is a regression model, with additive subgaussian noise, where the regression function is the indicator function of an unknown convex body. In the first model, our goal is to estimate the unknown support of the common uniform density of the observed points. In the second model, we aim either to estimate the support of the regression function, or to detect whether this support is nonempty, i.e., the regression function is nonzero. In both models, we investigate the cases when the unknown set is a convex polytope, and when we know the number of vertices. If this number is not known, we propose an adaptive method which allows us to obtain a statistical procedure performing asymptotically as well as in the case of perfect knowledge of that number. In addition, this procedure allows misspecification, i.e., provides an estimator of the unknown set, which is optimal in a minimax sense, even if the unknown set is not polytopal, in the contrary to what may have been thought.

We prove a universal deviation inequality for the volume of the convex hull of the observations in the first model. We show that this inequality allows one to derive tight bounds on the moments of the missing volume of this convex hull, as well as on the moments of the number of its vertices.

In the one-dimensional case, in the second model, we compute the asymptotic minimal size of the unknown set so that it can be detected by some statistical procedure, and we propose a decision rule which allows consistent testing of whether of that set is empty.

# Résumé (Français)

Dans ce travail, nous nous intéressons à l'estimation d'ensembles convexes dans l'espace Euclidien  $\mathbb{R}^d$ , en nous penchant sur deux modèles. Dans le premier modèle, nous avons à notre disposition un échantillon de  $n$  points aléatoires, indépendants et de même loi, uniforme sur un ensemble convexe inconnu. Le second modèle est un modèle additif de régression, avec bruit sous-gaussien, et dont la fonction de régression est l'indicatrice d'Euler d'un ensemble convexe ici aussi inconnu. Dans le premier modèle, notre objectif est de construire un estimateur du support de la densité des observations, qui soit optimal au sens minimax. Dans le second modèle, l'objectif est double. Il s'agit de construire un estimateur du support de la fonction de régression, ainsi que de décider si le support en question est non vide, c'est-à-dire si la fonction de régression est effectivement non nulle, ou si le signal observé n'est que du bruit.

Dans ces deux modèles, nous nous intéressons plus particulièrement au cas où l'ensemble inconnu est un polytope convexe, dont le nombre de sommets est connu. Si ce nombre est inconnu, nous montrons qu'une procédure adaptative permet de construire un estimateur atteignant la même vitesse asymptotique que dans le cas précédent. Enfin, nous démontrons que ce même estimateur pallie à l'erreur de spécification du modèle, consistant à penser à tort que l'ensemble convexe inconnu est un polytope.

Nous démontrons une inégalité de déviation pour le volume de l'enveloppe convexe des observations dans le premier modèle. Nous montrons aussi que cette inégalité implique des bornes optimales sur les moments du volume manquant de cette enveloppe convexe, ainsi que sur les moments du nombre de ses sommets.

Enfin, dans le cas unidimensionnel, pour le second modèle, nous donnons la taille asymptotique minimale que doit faire l'ensemble inconnu afin de pouvoir être détecté, et nous proposons une règle de décision, permettant un test consistant du caractère non vide de cet ensemble.





# Notation

$\rho_d$  : Euclidean distance in  $\mathbb{R}^d$ .

$B_2^d$  : Euclidean centered unit closed ball in  $\mathbb{R}^d$ .

$B_2^d(a, r)$  : Euclidean closed ball with center  $a \in \mathbb{R}^d$  and radius  $r \geq 0$ .

$\beta_d$  : Volume of  $B_2^d$ .

$\bar{G}, \overset{\circ}{G}, \partial G$  : Closure, interior and boundary of the set  $G$ .

$CH(S)$  : Convex hull of the set  $S$ .

$\mathcal{K}_d$  : Class of all convex bodies in  $\mathbb{R}^d$ .

$\mathcal{K}_d^{(1)}$  :  $\{G \in \mathcal{K}_d : G \subseteq [0, 1]^d\}$ .

$\mathcal{P}_r$  : Class of all convex polytopes in  $\mathbb{R}^d$  with at most  $r$  vertices, and with positive volume  
( $r \geq d + 1$ ).

$\mathcal{P}_r^{(1)}$  :  $\{P \in \mathcal{P}_r : P \subseteq [0, 1]^d\}, r \geq d + 1$ .

$\mathcal{P}_{r,n}^{(1)}$  : Class of all polytopes in  $\mathcal{P}_r^{(1)}$ , with vertices whose coordinates are integer multiples of

$1/n$  ( $r \geq d + 1, n \in \mathbb{N}^*$ )

$\mathcal{P}_\infty^{(1)} = \mathcal{K}_d^{(1)}$ .

$\mathcal{P} = \bigcup_{r \geq d+1} \mathcal{P}_r$ .

$\mathcal{P}^{(1)} = \bigcup_{r \geq d+1} \mathcal{P}_r^{(1)}$ .

$f_k(P)$  : number of  $k$ -faces of the polytope  $P$ .

$f_k(r, d)$  : number of  $k$ -faces of a  $d$ -dimensional cyclic polytope with  $r$  vertices.

$G^\epsilon$  : Closed  $\epsilon$ -neighborhood of the set  $G$ , defined as  $G + \epsilon B_2^d$ .

$\mathbb{1}(\cdot \in G)$  : Indicator function of the set  $G$ .

$|\cdot|_d$  : Lebesgue measure in  $\mathbb{R}^d$ .

$G_1 \triangle G_2$  : Symmetric difference between the two sets  $G_1$  and  $G_2$ .

$d_H(G_1, G_2)$  : Hausdorff distance between the two sets  $G_1$  and  $G_2$ .

$\mathbb{P}_G$  : Probability measure associated to the set  $G$ .

$\mathbb{P}_f$  : Probability measure associated to the density  $f$ .

$\mathbb{E}$  : Expectation operator.

$\mathbb{E}_G$  : Expectation operator corresponding to  $\mathbb{P}_G$ .

$\mathbb{E}_f$  : Expectation operator corresponding to  $\mathbb{P}_f$ .

$\mathbb{V}$  : Variance operator.

$\mathbb{V}_G$  : Variance operator corresponding to  $\mathbb{P}_G$ .

$\mathbb{P}^{\otimes n}$  :  $n$ -product of the probability measure  $\mathbb{P}$ .

$\mathbb{E}^{\otimes n}$  : Expectation operator corresponding to  $\mathbb{P}^{\otimes n}$ .

$H(\mathbb{P}, \mathbb{Q})$  : Hellinger distance between the two probability measures  $\mathbb{P}$  and  $\mathbb{Q}$ .

$\lfloor \cdot \rfloor$  : Integer part.

$L^p(E)$  : Set of real valued and Lebesgue-measurable functions defined on the Borel set  $E$ , such that  $\int_E |f|^p < \infty, p \geq 1$ .

$\|\cdot\|_p$  :  $L^p$ -norm,  $1 \leq p < \infty$ .

$\|\cdot\|_\infty$  :  $L^\infty$ -norm.

$f^+ = \max(f, 0)$ .

$$\binom{N}{m} = \frac{N!}{m!(N-m)!}.$$

$\mathcal{R}_n(\hat{G}_n; \mathcal{C})$  : Risk of the estimator  $\hat{G}_n$  on the class  $\mathcal{C}$ .

$\mathcal{R}_n(\mathcal{C})$  : Minimax risk on the class  $\mathcal{C}$ .

$\mathcal{Q}_n(\mathcal{C})$  : Minimax weighted risk on the class  $\mathcal{C}$ .

In this work, we are generally not interested in the explicit form of the constants, but only in their dependence on the parameters. The sign ‘ $c$ ’, followed by some indexes or arguments, will be used for the constants, and its value may vary along the thesis.

Most of the results presented in this thesis are based on [Bru13], [Bru14c], [Bru14a] and [Bru14b]. In addition, some unpublished results, and reviews of the existing literature are included.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Why set estimation ? . . . . .	1
1.2	Definitions and Notation . . . . .	3
1.2.1	Set-valued estimators . . . . .	3
1.2.2	Notation and first definitions . . . . .	4
1.2.3	The minimax setup . . . . .	8
1.2.4	Adaptation and misspecification . . . . .	10
1.3	The statistical models . . . . .	11
<b>2</b>	<b>Estimation of the convex or polytopal support of a uniform density</b>	<b>14</b>
2.1	Random polytopes . . . . .	15
2.1.1	The wet part and the floating convex body . . . . .	16
2.1.2	Convergence of the random convex hull . . . . .	18
2.1.3	Asymptotic facial structure . . . . .	21
2.2	Minimax estimation . . . . .	22
2.2.1	Estimation of polytopes . . . . .	23
2.2.2	Estimation of convex bodies . . . . .	27
2.2.3	Adaptation and misspecification . . . . .	27
2.3	Further results and open problems . . . . .	29
2.3.1	A universal deviation inequality for the random convex hull . . . . .	29
2.3.2	Moments of the number vertices of the convex hull estimator . . . . .	31
2.3.3	Efron's identity revisited . . . . .	33
2.3.4	Non uniform distributions . . . . .	38
2.3.5	Estimation of log-concave densities . . . . .	41
2.4	Proofs . . . . .	42
2.5	Appendix to Chapter 2 . . . . .	48

2.5.1	Lemmata . . . . .	48
2.5.2	Proofs of Lemmata and others . . . . .	49
<b>3</b>	<b>A nonparametric regression model</b>	<b>53</b>
3.1	Estimation of polytopes . . . . .	54
3.1.1	Upper bound . . . . .	54
3.1.2	Lower bound . . . . .	55
3.2	Estimation of convex bodies . . . . .	56
3.3	Adaptation and misspecification . . . . .	58
3.4	Discussion . . . . .	59
3.5	The one-dimensional case and the change point problem . . . . .	61
3.5.1	Detection of a segment . . . . .	65
3.5.2	Estimation of a segment . . . . .	66
3.5.3	Conclusion and discussion . . . . .	71
3.6	Proofs . . . . .	72
3.7	Appendix to Chapter 3 . . . . .	95
3.7.1	Lemmata . . . . .	95
3.7.2	Proof of the lemmata . . . . .	95
<b>4</b>	<b>Estimation of functionals</b>	<b>98</b>
4.1	The density support model . . . . .	99
4.1.1	Estimation of the volume of a convex body . . . . .	99
4.1.2	Estimation of the area of a triangle . . . . .	103
4.1.3	Estimation of intrinsic volumes . . . . .	104
4.2	The regression model . . . . .	106
<b>5</b>	<b>Conclusion, contributions</b>	<b>108</b>

# Chapter 1

## Introduction

### 1.1 Why set estimation ?

Geometry and probability have been related since the eighteenth century, when some fundamental questions were posed. Here are two examples of questions that received much attention. If a needle is dropped on a wooden floor, what is the probability that it does not intersect a gap of that floor ? If four points are chosen at random in the plane, what is the probability that one of them is in the triangle made by the three others ? The first question is known for having been addressed by the French Georges-Louis Leclerc, Comte de Buffon. The second question is well known under the name of *Sylverster's four-point problem*. Contradictory answers were brought to each of these two questions, and despite their differences, they all seemed correct. The reason was that these questions were not well posed. In particular, it was necessary to define the probability measures which described best the randomness of the experiments, in order to give a strict answer. This need was beyond the scope of the theory of random numbers. Probability measures were to be defined for geometrical objects. This was more or less the birth of what we call today *stochastic geometry*. Quite rapidly, new challenging questions were posed, related to this new field. First leading works were those of Rényi, Sulanke and Efron, in the 1960's. The study of probabilistic and geometrical properties of random geometrical objects - e.g. random points, Poisson point processes, random sets - became a hot topic. However, statistical inference on such objects appeared much later.

One purpose of statistics is to get as precise information as possible about an intricate system, of which only a little information can be obtained, whether clean or noisy. For instance, the blurred aspect of a picture can be modeled as the result of a random process, whose properties can be described by few parameters only. In this case, statistical modeling is an artificial tool,

which allows one to describe as simply as possible a macroscopic system, whose microscopic details behave erratically. Blurred images can be recovered this way: assuming the *noise* is random and homogeneous all over the image. In a survey, when the respondents provide ranks instead of precise values to certain questions, or in order to deal with missing answers, the theory of partial identification, based on the theory of random sets, allows one to recover some information [Man03, BMM10].

Statistical inference on sets has gained attractiveness in the last three decades. In particular, estimation of the support of a density, or more generally, of the level sets of a density, detection of an object in a blurred - or partially observed - image, or detection of a signal, classification and clustering, all these possible applications and others, made indispensable statistical inference on sets, in particular set estimation. Set estimation was probably for the first time introduced by Geffroy in 1964 [Gef64], followed by Chevalier [Che76]. Chevalier's work deals with estimation of the support of a density in general metric spaces, under very general assumptions. A simple estimator, which was also studied by Devroye and Wise in 1980 [DW80], consists of the union of small Euclidean balls centered at the points of the sample. Consistency of this estimator is ensured if the radius  $\epsilon_n$  of these balls -  $n$  being the total number of available observations - satisfies both  $\epsilon_n \rightarrow 0$  and  $n\epsilon_n^d \rightarrow \infty$ , as  $n \rightarrow \infty$ , where  $d$  is the dimension of the ambient Euclidean space. Notably, these two conditions remind those for the choice of bandwidth in Kernel density estimation, leading to consistency. In fact, this estimator is exactly the support of the Kernel density estimator, with Kernel  $K(x) = (\rho_d(x, 0) \leq \epsilon), x \in \mathbb{R}^d$ . Estimation of boundary fragments, i.e., subsets  $G$  of  $\mathbb{R}^d$  which can be described as the subgraph of a positive function  $g : [0, 1]^{d-1} \rightarrow [0, 1]$ :

$$G = \left\{ (x_1, \dots, x_d) \in [0, 1]^d : 0 \leq x_d \leq g(x_1, \dots, x_{d-1}) \right\}, \quad (1.1)$$

attracted much attention in the 1980's, see for example [KT93a] where  $g$  satisfies Hölder conditions, or [KST95] where  $g$  is monotone or convex. In particular, the function  $g$  in (1.1) can be interpreted as the efficiency frontier of the productivity function of a firm. Estimation of more general sets, under shape restrictions, was considered as well. In [BCJ00, BC01], consistency of similar estimators to that of [DW80] is proved, when the target is connected, or star-shaped. Estimation of convex sets is studied as a particular case of estimation of sets with smooth boundaries in [MT95], and analytical methods, based on functional analysis and not on geometrical considerations, are developed there. Optimality of the estimators in a minimax



sense (see Section 1.2.3 for precise definitions) is proved in that work. We refer to [Tsy94] where estimation of sets is discussed under various statistical models and different kinds of assumptions on the classes of unknown sets.

## 1.2 Definitions and Notation

### 1.2.1 Set-valued estimators

An important part of statistics deals with estimation. From a sample of observations, say,  $\mathcal{X}$ , one may expect to be able to understand better the underlying mechanism, which created that sample. This mechanism may be Nature, or some intricate machinery. For instance, take  $\mathcal{X} = \{X_1, \dots, X_n\}$ , where  $X_i$  is a binary variable giving the sex of the  $i$ -th newborn from a sample of  $n$  newborns. Then  $X_i$  can be - maybe artificially - seen as the realization of a random variable, whose distribution is Bernoulli, with some parameter  $p_i \in (0, 1)$ . A common modeling of this situation consists in assuming that the parameters  $p_i$  take one common value for all  $i$ , say  $p$ , and that the random variables  $X_1, \dots, X_n$  are mutually independent. In other words, the observations are assumed to be the realization of one and the same phenomenon repeated  $n$  times. Given this modeling, the only unknown quantity is the value of the parameter  $p$ . The statistician would like to use the observations  $X_1, \dots, X_n$ , in order to have a precise idea of the value of that number  $p$ . One way to do this is to estimate  $p$ . An estimator of  $p$  is a random variable depending on the observations. More precisely, in this example, an estimator can be written as  $S_n(X_1, \dots, X_n)$ , where  $S_n$  is a measurable function from  $\{0, 1\}^n$  onto  $(0, 1)$ .

Let us now be more general. Consider a finite sample  $\mathcal{X}$ , consisting of  $n$  observations, each of which belongs to a given measurable set  $\mathcal{E}$ . Assume that these observations are realizations of random variables, whose distribution depends on one unknown quantity  $\theta \in \Theta$ , where  $\Theta$  is a measurable set. An estimator of the unknown  $\theta$  is a random variable of the form  $S_n(X_1, \dots, X_n)$ , where  $S_n : \mathcal{E}^n \rightarrow \Theta$  is a measurable function. Therefore, constructing an estimator requires a structure on the set which contains the unknown quantity of interest. In this thesis, we deal with estimation of sets and, more precisely, of convex bodies. Hence, we need to allow  $\Theta$  to be the class of all convex bodies in  $\mathbb{R}^d$ , where  $d$  is a given positive integer. However, there is no natural way to provide a non trivial measurability structure on this class. A particular definition of set-valued random variables is given in [Mol05]. For a given topological space  $E$ , consider the class  $\mathcal{F}$  of all closed subsets of  $E$ . The Effros  $\sigma$ -algebra on  $\mathcal{F}$  is defined as the  $\sigma$ -algebra generated by  $\{F \in \mathcal{F} : F \cap K \neq \emptyset\}$ , where  $K$  runs through the family of all compact

subsets of  $E$ . Denote the Effros  $\sigma$ -algebra by  $\mathfrak{B}(\mathcal{F})$ . Let  $(\Omega, \mathfrak{F}, \mathbb{P})$  be a probability space. A map  $X : \Omega \rightarrow \mathcal{F}$  is measurable with respect to  $\mathfrak{F}$  and  $\mathfrak{B}(\mathcal{F})$  if, and only if, for every compact subset  $K$  of  $E$ ,

$$\{\omega \in \Omega : X(\omega) \cap K \neq \emptyset\} \in \mathfrak{F}. \quad (1.2)$$

If the latest condition is satisfied, the map  $X$  is a random variable, valued in  $\mathcal{F}$ , and is called a random set. Note that this definition restricts the random sets to be closed subsets of  $E$ , but it is not too restrictive for our purposes, since we only consider compact sets. However, Condition (1.2) may be too difficult to check, sometimes, and it is not very useful for our purposes. Indeed, this definition allows one to extend the usual properties of random variables to random sets. For example, the expectation of a random set is defined in [Mol05]. The random sets that we consider are estimators, and we are interested in their pointwise accuracy, defined as their distance to their target. This distance is real-valued, and what is important to us is that it is a random variable. Indeed, this would allow us to compute its expectation, its probability deviations, etc.

We can bypass Condition (1.2) by defining a set-valued estimator, as a set-valued mapping, such that its accuracy, measured with a given distance, is a random variable. This will be the case most of the time. In order to make this even more simple, we prefer to consider outer probabilities, in order to avoid measurability conditions.

### 1.2.2 Notation and first definitions

Before going further in this introduction, we need some notation. In the whole thesis,  $d \geq 1$  denotes a positive integer, and the ambient space is the Euclidean space  $\mathbb{R}^d$ .

**General notation and definitions** Denote by  $\rho_d$  - or simply  $\rho$ , when there is no ambiguity - the *Euclidean distance* in  $\mathbb{R}^d$ , by  $B_2^d$  the *Euclidean unit closed ball* in  $\mathbb{R}^d$ , and by  $\beta_d$  its volume. Denote by  $B_2^d(a, r)$  the Euclidean ball centered at the point  $a \in \mathbb{R}^d$  and of radius  $r \geq 0$ . If  $x \in \mathbb{R}^d$  and  $G \subseteq \mathbb{R}^d$ , we denote by  $\rho_d(x, G)$  the Euclidean distance from  $x$  to the set  $G$ , i.e.,

$$\rho_d(x, G) = \inf_{y \in G} \rho_d(x, y).$$

We use the convention that the infimum on the empty set is infinite.

If  $G \subseteq \mathbb{R}^d$ , we denote by  $\bar{G}$  its *closure*,  $\mathring{G}$  its *interior* and  $\partial G$  its *boundary*, i.e.,  $\partial G = \bar{G} \setminus \mathring{G}$ .

If  $G$  is a closed subset of  $\mathbb{R}^d$  and  $\epsilon$  is a positive number, we denote by  $G^\epsilon$  the *closed*

$\epsilon$ -neighborhood of  $G$ , i.e., the set of all  $x \in \mathbb{R}^d$  such that  $\rho(x, G) \leq \epsilon$  or, in other words,  $G^\epsilon = G + \epsilon B_2^d$ . If  $G$  is any set,  $\mathbb{1}(\cdot \in G)$  stands for the *indicator function* of  $G$ .

The *Lebesgue measure* in  $\mathbb{R}^d$  is denoted by  $|\cdot|_d$ . For brevity, we omit the dependence on  $d$  when there is no ambiguity.

For two sets  $G_1$  and  $G_2$ , their symmetric difference  $G_1 \triangle G_2$  is defined as  $(G_1 \setminus G_2) \cup (G_2 \setminus G_1)$ .

The *Nikodym pseudo-distance* between two measurable subsets  $G_1$  and  $G_2$  of  $\mathbb{R}^d$  is defined as the Lebesgue measure of their symmetric difference, namely  $|G_1 \triangle G_2|_d$ . We will often refer to this pseudo-distance as the Nikodym distance.

The *Hausdorff distance* between two subsets  $G_1$  and  $G_2$  of  $\mathbb{R}^d$  is denoted by  $d_H(G_1, G_2)$ , and is defined as

$$d_H(G_1, G_2) = \inf\{\epsilon > 0 : G_1 \subseteq G_2^\epsilon, G_2 \subseteq G_1^\epsilon\}.$$

Equivalently, one has:

$$d_H(G_1, G_2) = \max \left\{ \inf_{x \in G_1} \rho(x, G_2), \inf_{y \in G_2} \rho(y, G_1) \right\}.$$

If  $x \in \mathbb{R}$ , we denote by  $\lfloor x \rfloor$  its integer part, i.e., the greatest integer smaller or equal to  $x$ .

For  $1 \leq p < \infty$ , and  $E$  a measurable subset of  $\mathbb{R}^d$ , denote by  $L^p(E)$  the set of real valued and measurable functions  $f$ , such that  $\int_E |f|^p < \infty$ . The integral is defined with respect to the Lebesgue measure. The  $L^p(E)$ -norm is denoted by  $\|\cdot\|_p$ , and is defined as  $\|f\|_p = \left( \int_E |f|^p \right)^{\frac{1}{p}}$ . If  $f$  is a bounded function, we set  $\|f\|_\infty = \text{esssup}_{x \in E} |f(x)|$ . If  $f$  is a real-valued function, define  $f^+ = \max(f, 0)$ .

**Geometric notation and definitions** Let  $S$  be a subset of  $\mathbb{R}^d$ . Its *convex hull*, denoted by  $CH(S)$ , is the smallest convex set containing  $S$ . Equivalently, it is the intersection of all convex sets which contain  $S$ . Its *affine hull* is the intersection of all affine subspaces which contain  $S$ .

A convex set is said *k-dimensional* if its affine hull has dimension  $k$  ( $k \in \mathbb{N}$ ). A *convex body* in  $\mathbb{R}^d$  is a compact and convex subset of  $\mathbb{R}^d$  with positive volume. We denote by  $\mathcal{K}_d$  the class of all convex bodies in  $\mathbb{R}^d$ , and by  $\mathcal{K}_d^{(1)}$  the class of those convex bodies that are included in  $[0, 1]^d$ . Note that a convex set in  $\mathbb{R}^d$  has a positive volume if and only if it is  $d$ -dimensional. Thus, considering convex sets of positive volume is equivalent to considering *real d-dimensional* convex sets.

A convex body  $K \in \mathcal{K}^d$  is said to have a smooth boundary if and only if its boundary  $\partial K$  is  $k$  times continuously differentiable - in the sense of differential submanifolds, see [Sch93a,

Section 2.5] -, for some integer  $k \geq 2$ .

A *convex polytope* is the convex hull of a finite set. If  $P = CH(S)$  is a convex polytope,  $S$  being a finite subset of  $\mathbb{R}^d$ , the vertices of  $P$  are those  $x \in S$  such that  $CH(S \setminus \{x\}) \neq P$ . By *polytope*, we will always mean convex polytope.

For an integer  $r \geq d + 1$ , we denote by  $\mathcal{P}_r$  the class of all convex polytopes in  $\mathbb{R}^d$  with at most  $r$  vertices and with positive volume, and by  $\mathcal{P}_r^{(1)}$  the class of those  $P \in \mathcal{P}_r$  such that  $P \subseteq [0, 1]^d$ . We define  $\mathcal{P}_{r,n}^{(1)}$  as the class of all polytopes in  $\mathcal{P}_r^{(1)}$  with vertices whose coordinates are integer multiples of  $1/n$ . We denote by  $\mathcal{P} = \bigcup_{r \geq d+1} \mathcal{P}_r$  the class of all convex polytopes in  $\mathbb{R}^d$ , and by  $\mathcal{P}^{(1)} = \bigcup_{r \geq d+1} \mathcal{P}_r^{(1)}$  the class of all convex polytopes in  $[0, 1]^d$ . We will also use the notation  $\mathcal{P}_\infty^{(1)}$  for the class  $\mathcal{K}_d^{(1)}$ . Note that  $\mathcal{P}^{(1)} \subsetneq \mathcal{P}_\infty^{(1)}$ .

A polytope is the convex hull of finitely many points. The dual form of this definition states that a polytope is a bounded intersection of finitely many closed halfspaces. A polytope can be described either by its vertices - a polytope is the set of convex combinations of its vertices - or as intersection of halfspaces - a polytope is a set of points satisfying a finite number of linear inequalities -. The boundary of a polytope is entirely determined by these halfspaces, and its structure can be viewed from a combinatorial point of view.

**Definition 1.1.** "*Supporting hyperplane*"

Let  $K \in \mathcal{K}_d$  and let  $H$  an affine hyperplane of  $\mathbb{R}^d$ . We call  $H$  a *supporting hyperplane* of  $K$  if and only if  $H \cap K \neq \emptyset$  and there exists a non zero vector  $e$ , orthogonal to  $H$ , such that for all  $t > 0$ ,  $(H + te) \cap K = \emptyset$ .

Let  $x \in \partial K$ . A *supporting hyperplane* of  $K$  at  $x$  is a supporting hyperplane of  $K$  which contains the point  $x$ .

Note that the supporting hyperplane of a convex body at a point of its boundary need not be unique.

**Definition 1.2.** Let  $P$  be a polytope and  $k \in \{0, \dots, d-1\}$  be an integer. A *face* of  $P$  is the intersection of  $P$  with a supporting hyperplane. A *k-face* is a face whose affine hull has dimension  $k$ . The vertices of  $P$  are exactly the 0-faces of  $P$ . The 1-faces of  $P$  are called the *edges* of  $P$ . The  $(d-1)$ -faces of  $P$  are called the *facets* of  $P$ .

There are some universal relations between the numbers of  $k$ -faces of a polytope. The most simple one is Euler's formula. Let  $P$  be any  $d$ -dimensional polytope. Let us denote by  $f_k$  the number of  $k$ -faces of  $P$ , for  $k \in \{0, \dots, d-1\}$ . When  $d = 2$ , it is clear that  $f_0 = f_1$ . If  $d = 3$ , Euler's formula states that:

$$f_0 - f_1 + f_2 = 2. \quad (1.3)$$

This formula still holds in any finite dimension (see [BM71] for one of the first proof of this identity, after several unsuccessful attempts by other mathematicians in the nineteen century):

$$\sum_{k=0}^{d-1} (-1)^k f_k = 1 + (-1)^{d-1}. \quad (1.4)$$

Other combinatorial equations exist, in specific cases, e.g. Dehn-Sommerville equations for simplicial polytopes. We do not provide the details, because we do not consider such cases in this thesis. We refer to [Bro83, Zie95] for a detailed account.

In this thesis, we deal with the classes  $\mathcal{P}_r, r \geq d + 1$ , i.e., the polytopes in  $\mathbb{R}^d$  with known value of  $f_0$  - or, to be more precise, with known upper bound on  $f_0$  -. In dimension 2, the knowledge of  $f_0$  determines completely the value of  $f_1$ :  $f_0 = f_1$ . In higher dimensions, this is not true anymore, and a given value of  $f_0$  may be compatible with different values of the  $f_k$ 's,  $k = 1, \dots, d-1$ . We are interested in upper bounds on  $f_k$ , i.e., in controlling the possible values of  $f_k$ , given  $f_0 \leq r$ . In particular, for  $k = d-1$ , it will be useful in Section 2.2.1 to know if  $f_{d-1}$  can be bounded from above, uniformly on the class  $\mathcal{P}_r$ . The answer is positive, and it is given by Mc Mullen's bound. In order to be more specific, let us denote by  $f_k(P)$  the number of  $k$ -faces of a given polytope  $P \in \mathcal{P}$ . The question addressed in [McM70] consists in maximizing  $f_k(P)$  on the class  $\mathcal{P}_r$ , i.e., finding the maximal number - if not infinite - of  $k$ -faces of a  $d$ -dimensional polytope having  $r$  vertices. Let  $f_k(r, d)$  be the number of  $k$ -faces of a cyclic polytope with  $r$  vertices. A cyclic polytope with  $r$  vertices is the convex hull of  $r$  distinct points on the moment curve  $\{(t, t^2, \dots, t^d) : t \in \mathbb{R}\}$ . Grünbaum [Gru67, Section 4.7] proved that the combinatorial structure of a cyclic polytope with  $r$  vertices does not depend on the choice of the vertices on the moment curve. The *upper bound conjecture*, proved by McMullen in [McM70], states that

$$\max_{P \in \mathcal{P}_r} f_k(P) = f_k(r, d), \quad \forall 1 \leq k < d < r. \quad (1.5)$$

This equation states that among all  $d$ -dimensional polytopes with  $r$  vertices, the cyclic polytopes are those with the largest number of  $k$ -faces, for all  $k = 1, \dots, d-1$ . The number of  $k$ -faces of a cyclic polytope with  $r$  vertices is  $\binom{r}{k}$  for  $k = 0, \dots, \lfloor d/2 \rfloor$ , and is given by the Dehn-Sommerville formula for the other values of  $k$ , see [Bro83, Zie95].

**Probabilistic and statistical notation and definitions** Let  $(A, \mathcal{B})$  be a measurable space, and  $\mathcal{C}$  a class of sets in  $\mathbb{R}^d$ . If  $\{\mathbb{P}_G : G \in \mathcal{C}\}$  is a family of probability measures on  $(A, \mathcal{B})$ , we denote by  $\mathbb{E}_G$  the expectation operator associated with  $\mathbb{P}_G$ , and by  $\mathbb{V}_G$  the corresponding

variance operator, for  $G \in \mathcal{C}$ . We keep the same notation  $\mathbb{P}_G$  and  $\mathbb{E}_G$  for the corresponding outer probability and expectation when necessary, to avoid measurability issues.

If  $f$  is a density, we denote by  $\mathbb{P}_f$  and  $\mathbb{E}_f$  the corresponding probability measure and expectation operator.

Let  $\mathbb{P}$  be a probability measure and  $\mathbb{E}$  the corresponding expectation operator. For any positive integer  $n$ , we denote by  $\mathbb{P}^{\otimes n}$  and  $\mathbb{E}^{\otimes n}$  the  $n$ -product of  $\mathbb{P}$  and its corresponding expectation operator. In particular, if  $X_1, \dots, X_n$  are  $n$  independent, identically distributed (i.i.d.) random variables in  $\mathbb{R}^d$  with probability distribution  $\mathbb{P}$ , then  $\mathbb{P}^{\otimes n}$  is the probability distribution of the vector  $(X_1, \dots, X_n)$ . When there is no ambiguity, we may omit the superscript  $\otimes n$ .

Let  $\mathbb{P}$  and  $\mathbb{Q}$  be two probability measures defined on the same measurable space. Let  $p$  and  $q$  be their respective densities with respect to a common  $\sigma$ -finite dominating measure  $\nu$ . The Hellinger distance between  $\mathbb{P}$  and  $\mathbb{Q}$  is denoted by  $H(\mathbb{P}, \mathbb{Q})$ , and is defined as

$$H(\mathbb{P}, \mathbb{Q}) = \left( \int (\sqrt{p} - \sqrt{q})^2 d\nu \right)^{1/2},$$

Note that this definition does not depend on the choice of the dominating measure  $\nu$ .

### 1.2.3 The minimax setup

Let  $(\Omega, \mathfrak{F}, \mathbb{P})$  be a probability space, and  $(A, \mathcal{B})$  a measurable space. Let  $\mathcal{C}$  a class of measurable subsets of  $\mathbb{R}^d$ , and  $\{\mathbb{P}_G : G \in \mathcal{C}\}$  a family of probability measures on  $(A, \mathcal{B})$ . Let  $n$  be a positive integer, and  $Z_1, \dots, Z_n$  be i.i.d. random variables defined on  $(\Omega, \mathfrak{F})$ , taking values in  $(A, \mathcal{B})$ . Assume that the common probability distribution of  $Z_1, \dots, Z_n$  is  $\mathbb{P}_G$ , for some  $G \in \mathcal{C}$ . This means that for all  $B \in \mathcal{B}$ ,  $\mathbb{P}[Z_1 \in B] = \mathbb{P}[\{\omega \in \Omega : Z_1(\omega) \in B\}] = \mathbb{P}_G[B]$ . The vector  $(Z_1, \dots, Z_n)$  is interpreted as the observation of  $n$  independent realizations of a random variable  $Z$ , having  $\mathbb{P}_G$  as its probability distribution, and the set  $G$  is unknown to the statistician. Let  $\hat{G}_n$  be an estimator of  $G$ , i.e., a set-valued function of  $(Z_1, \dots, Z_n)$ .

We measure the accuracy of  $\hat{G}_n$  in a minimax framework, and we will always use the Nikodym distance. The pointwise error of  $\hat{G}_n$  is its distance to the target, namely  $|\hat{G}_n \triangle G|$ . Its pointwise risk is the expectation of its pointwise error, i.e.,  $\mathbb{E}_G^{\otimes n} [|\hat{G}_n \triangle G|]$ . The uniform risk, or simply the risk of  $\hat{G}_n$  on the class  $\mathcal{C}$ , is defined as

$$\mathcal{R}_n(\hat{G}_n; \mathcal{C}) = \sup_{G \in \mathcal{C}} \mathbb{E}_G [|\hat{G}_n \triangle G|].$$

The rate - a sequence depending on  $n$  - of the estimator  $\hat{G}_n$  on the class  $\mathcal{C}$  is the speed at which

its risk on  $\mathcal{C}$  converges to zero, when the number  $n$  of available observations tends to infinity. For all estimators defined in the sequel we will be interested in upper bounds on their risk, in order to get information about their rates.

The minimax risk on  $\mathcal{C}$ , when  $n$  observations are available, is defined as

$$\mathcal{R}_n(\mathcal{C}) = \inf_{\hat{G}_n} \mathcal{R}_n(\hat{G}_n; \mathcal{C}),$$

where the infimum is taken over all set estimators depending on  $n$  observations. If  $\mathcal{R}_n(\mathcal{C})$  converges to zero, we call a minimax rate of convergence on the class  $\mathcal{C}$  the speed at which  $\mathcal{R}_n(\mathcal{C})$  tends to zero, i.e., a positive sequence  $\psi_n$  converging to zero and such that the positive sequence  $\psi_n^{-1} \mathcal{R}_n(\mathcal{C})$  is both bounded from above, and bounded from below by a positive constant.

It is interesting to provide a lower bound for  $\mathcal{R}_n(\mathcal{C})$ : By definition, no estimator can achieve a better rate on  $\mathcal{C}$  than that of the lower bound. This bound gives also information on how close the risk of a given estimator is to the minimax risk. If the rate of the upper bound on the risk of an estimator matches the rate of the lower bound on the minimax risk on the class  $\mathcal{C}$ , then the estimator is said to have the minimax rate of convergence on this class.

Similarly to the minimax risk, we define the weighted minimax risk as follows. Let  $\mathcal{C}$  be a class of measurable subsets of  $\mathbb{R}^d$ , of positive volume. For an estimator  $\hat{G}_n$ , we define its weighted risk on the class  $\mathcal{C}$  as

$$\mathcal{Q}_n(\hat{G}_n; \mathcal{C}) = \sup_{G \in \mathcal{C}} \mathbb{E}_G \left[ \frac{|G \Delta \hat{G}_n|}{|G|} \right],$$

and the minimax weighted risk on the class  $\mathcal{C}$ , when  $n$  observations are available, is defined as

$$\mathcal{Q}_n(\mathcal{C}) = \inf_{\hat{G}_n} \mathcal{Q}_n(\hat{G}_n; \mathcal{C}).$$

Let us mention an alternative to this minimax setup. The minimax risk on a class  $\mathcal{C}$  is defined for each number  $n$  of available observations. The supremum is taken over all  $G \in \mathcal{C}$ , and therefore  $G$  is allowed to depend on  $n$ . In particular, usual methods for proving lower bounds on the minimax risk are based on hypotheses testing, and the hypotheses often depend on  $n$  in the proofs. This might seem unrealistic, since adding new observations should not change the unknown distribution measure from which they are generated. The alternative to this minimax setup consists in defining individual lower rates of convergence on a class  $\mathcal{C}$ . A sequence of positive numbers  $(a_n)$  is called an individual lower rate of convergence on the class  $\mathcal{C}$  if and only

if

$$\inf_{(\hat{G}_n)_{n \in \mathbb{N}}} \sup_{G \in \mathcal{C}} \limsup_{n \rightarrow \infty} \frac{\mathbb{E}_G \left[ |G \Delta \hat{G}_n| \right]}{a_n} > 0, \quad (1.6)$$

where the infimum is taken over all sequences  $(\hat{G}_n)_{n \in \mathbb{N}}$  of estimators. This alternative is proposed in [GKKW02]. It will not be much discussed in this thesis, in which we prefer to focus on the classical minimax setup. This preference will be motivated when we compare estimation and testing, in Section 3.5.

#### 1.2.4 Adaptation and misspecification

In our minimax setup, the target set  $G$  that we aim to estimate is seen as a member of a class  $\mathcal{C}$ , and the risk of estimation is defined uniformly on this class. This class can be very big. Assume it can be divided into several subclasses:

$$\mathcal{C} = \bigcup_{\tau \in \mathfrak{T}} \mathcal{C}_\tau,$$

where  $\mathfrak{T}$  is a set of parameters. This decomposition may be interesting, for example, if the subclasses  $\mathcal{C}_\tau$  are more simple than the whole class  $\mathcal{C}$ , i.e., much smaller. If so, the minimax rate on a subclass  $\mathcal{C}_\tau$  might be much smaller than that on the whole class. The target set  $G$  belongs to  $\mathcal{C}_\tau$  for some  $\tau \in \mathfrak{T}$ . If the value of  $\tau$  is known, a natural approach consists in constructing an estimator which *knows* the value of  $\tau$ , and in bounding from above its risk on the class  $\mathcal{C}_\tau$ . However, the value of  $\tau$  may be unknown. As a consequence, the estimation procedure cannot take into account the value of this parameter. Yet, we might be interested in achieving or, at least, approaching, the minimax rate  $\mathcal{R}_n(\mathcal{C}_\tau)$ , by constructing an estimator which does not depend on  $\tau$ . Such an estimator is said to be adaptive with respect to  $\tau$ , or to adapt to  $\tau$ . It would have exactly, or approximately, the minimax rate of convergence, simultaneously on all subclasses  $\mathcal{C}_\tau$ , for  $\tau$  ranging in  $\mathfrak{T}$ .

Let  $\hat{G}_n$  be an estimator, based on  $n$  random observations, of the target set  $G$  belonging to a class  $\mathcal{C} = \bigcup_{\tau \in \mathfrak{T}} \mathcal{C}_\tau$ . For  $\tau \in \mathfrak{T}$ , let  $\psi_{n,\tau}$  be the minimax rate of convergence on  $\mathcal{C}_\tau$ . The estimator  $\hat{G}_n$  is said to be adaptive minimax with respect to  $\tau$  if and only if

$$c_1 \leq \sup_{\tau \in \mathfrak{T}} \psi_{n,\tau}^{-1} \mathcal{R}_n(\hat{G}_n; \mathcal{C}_\tau) \leq c_2$$

for  $n$  large enough, where  $c_1$  and  $c_2$  are two positive constants. It is not always possible to construct an adaptive minimax estimator. In particular, not knowing in advance the value of



$\tau$  may unavoidably lead to a loss in the rate of convergence. Such loss often occurs when one aims to estimate a function at a given point, e.g. a density or a regression function, whose smoothness is characterized by the parameter  $\tau$  (see [BL92] for instance).

In this thesis, we will do adaptation when estimating a polytope in  $\mathcal{P}_r$  - or in  $\mathcal{P}_r^{(1)}$  - without knowing the true value of  $r$ , i.e., the number of vertices of the target polytope. In this case,  $\mathcal{C} = \mathcal{P} = \bigcup_{r \geq d+1} \mathcal{P}_r$  or  $\mathcal{C} = \mathcal{P}^{(1)} = \bigcup_{r \geq d+1} \mathcal{P}_r^{(1)}$ , and  $\mathfrak{T} = \{d+1, d+2, \dots\}$ .

**Misspecification** Let us now add a point, denoted by  $\infty$ , to the set  $\mathfrak{T}$ , and consider a new class  $\mathcal{C}_\infty$ , which is not contained in  $\mathcal{C}$ . Denote by  $\mathcal{C}' = \mathcal{C} \cup \mathcal{C}_\infty = \bigcup_{\tau \in \mathfrak{T} \cup \{\infty\}} \mathcal{C}_\tau$ . Assume now that the target to be estimated belongs to this bigger class  $\mathcal{C}'$ , and not necessarily to  $\mathcal{C}$ . This situation corresponds to possible misspecification of the model. Adaptive estimation with respect to  $\tau \in \mathfrak{T} \cup \{\infty\}$  allows one to bypass misspecification, if we are able to construct an estimator which adapts to  $\tau \in \mathfrak{T} \cup \{\infty\}$ , and which is minimax simultaneously on all subclasses  $\mathcal{C}_\tau, \tau \in \mathfrak{T}$ , and, on the class  $\mathcal{C}_\infty$ .

Typically, in this thesis,  $\mathfrak{T}$  is the set of all integers greater than the dimension  $d$ , and  $\mathcal{C}_\tau = \mathcal{P}_\tau$  or  $\mathcal{P}_\tau^{(1)}$ , while  $\mathcal{C}_\infty = \mathcal{K}_d$  or  $\mathcal{K}_d^{(1)}$ : The class  $\mathcal{C}_\infty$  contains convex bodies which are not polytopes. The class  $\mathcal{C}$  is  $\mathcal{K}_d$  itself, or  $\mathcal{K}_d^{(1)}$ . In this framework, misspecification consists in mistakenly believing that the target - a convex body - is a polytope. As we will see, there is a big gap between the minimax risks on the classes of polytopes with given number of vertices, and on the class of all convex bodies.

### 1.3 The statistical models

**The density support model (DS)** The density support model that we consider in this thesis consists of the observation of a sample of  $n$  i.i.d. random variables  $X_i, i = 1, \dots, n$ , with uniform distribution on some compact subset  $G$  of  $\mathbb{R}^d$ . In this setup,  $\mathbb{P}_G$  stands for the uniform probability measure on  $G$ .

**The regression model (RM)** Consider the following regression model:

$$Y_i = \mathbb{1}(X_i \in G) + \xi_i, i = 1, \dots, n,$$

where  $G$  is an unknown set in  $[0, 1]^d$ . The set of points  $X_i, i = 1, \dots, n$  is called the design, and it is observed. Unless we mention otherwise, the design is assumed to be i.i.d. uniformly distributed in the hypercube  $[0, 1]^d$ . The errors  $\xi_i$  are i.i.d. random variables, independent of

the design. In other words, some points in  $[0, 1]^d$  are labeled, with label 1 if the point belongs to some set  $G$  and 0 otherwise. Instead of observing directly the correct labels, one has access to their noisy version. From these observations, one wishes to recover  $G$  as accurately as possible. This model can be interpreted as a partial and noisy observation of an image,  $[0, 1]^d$ , in which some unknown object  $G$  is to be recovered. If there is no noise, i.e.,  $\xi_i = 0, i = 1, \dots, n$ , the observed image has only black and white pixels, the black ones belonging to  $G$ . If the signal is noisy, the image has different levels of gray. In this setup,  $\mathbb{P}_G$  stands for the probability measure associated to the random couple  $(X_1, Y_1)$  of (1.3).

We assume, along this thesis, that the errors  $\xi_i$  are i.i.d. and subgaussian, i.e., we make the following assumption.

**Assumption A.** *The random variables  $\xi_i, i = 1, \dots, n$ , are i.i.d. and satisfy the following exponential inequality.*

$$\mathbb{E}[e^{u\xi_i}] \leq e^{\frac{u^2\sigma^2}{2}}, \forall u \in \mathbb{R},$$

where  $\sigma$  is a positive number.

In particular, the  $\xi_i$ 's are necessarily centered. Note that i.i.d. zero-mean Gaussian random variables satisfy Assumption A.

In both setups (DS) and (RM), the set  $G$  is unknown and our purpose is to do statistical inference about this set. In Model (DS), we aim to estimate  $G$ . In Model (RM), we want either to estimate  $G$ , or to detect it, i.e., to be able to decide whether  $G$  is nonempty. We focus on two cases. First, we assume that the unknown set  $G$  is a polytope, and that it belongs to the class  $\mathcal{P}_r$ , for a given and known  $r$ . For each model we define an estimator which achieves the minimax - or nearly minimax, up to a logarithmic factor, in the case of Model (DS) - rate of convergence on the class  $\mathcal{P}_r^{(1)}$ , and another estimator which also achieves the minimax rate of convergence on the class  $\mathcal{K}_d^{(1)}$ . On the other hand, if  $G \in \mathcal{P}_r^{(1)}$  but the integer  $r$  is not known, we propose an adaptive method to estimate  $G$ , and we get the minimax - or, again, nearly minimax in Model (DS), up to a logarithmic factor - rate of convergence, simultaneously on all classes  $\mathcal{P}_r^{(1)}, r \geq d + 1$ . In addition, we prove that the same estimator achieves the minimax rate of convergence on the class  $\mathcal{K}_d^{(1)}$ . These results are presented in Chapter 2 for Model (DS), and Chapter 3 for Model (RM). In Chapter 2, we prove new additional probabilistic results on the proposed estimator. In Chapter 3, we also focus on the one-dimensional case in Model (RM). In this case, a convex body is simply a segment. We prove that our first estimator can be improved under an assumption either on the location of the unknown set  $G$ , or on its size. We also propose decision rules for testing whether  $G$  is nonempty, and give the asymptotic minimal

size of  $G$  so it can be detected by some decision rule. Chapter 4 is mostly devoted to a review of the literature on estimation of functionals of  $G$ , such as its volume, perimeter, or other intrinsic volumes.

## Chapter 2

# Estimation of the convex or polytopal support of a uniform density

Here, we consider the density support model (DS) defined in Section 1.3, and we focus on estimation of  $G$ . This problem is of interest in connection with the detection of abnormal behavior, see [DW80]. In image recovering, when an object is only partially observed, e.g. if only some pixels are available, one would like to recover the object as accurately as possible. In the case of polytopal support  $G$ , we will propose an estimator when the number of vertices is known, and an estimator which is adaptive to the number of vertices if that number is unknown. In the general case, a natural estimator is the convex hull of the available observations. We study its optimality in a minimax setup, and we analyze asymptotic properties of this random polytope, as the sample size tends to  $\infty$ .

The density of the observations is assumed to be uniform and only the support is unknown. This is a very strong assumption, and it may be questionable in some practical cases. However, this is an important setup, which already raises to many questions, both in probabilistic and statistical perspectives.

Assume that  $G$  is a convex body, and denote it by  $K$ . The convex hull of the sample is quite a natural estimator of  $K$ . The properties of this random subset of  $\mathbb{R}^d$  have been extensively studied since the early 1960's, from a geometric and probabilistic perspective. The very original question associated to this object was the famous Sylvester four-point problem: what is the probability that one of the four points chosen at random in the plane is inside the triangle formed by the three others ? We refer to [Bà01] for a historical survey and extensions of Sylvester problem.

Of course, this question is not well-posed, and one should specify the probability measure of those four points. The many answers that were proposed, in the 18th century, accompanied the birth of a new field: geometrical probability, and, later, in the twentieth century, stochastic geometry. In 1963 and 1964, Rényi and Sulanke [RS63, RS64] studied some basic properties of the convex hull of  $n$  i.i.d. random points in the plane ( $d = 2$ ), uniformly distributed in some convex body  $K$ . More specifically, if this convex hull is denoted by  $\hat{K}_n$ , its number of vertices by  $V_n$  and its missing area  $|K \setminus \hat{K}_n|$  by  $A_n$ , they investigated the asymptotics of the expectations  $\mathbb{E}[V_n]$  and  $\mathbb{E}[A_n]$ . Their results are highly dependent on the structure of the boundary of  $K$ . The expected number of vertices is of the order  $n^{1/3}$  when the boundary of  $K$  is smooth enough, and  $r \ln n$  when  $K$  is a convex polygon with  $r$  vertices,  $r \geq 3$ . The expected missing area is of the order  $n^{-2/3}$  in the first case and, if  $K$  is a square, it is of the order  $(\ln n)/n$ . May the square be arbitrarily large or small, only the constants and not the rates are affected by a scale factor. Rényi and Sulanke [RS63, RS64] provided asymptotic expansions of these expectations with the explicit constants up to two or three terms. In 1965, Efron [Efr65] showed a very simple equality which connects the expected value of the number of vertices  $V_{n+1}$  and that of the missing area  $A_n$ . Namely, one has

$$\mathbb{E}_K[A_n] = \frac{|K| \mathbb{E}_K[V_{n+1}]}{n+1}, \quad (2.1)$$

independently of the structure of the boundary of  $K$ . In particular, (2.1) allows one to extend the results of [RS63, RS64] about the missing area to any convex polygon with  $r$  vertices. If  $K = P$  is such a polygon,  $\mathbb{E}_P[A_n]$  is of the order  $r(\ln n)/n$ , up to a factor of the form  $c|P|_2$ , where  $c$  is positive and does not depend on  $r$  or  $P$ . Notably, Efron's identity (2.1) holds in any dimension  $d \geq 1$ . Thus, all results concerning the expectation of the missing volume of  $\hat{K}_n$  can be adapted for the expected number of vertices of  $\hat{K}_{n+1}$  and conversely. We give a proof of this identity in Section 2.5.2.

More recently, many efforts were made to extend Rényi and Sulanke's results to dimensions 3 and higher. We refer to [Wie78], [Gro74], [Dwy88] and the references therein. The most important results of the literature are presented in the next section.

## 2.1 Random polytopes

The random set  $\hat{K}_n$  is a random polytope, called the random convex hull. It is indeed a polytope, because it is the convex hull of a finite sample of points, and it is random since those points are

random. There are several other ways to construct random polytopes. For instance, by taking the intersection of randomly chosen half spaces, or by taking the convex hull of a Poisson point process. This last example is closely related to the model we are dealing with, and this relation has been used in the literature (e.g. [Par12]) to derive asymptotic results for  $\hat{K}_n$ . An extensive survey on the different kinds of random polytopes is made in [MCRF10]. Let us focus on the properties of  $\hat{K}_n$ .

### 2.1.1 The wet part and the floating convex body

In the late 1980's, B  r  ny and Larman [BL88] (see [BBSW07] for a review) proposed a generalization of the results on the asymptotic expected missing volume of  $\hat{K}_n$  that were known so far in particular cases. They made no assumption on the dimension  $d$ , on symmetry properties of  $K$  and on the structure of its boundary. They considered the  $\varepsilon$ -wet part of  $K$ , as defined by Dupin [Dup22] in fluid mechanics, and later by Blaschke [Bla23]. Let us call a cap of  $K$  the intersection of a closed halfspace of  $\mathbb{R}^d$  with  $K$ .

**Definition 2.1.** *Let  $K \in \mathcal{K}_d$  be a convex body and  $\varepsilon \in (0, 1)$ . The  $\varepsilon$ -wet part of  $K$ , denoted by  $K(\varepsilon)$ , is the union of all caps of  $K$  of volume less or equal to  $\varepsilon|K|$ . The  $\varepsilon$ -floating body of  $K$  is  $K \setminus K(\varepsilon)$ .*

To understand this definition and why it was introduced in fluid mechanics, let us set  $d = 2$ . One should imagine that  $\mathbb{R}^2$  is an ocean, with an iceberg in it. That iceberg is seen from above, and  $K$  is what is seen of the iceberg, i.e., its projection on the horizontal plane. The part of the iceberg inside the water is the wet part of  $K$ , and the floating part of the iceberg is the floating body of  $K$ .

B  r  ny and Larman [BL88] proved that if  $K$  is of volume one, then the expected missing volume of  $\hat{K}_n$ , i.e.,  $\mathbb{E}_K \left[ |K \setminus \hat{K}_n| \right]$ , is of the same order as the volume of the  $1/n$ -wet part.

**Theorem 2.1** ([BL88]). *Assume that  $K \in \mathcal{K}_d$  has volume 1. Then,*

$$c_1|K(1/n)| \leq \mathbb{E}_K \left[ |K \setminus \hat{K}_n| \right] \leq c_2(d)|K(1/n)|, \quad \forall n \geq n_0(d), \quad (2.2)$$

where  $c_1$  is a universal positive constant,  $c_2(d)$  is a positive constant which depends on  $d$  only, and  $n_0(d)$  is a positive integer which depends on  $d$  only.

Note that it is not necessary to assume that  $K$  is of volume one, and by rescaling, (2.2) still holds for all  $K \in \mathcal{K}_d$ .

As a consequence of (2.2), the problem of computing the expected missing volume of the random polytope  $\hat{K}_n$  becomes analytical, and studying its asymptotic properties reduces to

analyzing those of the wet part of  $K$ . The wet part has been studied extensively in convex analysis and geometry. We state the following result, taken from [BL88], which gives the smallest value of  $|K(\epsilon)|$ , for all possible  $K \in \mathcal{K}_d$  of volume 1, and shows that up to positive constants independent of  $\epsilon$ , this value is attained by polytopes.

**Theorem 2.2** ([BL88]). *Let  $K \in \mathcal{K}_d$ , such that  $|K| = 1$ , and  $\epsilon \in (0, 1)$ . There exists a positive constant  $c(d)$ , which depends on  $d$  only, such that*

$$c(d)\epsilon \left( \ln \left( \frac{1}{\epsilon} \right) \right)^{d-1} \leq |K(\epsilon)|.$$

*In addition, if  $K$  is a polytope, then*

$$|K(\epsilon)| \leq c(K)\epsilon \left( \ln \left( \frac{1}{\epsilon} \right) \right)^{d-1},$$

*for some positive constant  $c(K)$  which depends on  $K$ .*

The first term of the asymptotic expansion of the  $\epsilon$ -wet part of a convex body  $K$  in  $\mathbb{R}^d$ , when  $\epsilon \rightarrow 0$ , is given with precise constants when  $\partial K$  is smooth enough, in [SW90, Sch93b]. Note that the volume of the  $\epsilon$ -wet part of  $K$ , renormalized by the volume of  $K$ , is invariant under invertible affine transformations, that is, if  $T$  is such a transformation, and  $K' = TK$ ,

$$\frac{|K(\epsilon)|}{|K|} = \frac{|K'(\epsilon)|}{|K'|}.$$

The same property holds for the expected ratio  $\mathbb{E}_K \left[ \frac{|K \setminus \hat{K}_n|}{|K|} \right]$ . Indeed, if  $X_1, \dots, X_n$  are i.i.d. uniformly distributed in  $K$ , then the random points  $TX_1, \dots, TX_n$  are i.i.d., uniformly distributed in  $TK$ . In addition,

$$CH(TX_1, \dots, TX_n) = T(CH(X_1, \dots, X_n)).$$

Therefore,

$$\mathbb{E}_K \left[ \frac{|K \setminus \hat{K}_n|}{|K|} \right] = \mathbb{E}_{K'} \left[ \frac{|K' \setminus \hat{K}_n|}{|K'|} \right].$$

Hence, by taking  $T = |K|^{-1/d} I_d$ , where  $I_d : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is the identity function, so that  $TK$  has volume one, we see that it is sufficient to consider convex bodies of volume one.

The next section is devoted to the asymptotic properties of the expected missing volume of  $\hat{K}_n$ .

## 2.1.2 Convergence of the random convex hull

### Rate of convergence

Together with Theorem 2.1, Theorem 2.2 proves that the expected missing volume of  $\hat{K}_n$  converges to zero at a slower rate than  $(\ln n)^{d-1}/n$ , for all  $K \in \mathcal{K}_d$ , and this rate is achieved if  $K$  is a polytope.

Conversely, Groemer [Gro74] proved the following:

**Theorem 2.3** ([Gro74]). *Let  $n$  be a positive integer. Let  $K \in \mathcal{K}_d$ , and let  $B$  be a Euclidean ball in  $\mathbb{R}^d$  with the same volume as  $K$ . Then*

$$\mathbb{E}_K \left[ |K \setminus \hat{K}_n| \right] \leq \mathbb{E}_B \left[ |B \setminus \hat{K}_n| \right],$$

with equality if and only if  $K$  is an ellipsoid.

It is not hard to see (cf. [Wer06] for instance) that if  $B$  is a volume one Euclidean ball in  $\mathbb{R}^d$ , then

$$c_1(d)\epsilon^{2/(d+1)} \leq |B(\epsilon)| \leq c_2(d)\epsilon^{2/(d+1)}, \quad \forall \epsilon \in (0, 1/2), \quad (2.3)$$

for some positive constants  $c_1(d)$  and  $c_2(d)$  which depend on  $d$  only.

Therefore, using (2.2), Theorem 2.3 and (2.3), the expected missing volume of  $K$  converges to zero at a faster rate than  $n^{-2/(d+1)}$ , and this rate is achieved when  $K$  is an ellipsoid.

It is now clear that for  $n \geq n_0(d)$ ,

$$c_1(d) \frac{(\ln n)^{d-1}}{n} \leq \mathbb{E}_K \left[ \frac{|K \setminus \hat{K}_n|}{|K|} \right] \leq c_2(d) n^{\frac{-2}{d+1}}, \quad \forall K \in \mathcal{K}_d, \quad (2.4)$$

where  $n_0(d)$  is a positive integer which depends on  $d$  only, and  $c_1(d)$  and  $c_2(d)$  are positive constants which depend on  $d$  only. The lower bound is tight when  $K$  is a polytope, and the upper bound is tight when  $K$  is an ellipsoid.

We end this section with a remark. As we saw previously, it was shown by Groemer [Gro74] that if  $K \in \mathcal{K}_d$ , then

$$\mathbb{E}_K \left[ |K \setminus \hat{K}_n| \right] \leq \mathbb{E}_B \left[ |B \setminus \hat{K}_n| \right],$$

where  $B$  is a Euclidean ball in  $\mathbb{R}^d$  with the same volume as  $K$ . On the other hand, it was shown that the equality holds if and only if  $K$  is an ellipsoid. In other words, among all convex bodies of a given volume, the ellipsoids are the hardest to estimate by  $\hat{K}_n$ . We also saw that the rate of convergence of the convex hull  $\hat{K}_n$  is the best for polytopes. Is there such an inequality as



in Theorem 2.3, which specifies for which  $K \in \mathcal{K}_d$  the minimum value of the expected missing volume of  $\hat{K}_n$  is achieved? Barany and Buchta [BB93] partially answered this question. They proved that among all convex bodies of a given volume, the simplices are the easiest to estimate by  $\hat{K}_n$  in the following sense.

**Theorem 2.4** ([BB93]). *Assume  $d \geq 2$ . Let  $K \in \mathcal{K}_d$  and  $\Delta$  be a simplex in  $\mathbb{R}^d$  with the same volume as  $K$ . Then,*

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{E}_K \left[ |K \setminus \hat{K}_n| \right]}{\mathbb{E}_\Delta \left[ |\Delta \setminus \hat{K}_n| \right]} \geq 1 + \frac{1}{d+1},$$

*unless  $K$  is itself a simplex.*

We recall that a simplex is an (invertible) affine transformation of the regular simplex, defined as the convex hull of the origin and the canonical base vectors in  $\mathbb{R}^d$ . This result is asymptotic, so it is weaker than that of Groemer [Gro74]. The authors in [BB93] conjectured the following inequality, of the type of Groemer's result:

$$\mathbb{E}_K \left[ |K \setminus \hat{K}_n| \right] \geq \mathbb{E}_\Delta \left[ |\Delta \setminus \hat{K}_n| \right], \quad \forall n \geq 1,$$

as soon as  $K \in \mathcal{K}_d$  and  $\Delta$  is a simplex with the same volume as  $K$ . Dimension  $d$  would be any positive integer. To our knowledge, this stronger result has not been proved yet.

### Exact rates of convergence

The first term of the asymptotic expansion of the expected missing volume of  $\hat{K}_n$ , in a convex body  $K \subseteq \mathbb{R}^d$ , is known when  $\partial K$  is smooth enough [Bà92] or under weaker assumptions [Bà92], and when  $K$  is a polytope [BB93]. The two following theorems can be found in the references above.

**Theorem 2.5** ([Sch94]). *Let  $K \in \mathcal{K}_d$  be a convex body. Denote by  $\kappa(x)$  the generalized Gauss-Kronecker curvature of  $\partial K$  at point  $x$ , and by  $\mu$  the Lebesgue measure on  $\partial K$ . Then,*

$$c(d) \lim_{n \rightarrow \infty} \frac{\mathbb{E}_K \left[ |K \setminus \hat{K}_n| \right]}{\left( \frac{|K|}{n} \right)^{\frac{2}{d+1}}} = \int_{\partial K} \kappa(x)^{\frac{1}{d+1}} d\mu(x),$$

where

$$c(d) = 2 \left( \frac{\beta_{d-1}}{d+1} \right)^{\frac{2}{d+1}} \frac{(d+3)(d+1)!}{(d^2+d+2)(d^2+1)\Gamma\left(\frac{d^2+1}{d+1}\right)}.$$

In this theorem,  $\Gamma$  denotes Euler's Gamma function, defined as  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \forall x > 0$ . For a definition of the Gauss-Kronecker curvature, we refer to [Sch93a, Section 2.5, p. 106].

The notion of generalized Gauss-Kronecker curvature is introduced in [Sch94]. If  $K = rB_2^d$  is a Euclidean ball of radius  $r > 0$ , then  $\kappa(x) = r^{-(d-1)}, \forall x \in \partial K$ , and the last theorem implies that the expected missing volume of  $\hat{K}_n$  is asymptotically equivalent to  $c'(d)r^d n^{-2/(d+1)}$ , i.e., a positive constant which depends on  $d$ , multiplied by the volume of  $K$ , times the rate  $n^{-2/(d+1)}$ . Theorem 2.5 shows that the rate  $n^{-2/(d+1)}$  is attained by  $\hat{K}_n$  not only when  $K$  is an ellipsoid, but also as soon as the set of points  $x \in \partial K$  at which the generalized Gauss-Kronecker curvature is positive, is of positive measure with respect to  $\mu$ .

The next theorem concerns convex polytopes. For a polytope  $P \in \mathcal{P}$ , a tower of  $P$  is any increasing sequence  $F_0 \subseteq F_1 \subseteq \dots \subseteq F_{d-1}$ , where  $F_k$  is a  $k$ -face of  $P$ , for  $k = 0, \dots, d-1$ . Denote by  $T(P)$  the number of such towers of  $P$ .

**Theorem 2.6** ([BB93]). *Let  $d \geq 2$  and  $P \in \mathcal{P}$ . Then,*

$$\lim_{n \rightarrow \infty} \frac{n}{(\ln n)^{d-1}} \mathbb{E}_P[|P \setminus \hat{K}_n|] = \frac{|P|T(P)}{(d+1)^{d-1}(d-1)!}.$$

The positive constant appearing in the limit in the last theorem is the volume of  $P$  multiplied by a constant, which depends on  $P$  through its facial structure only, i.e., through its boundary.

Both theorems 2.5 and 2.6 suggest that the expected missing volume of  $\hat{K}_n$ , normalized by the volume  $K$ , depends, asymptotically, only on the facial structure of  $K$ , i.e., on the algebraic structure of its boundary.

We have shown that the best rate of convergence of the expected missing volume of  $\hat{K}_n$  is achieved when  $K$  is a polytope, and this rate is of the order  $(\ln n)^{d-1}/n$ . This remark motivates our focus on polytopes. On the other hand, polytopes are very important objects in convex geometry and convex analysis. They are the most simple convex bodies. They can be described by a finite number of inequalities. Equivalently, they are defined by the knowledge of a finite number of parameters - the coordinates of their vertices. In particular, polytopes can be used to approximate convex bodies. We refer to [Gru93] for results on best approximation of convex bodies by polytopes, and to [Rei03] for an interesting comparison of best polytopal approximation of convex bodies and their approximation by random polytopes.

In the next section, we present the main result on the asymptotic structure of the random polytope  $\hat{K}_n$ , due to B  r  ny and Buchta [BB93].

### 2.1.3 Asymptotic facial structure

As we mentioned above, the asymptotics for expectation of the missing volume of  $\hat{K}_n$  is well known. Thus, thanks to Efron's identity (2.1), it is easy to get the asymptotics for expectation of the number of vertices of  $\hat{K}_n$ . This object,  $\hat{K}_n$ , has been of much interest in the probabilistic and geometric literature since the leading works of Rényi and Sulanke [RS63, RS64], which were an important step in stochastic geometry. The volume, the perimeter, and the number of vertices of  $\hat{K}_n$ , in dimension  $d = 2$ , were investigated. Naturally, the results were extended to higher dimensions. In particular, in dimension  $d \geq 3$ , the facial structure of  $\hat{K}_n$  becomes more intricate. An easy fact is that every  $k$ -face of  $\hat{K}_n$ , for  $k = 1, \dots, d-1$ , has exactly  $k+1$  vertices,  $\mathbb{P}_K$ -almost surely. Indeed, a vertex of any  $k$ -face is necessarily a vertex of  $\hat{K}_n$ , and therefore it belongs to the sample  $\{X_1, \dots, X_n\}$ . Therefore,  $\mathbb{P}_K$ -almost surely, no more than  $k+1$  of the points  $X_1, \dots, X_n$  can belong to the same affine subspace of dimension  $k$ . Bárány and Buchta, in [BB93], investigated, in Model (DS), the expected number of  $k$ -faces of  $\hat{K}_n$ ,  $1 \leq k \leq d-1$ , when  $G = P$  is a polytope. They showed that this expected number is of the same order for all values of  $k = 0, \dots, d-1$ :

**Theorem 2.7** ([BB93]). *Let  $P \in \mathcal{P}$  and  $0 \leq k \leq d-1$ . For all  $n \geq 1$ ,*

$$\mathbb{E}_P[f_k(\hat{K}_n)] = c(d, k)T(P)(\ln n)^{d-1} + \epsilon_n,$$

where  $c(d, k)$  is a positive constant which depends on  $d$  and  $k$ , and the residual term  $\epsilon_n$  satisfies  $\epsilon_n \leq c_0(d, k, P)(\ln n)^{d-2} \ln \ln n$ , for some positive constant  $c_0(d, k, P)$  which depends only on  $d$ ,  $k$  and  $P$ .

It is clear from this theorem that  $f_k(\hat{K}_n)$  is not a consistent estimator of  $f_k(P)$ . In Section 2.2.3, we will propose an estimator  $\hat{r}$  of  $f_0(P)$ . Again, this estimator will not be shown to be consistent. However, it will allow us to select among a family of preliminary estimators of  $P \in \mathcal{P}$  one which achieves a reasonable rate of convergence in the case of unknown value of  $f_0(P)$ . It is actually not possible to estimate  $f_0(P)$  consistently and uniformly on the class  $\mathcal{P}$ , or even on a smaller class  $\mathcal{P}_r$  or  $\mathcal{P}_r^{(1)}$ . On these two classes,  $f_0(P) \leq r$ . To make the ideas clear, assume that  $d = 2$ . A vertex may be undetectable if its angle is close to 180 degrees. An open question would be whether it is possible to estimate, say for  $d = 2$ , the number of vertices of  $P \in \mathcal{P}$ , provided that at each of its vertices, the angle is less than  $180 - \alpha$ , where  $\alpha > 0$  is given and known. Statistical inference on the facial structure of polytopes has not been investigated yet.

## 2.2 Minimax estimation

All the works discussed in the previous sections were developed in a geometric and probabilistic prospective. No efforts were made at this stage to understand whether the  $\hat{K}_n$  is optimal if seen as an estimator of the unknown support. Only in the 1990's, this question was invoked in the statistical literature. Mammen and Tsybakov [MT95] showed that under some restrictions on the volume of  $K$ , the convex hull is optimal in a minimax sense (see the next section for details). Korostelev and Tsybakov [KT93b] give a detailed account of the topic of set estimation. Cuevas and Rodriguez-Cazal [CRC04], and Pateiro Lopez [PL08], studied the properties of set estimators of the support of a density under several geometric assumptions on the boundary of the unknown set. See also [Cue09], [CF10], [CRC04], [Gun12], for an overview of recent developments about estimation of the support of a probability measure.

Let us now look at the random polytope  $\hat{K}_n$  as a set estimator, in the minimax setup. The right side of (2.4) shows that  $n^{-2/(d+1)}$  is an upper bound of the minimax weighted risk of  $\hat{K}_n$  on the class  $\mathcal{K}_d$ .

Hence, this provides both an upper bound for the minimax weighted risk  $\mathcal{Q}_n(\mathcal{K}_d)$  on  $\mathcal{K}_d$ , and an upper bound for the minimax risk  $\mathcal{R}_n(\mathcal{K}_d^{(1)})$  on  $\mathcal{K}_d^{(1)}$  (see Section 1.2.3 for the definition of the risks):

**Theorem 2.8** ([Bru14a]). *There exists a positive integer  $n_0(d)$ , which depends on  $d$  only, such that for all  $n \geq n_0(d)$*

$$\mathcal{Q}_n(\mathcal{K}_d) \leq c_1(d)n^{-\frac{2}{d+1}},$$

and

$$\mathcal{R}_n(\mathcal{K}_d^{(1)}) \leq c_1(d)n^{-\frac{2}{d+1}},$$

where  $c_1(d)$  is a positive constant which depends on  $d$  only.

A natural question is whether these upper bounds are sharp, i.e., if the minimax weighted risk on  $\mathcal{K}_d$  and the minimax risk on  $\mathcal{K}_d$  are of this order. This question will be answered positively in Section 2.2.2. In addition, Theorem 2.6 suggests that the minimax weighted risk on the class of polytopes with uniformly bounded number of towers, is at most of the order  $(\ln n)^{d-1}/n$ . Theorem 2.2 shows that the convex hull estimator cannot perform better, in terms of rate of convergence, on that class. In Section 2.2.1, we will show that the convex hull estimator does not have the minimax rate of convergence on the classes  $\mathcal{P}_r, r \geq d+1$ , and we will construct a new random polytope, which nearly attains the minimax rate of convergence on the classes  $\mathcal{P}_r^{(1)}, r \geq d+1$ , up to a single logarithmic factor.

### 2.2.1 Estimation of polytopes

#### Upper bound

The case of polytopes has not been investigated from the statistical prospective. In Section 2.1.1, we saw that if  $K$  is a polytope, denoted by  $P$ , then the expected missing volume of the convex hull estimator  $\hat{K}_n$  is of the order  $(\ln n)^{d-1}/n$ . However, the constant, in the upper bound, depends on  $P$ , not only through its volume, but also through its facial structure:

$$\mathbb{E}_P[|P \setminus \hat{K}_n|] \leq c(d, P) \frac{(\ln n)^{d-1}}{n}, \quad \forall n \geq n_0(d). \quad (2.5)$$

It seems quite clear that  $\mathbb{E}_P \left[ \frac{|P \setminus \hat{K}_n|}{|P|} \right]$  cannot be uniformly bounded from above by a positive constant times  $(\ln n)^{d-1}/n$  on the class  $\mathcal{P}$  of polytopes. The reason is that, since every convex body can be arbitrarily well approximated by a polytope, it seems quite reasonable to think that  $\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[ \frac{|P \setminus \hat{K}_n|}{|P|} \right] = \sup_{K \in \mathcal{K}_d} \mathbb{E}_K \left[ \frac{|K \setminus \hat{K}_n|}{|K|} \right]$ , and Corollary 2.3 would contradict a uniform upper bound of  $\mathcal{Q}_n(\mathcal{P})$  of the order  $(\ln n)^{d-1}/n$ .

Instead of considering the whole class  $\mathcal{P}$ , let us restrict ourselves to  $\mathcal{P}_r$ , for a given  $r \geq d+1$ . Intuitively, the convex hull estimator can be improved, and the logarithmic factor can be, at least partially, dropped. Indeed, a polytope with a given and known number of vertices is completely determined by the coordinates of its vertices, and therefore belongs to some parametric family. Hence, the rate may be expected to be  $1/n$ , instead of  $(\ln n)^{d-1}/n$ .

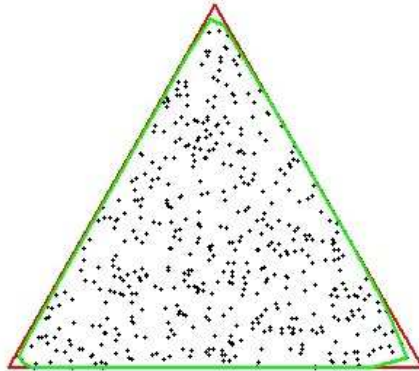


Figure 2.1: Convex hull of  $n = 500$  random points in a triangle

Figure 2.1 illustrates why there is a logarithmic factor in the risk of  $\hat{K}_n$ . Let us set the dimension  $d$  to 2, for simplicity's sake. Many of the points are expected to be close to the edges of  $P$  - a triangle, in Figure 2.1 -, at a distance of the order  $1/n$ , as for the minimum

of a sample of i.i.d. random variables, uniformly distributed on  $[0, 1]$ , expected to be of the order  $1/n$ . However, close to the vertices of  $P$ , there is less room for the  $X_i$ 's, and one can see on Figure 2.1 that  $\hat{K}_n$  underestimates  $P$  around the vertices, although it is locally much more accurate near the edges, far from the vertices.

Let  $P \in \mathcal{P}_r^{(1)}$ . The estimator  $\hat{K}_n$  is the convex hull of the random sample. It is a convex polytope, and using (2.4) and Efron's identity (2.1), its expected number of vertices is greater than  $(\ln n)^{d-1}$ , and thus much greater than  $r$ . On Figure 2.1, one can see that  $\hat{K}_n$  has, indeed, more than 3 vertices, and that they are much concentrated around the vertices of  $P$ . Instead of taking  $\hat{K}_n$ , which does not use take into consideration that the number of vertices of  $P$  is known, let us define  $\hat{P}_n^{(r)}$  as one of the polytopes in  $\mathcal{P}_r^{(1)}$  of smallest volume, which contains all the sample points:

$$\hat{P}_n^{(r)} \in \underset{P \in \mathcal{P}_r^{(1)}, X_i \in P, i=1, \dots, n}{\operatorname{argmin}} |P|. \quad (2.6)$$

The existence of such a polytope is ensured by compactness arguments. Note that  $\hat{P}_n^{(r)}$  needs not be unique. The next theorem establishes an exponential deviation inequality for the estimator  $\hat{P}_n^{(r)}$ .

**Theorem 2.9** ([Bru14a]). *Let  $r \geq d + 1$  be an integer, and  $n \geq 2$ . Then,*

$$\sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{P}_P \left[ n \left( |\hat{P}_n^{(r)} \triangle P| - \frac{4dr \ln n}{n} \right) \geq x \right] \leq c(d) e^{-x/2}, \quad \forall x > 0,$$

for some positive constant  $c(d)$ , which depends on  $d$  only.

We restrict ourselves to  $\mathcal{P}_r^{(1)}$  instead of the whole class  $\mathcal{P}_r$  for technical reasons only. However, we believe that such an inequality as that of Theorem 2.9 could be achieved, by replacing  $\mathcal{P}_r^{(1)}$  with  $\mathcal{P}_r$ , and  $|\hat{P}_n^{(r)} \triangle P|$  with  $|\hat{P}_n^{(r)} \triangle P|/|P|$ .

From the deviation inequality of Theorem 2.9 one can easily derive that the risk of the estimator  $\hat{P}_n^{(r)}$  on the class  $\mathcal{P}_r^{(1)}$  is of the order  $\frac{\ln n}{n}$ . Indeed, we have the next corollary.

**Corollary 2.1.** *Let the assumptions of Theorem 2.9 be satisfied. Then, for any positive number  $q$ , there exists a constant  $c(d, q)$ , which depends on  $d$  and  $q$  only, such that,*

$$\sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{E}_P \left[ |\hat{P}_n^{(r)} \triangle P|^q \right] \leq c(d, q) \left( \frac{r \ln n}{n} \right)^q.$$

Corollary 2.1 shows that the risk  $\mathcal{R}_n(\hat{P}_n^{(r)}; \mathcal{P}_r^{(1)})$  of the estimator  $\hat{P}_n^{(r)}$  on the class  $\mathcal{P}_r^{(1)}$  is bounded from above by  $\frac{r \ln n}{n}$ , up to some positive constant which depends on  $d$  only. Therefore

we have the following upper bound for the minimax risk on the class  $\mathcal{P}_r^{(1)}$ :

$$\mathcal{R}_n(\mathcal{P}_r^{(1)}) \leq c(d) \frac{r \ln n}{n}, \forall n \geq 2. \quad (2.7)$$

### Lower bound

It is now natural to ask whether the rate  $\frac{\ln n}{n}$  is minimax, i.e. whether it is possible to find a lower bound for  $\mathcal{R}_n(\mathcal{P}_r^{(1)})$  which converges to zero at the rate  $\frac{\ln n}{n}$ , or the logarithmic factor should be dropped.

In the 2-dimensional case, we provide a lower bound of the order  $1/n$ , with a factor which depends linearly on the number of vertices  $r$ . Namely, the following theorem holds.

**Theorem 2.10** ([Bru14a]). *Let  $r \geq 9$  be an integer, and  $n \geq r$ . Assume  $d = 2$ . Then,*

$$\mathcal{R}_n(\mathcal{P}_r^{(1)}) \geq \frac{cr}{n},$$

for some positive universal constant  $c$ .

Combined with (2.7), this bound shows that, as a function of  $r$ ,  $\mathcal{R}_n(\mathcal{P}_r^{(1)})$  behaves linearly in  $r$  in dimension two. In greater dimensions, it is quite easy to show that  $\mathcal{R}_n(\mathcal{P}_r^{(1)}) \geq \frac{c(d)}{n}$ , for some positive constant  $c(d)$ , which depends on  $d$  only. But this lower bound does not show the dependency on  $r$ . However, the upper bound (2.7) shows that  $\mathcal{R}_n(\mathcal{P}_r^{(1)})$  is at most linear in  $r$ .

### The logarithmic factor in the upper bound

We conjecture that the logarithmic factor can be removed in the upper bound of  $\mathcal{R}_n(\mathcal{P}_r^{(1)})$ ,  $r \geq d + 1$ . In fact, for the class  $\mathcal{P}_r$ , we conjecture that, for the weighted risk,

$$\mathcal{Q}_n(\mathcal{P}_r) \leq \frac{c(d, r)}{n}, \forall n \geq 2,$$

where  $c(d, r)$  is a positive constant which depends on  $d$  and  $r$ . Our conjecture is motivated by Efron's identity (2.1). It turns out that we can follow almost all the proof of this identity when the underlying set is a polytope. Instead the estimator  $\hat{P}_n^{(r)}$ , defined in Section 2.2.1, consider

$$\tilde{P}_n^{(r)} \in \operatorname{argmin}_{P \in \mathcal{P}_r, X_i \in P, i=1, \dots, n} |P|.$$

The difference with  $\hat{P}_n^{(r)}$  is that it is not constrained to be included in  $[0, 1]^d$ .

Let  $r \geq d + 1$  be an integer and  $P \in \mathcal{P}_r$ . Let  $\tilde{P}_n^{(r)}$  be the estimator defined above. In this section, we denote this estimator simply by  $\tilde{P}_n$ : we omit the dependence on  $r$ , for simplicity's

sake. Note that this estimator does not satisfy the nice property  $\tilde{P}_n \subseteq P$ , unlike the convex hull. However, by construction,  $|\tilde{P}_n| \leq |P|$ , so  $|P \triangle \tilde{P}_n| \leq 2|P \setminus \tilde{P}_n|$ , and we have:

$$\begin{aligned} \mathbb{E}_P^{\otimes n}[|\tilde{P}_n \triangle P|] &\leq 2\mathbb{E}_P^{\otimes n}[|P \setminus \tilde{P}_n|] \\ &= 2|P|\mathbb{E}_P^{\otimes n}\left[\frac{1}{|P|}\int_P I(x \notin \tilde{P}_n)dx\right] \\ &= 2|P|\mathbb{E}_P^{\otimes n}\left[\mathbb{P}_P[X \notin \tilde{P}_n|X_1, \dots, X_n]\right], \end{aligned} \quad (2.8)$$

where  $X$  is a random variable of the same distribution as  $X_1$ , and independent of the sample  $X_1, \dots, X_n$ , and  $\mathbb{P}_P[\cdot|X_1, \dots, X_n]$  denotes the conditional distribution of  $X$  given  $X_1, \dots, X_n$ . We set  $X_{n+1} = X$ , and we consider the bigger sample  $X_1, \dots, X_{n+1}$ . For  $i = 1, \dots, n+1$ , we denote by  $\tilde{P}^{-i}$  the same estimator as  $\tilde{P}_n$ , but this time based on the sample  $X_1, \dots, X_{n+1}$  from which the  $i$ -th variable  $X_i$  is withdrawn. In other words,  $\tilde{P}^{-i}$  is a convex polytope with at most  $r$  vertices, which contains the whole sample  $X_1, \dots, X_{n+1}$  but maybe the  $i$ -th variable, of minimum volume. Then,  $\tilde{P}_n = \tilde{P}^{-(n+1)}$ , and by continuing (2.8),

$$\begin{aligned} \mathbb{E}_P^{\otimes n}[|\tilde{P}_n \triangle P|] &\leq 2|P|\mathbb{P}_P^{\otimes n+1}[X_{n+1} \notin \tilde{P}^{-(n+1)}] \\ &= \frac{2|P|}{n+1} \sum_{i=1}^{n+1} \mathbb{P}_P^{\otimes n+1}[X_i \notin \tilde{P}^{-(i)}] \\ &= \frac{2|P|}{n+1} \mathbb{E}_P^{\otimes n+1} \left[ \sum_{i=1}^{n+1} I(X_i \notin \tilde{P}^{-(i)}) \right] \\ &= \frac{2|P|\mathbb{E}_P^{\otimes n+1}[V'_{n+1}]}{n+1}, \end{aligned} \quad (2.9)$$

where  $V'_{n+1}$  stands for the number of points  $X_i$  falling outside  $\tilde{P}^{-i}$ ,  $i = 1, \dots, n+1$ . Note that in this description we assume the uniqueness of such a polytope, which we conjecture to hold  $\mathbb{P}_P$ -almost surely, as long as  $n$  is large enough. It is not clear that if a point  $X_i$  is not in  $\tilde{P}^{-i}$ , then  $X_i$  lies on the boundary of  $\tilde{P}_{n+1}$ . However, if this were true, then  $\mathbb{P}_P$ -almost surely  $V'_{n+1}$  would be less or equal to  $d+1$  times the number of facets of  $\tilde{P}_{n+1}$ . Indeed, a facet is supported by an affine hyperplane of  $\mathbb{R}^d$ , which, with probability one, cannot contain more than  $d+1$  points of the sample at a time. Besides, the maximal number of facets of a  $d$ -dimensional convex polytope with at most  $r$  vertices is bounded by McMullen's upper bound (1.5), and we could have our conjecture proved:

$$\mathbb{E}_P^{\otimes n} \left[ \frac{|\tilde{P}_n \triangle P|}{|P|} \right] \leq \frac{cf_{d-1}(r, d)}{n},$$

for some positive constant  $c$ . However, there might be some cases when some points  $X_i$  are



not in  $\tilde{P}^{-i}$ , though they do not lay on the boundary of  $\tilde{P}_{n+1}$ . So it may be of interest to work directly on the variable  $V'_{n+1}$ . This remains an open problem.

### 2.2.2 Estimation of convex bodies

In Theorem 2.8, we gave an upper bound of the minimax risk  $\mathcal{R}_n(\mathcal{K}_d^{(1)})$ , and on the minimax weighted risk  $\mathcal{Q}_n(\mathcal{K}_d)$ . These two upper bounds, of the order  $n^{-2/(d+1)}$ , were obtained using the convex hull estimator  $\hat{K}_n$ , whose both risk on the class  $\mathcal{K}_d^{(1)}$ , and weighted risk on the class  $\mathcal{K}_d$ , were of that order. In this section, we show that  $n^{-2/(d+1)}$  is exactly the rate of both  $\mathcal{R}_n(\mathcal{K}_d^{(1)})$  and  $\mathcal{Q}_n(\mathcal{K}_d)$ .

**Theorem 2.11** ([Bru14a]). *There exists a positive integer  $n_0(d)$ , which depends on  $d$ , such that the following statements hold.*

- *The minimax risk on the class  $\mathcal{K}_d^{(1)}$  satisfies*

$$c_1(d)n^{-\frac{2}{d+1}} \leq \mathcal{R}_n(\mathcal{K}_d^{(1)}) \leq c_2(d)n^{-\frac{2}{d+1}}, \quad \forall n \geq n_0(d),$$

*for some positive constants  $c_1(d)$  and  $c_2(d)$  which depend on  $d$  only. In addition, the convex hull estimator has the minimax rate of convergence on  $\mathcal{K}_d^{(1)}$ .*

- *The minimax weighted risk on the class  $\mathcal{K}_d$  satisfies*

$$c_3(d)n^{-\frac{2}{d+1}} \leq \mathcal{Q}_n(\mathcal{K}_d) \leq c_4(d)n^{-\frac{2}{d+1}}, \quad \forall n \geq n_0(d),$$

*for some positive constants  $c_3(d)$  and  $c_4(d)$  which depend on  $d$  only - with  $c_3(d) = c_1(d)$  - . In addition, the convex hull estimator has the minimax rate of convergence on  $\mathcal{K}_d$ , with respect to the minimax weighted risk.*

The proof of the lower bound for the minimax risk  $\mathcal{R}_n(\mathcal{K}_d^{(1)})$  can be found in [MT95]. For the minimax weighted risk, it is quite clear that  $\mathcal{R}_n(\mathcal{K}_d^{(1)}) \leq \mathcal{Q}_n(\mathcal{K}_d)$ , so the same lower bound still holds.

### 2.2.3 Adaptation and misspecification

In the previous sections, we proposed estimators which highly depend on the structure of the boundary of the unknown support. In particular, when the support was supposed to be polytopal with at most  $r$  vertices, for some known integer  $r$ , our estimator was also, by construction, a polytope with at most  $r$  vertices. Now we will construct an estimator which does not depend on any other knowledge than the convexity of the unknown support, and the fact that it is

included in  $[0, 1]^d$ . This estimator will achieve the same rate as the estimators of Section 2.2.1 in the polytopal case, that is,  $r \ln n / n$ , where  $r$  is the unknown number of vertices of the support, and the same rate as the convex hull estimator  $\hat{K}_n$ , in the case where the support is not polytopal, or is polytopal but with a large number of vertices. Note that if the unknown support is a polytope with  $r$  vertices, with  $r$  larger than  $(\ln n)^{-1} n^{\frac{d-1}{d+1}}$ , then the rate of convergence of the risk of  $\hat{K}_n$ , not larger than  $n^{-2/(d+1)}$ , is smaller than the rate  $r(\ln n)/n$  of the upper bound of the risk of  $\hat{P}_n^{(r)}$ . The idea which we develop here is inspired by Lepski's method [Lep91]. The classes  $\mathcal{P}_r^{(1)}$ ,  $r \geq d+1$ , are nested, that is,  $\mathcal{P}_r^{(1)} \subseteq \mathcal{P}_{r'}^{(1)}$ , for  $r \leq r'$ . So if the true value of  $r$  is unknown, it seems better to overestimate, than to underestimate it. Intuitively, it makes sense to fit some polytope with more vertices to  $P$ , while the opposite may be impossible (e.g. on any triangle, it is possible to fit well some quadrilateral, but on a square, it is impossible to fit well a triangle). We use this idea in order to select an estimator among the preliminary estimators  $\hat{P}_n^{(r)}$ ,  $r \geq d+1$ , and  $\hat{K}_n$ .

Set  $R_n = \lfloor n^{(d-1)/(d+1)} / (\ln n) \rfloor$ . For  $r = d+1, \dots, R_n - 1$ , denote by  $\hat{Q}_n^{(r)} = \hat{P}_n^{(r)}$  and define  $\hat{Q}_n^{(R_n)} = \hat{K}_n$ . Let  $C = \beta_d + \max(2d\beta_d, c_4)$ , where  $c_4$  is the constant appearing in (2.10) and define

$$\hat{r} = \min \left\{ r \in \{d+1, \dots, R_n\} : |\hat{Q}_n^{(r)} \triangle \hat{Q}_n^{(r')}| \leq \frac{2Cr' \ln n}{n}, \forall r' = r, \dots, R_n \right\}.$$

The integer  $\hat{r}$  is well defined ; indeed, the set in the brackets in the last display is not empty, since the inequality is satisfied for  $r = R_n$ . Note that the definition of  $\hat{r}$  requires the knowledge of the constant  $C$ , which can be made explicit by an investigation of the proofs, and depend on  $d$  only.

The adaptive estimator is defined as  $\hat{P}_n^{adapt} = \hat{Q}_n^{(\hat{r})}$ . Then, if we denote by  $\mathcal{P}_\infty = \mathcal{K}_1$ , we have the following theorem.

**Theorem 2.12.** *Let  $n \geq 2$ . Let  $\phi_{n,r} = \min\left(\frac{r \ln n}{n}, n^{-\frac{2}{d+1}}\right)$ , for all integers  $r \geq d+1$  and  $r = \infty$ . There exists a positive constant  $c(d)$ , which depends on  $d$  only, such that*

$$\sup_{d+1 \leq r \leq \infty} \sup_{P \in \mathcal{P}_r} \mathbb{E}_P \left[ \phi_{n,r}^{-1} |\hat{P}_n^{adapt} \triangle P| \right] \leq c(d),$$

Thus, we show that one and the same estimator  $\hat{P}_n^{adapt}$  attains the optimal rate, up to a logarithmic factor, simultaneously on all the classes  $\mathcal{P}_r$ ,  $r \geq d+1$ , and on the class  $\mathcal{K}_1$  of all convex bodies in  $[0, 1]^d$ . In particular, the adaptive estimator deals well with misspecification. If the model is misspecified, e.g., the true support is believed to be polytopal, although it is

not, the adaptive estimator  $\hat{P}_n^{adapt}$  will perform as well as  $\hat{K}_n$ , as far as the rate of convergence is concerned.

The proof of this theorem is similar to that of Theorem 3.6, in Section 3.6. It requires a deviation inequality for  $\hat{K}_n$ , which will be given (2.10), see Section 2.3.1.

## 2.3 Further results and open problems

### 2.3.1 A universal deviation inequality for the random convex hull

In this section, we are interested in deviation inequalities for the missing volume of  $\hat{K}_n$ , i.e., in bounding from above the probability

$$\mathbb{P}_K[|K \setminus \hat{K}_n| > \epsilon |K|],$$

for  $K \in \mathcal{K}_d$  and  $\epsilon > 0$ . This would yield, as a consequence, upper bounds for the moments  $\mathbb{E}_K[|K \setminus \hat{K}_n|^q]$ ,  $q > 0$ . In order to obtain a deviation inequality, we first prove that it is sufficient to obtain a deviation inequality for  $K \in \mathcal{K}_d^{(1)}$ , by a scaling argument. Then, we use the metric entropy of the class  $\mathcal{K}_d^{(1)}$ . The deviation inequality that we prove is uniform on the class  $\mathcal{K}_d$ , hence it is of much interest in a statistical framework, in the minimax setup.

As we saw in Section 2.1.2, many - asymptotic or not - properties of the expected missing volume are now very well-understood. Much less is known about its higher moments and deviation probabilities. Using a jackknife inequality for symmetric functions of  $n$  random variables, Reitzner [Rei03] proved that if  $K$  is a  $d$ -dimensional convex body with smooth boundary, the variance of the missing volume is bounded from above by a positive constant times  $n^{-(d+3)/(d+1)}$ , and he conjectured that this is the actual order of magnitude for the variance. In addition, he proved that the second moment of the missing volume is exactly of the order  $n^{-4/(d+1)}$ , with explicit constants in terms of the affine surface area of  $K$ . Vu [Vu05] obtained deviation inequalities for general convex bodies of volume one, involving quantities such as the volume of the wet part, and derived precise deviation inequalities in the cases when  $K$  is a polytope, and when it has a smooth boundary. These inequalities involve constants which depend on  $K$  in a non explicit way. The main tools are martingale inequalities. As a consequence, upper bounds on the moments of the missing volume are proved, again with implicit constants depending on  $K$ . If  $K$  has a smooth boundary and is of volume one, Vu [Vu05] showed the existence of positive constants  $c$  and  $\alpha$ , which depend on  $K$ , such that for any  $\lambda \in \left(0, (\alpha/4)n^{-\frac{(d-1)(d+3)}{(d+1)(3d+5)}}\right]$ ,

the following holds:

$$\mathbb{P} \left[ \left| |\hat{K}_n| - \mathbb{E}[|\hat{K}_n|] \right| \geq \sqrt{\alpha \lambda n^{-\frac{d+3}{d+1}}} \right] \leq 2 \exp(-\lambda/4) + \exp \left( -c n^{\frac{d-1}{3d+5}} \right).$$

This inequality leads to upper bounds on the variance and on the  $q$ -th moment of the missing volume, respectively of orders  $n^{-(d+3)/(d+1)}$  and  $n^{-2q/(d+1)}$ , for  $q > 0$ , for a *convex body*  $K$  of volume one with smooth boundary, up to constant factors depending on  $K$  in an unknown way. We propose, in this section, a universal deviation inequality and upper bounds on the moments of the missing volume, i.e., results with no restriction on the volume and boundary structure of  $K$ , and with constants which do not depend on  $K$ . The only assumptions on  $K$  are compactness and convexity.

**Theorem 2.13** ([Bru14c]). *There exist two positive constants  $c_1$  and  $c_2$ , which depend on  $d$  only, such that, for all  $n \geq 1$ ,*

$$\sup_{K \in \mathcal{K}_d} \mathbb{P}_K \left[ n \left( \frac{|K \setminus \hat{K}_n|}{|K|} - c_2 n^{-2/(d+1)} \right) > x \right] \leq c_1 e^{-x/(d^d)}, \quad \forall x > 0.$$

In Theorem 2.13, the constant  $c_2$  depends exponentially on the dimension  $d$ . This seems to be the price for getting a uniform deviation inequality on  $\mathcal{K}_d$ . Note that the missing volume is normalized here by the volume of  $K$ . Theorem 2.13 may be refined by normalizing the missing volume by another functional of  $K$ , which could be expressed in terms of the affine surface area of  $K$ , as in [Sch94] where only the first moment of the missing volume is considered.

Note that an easy adaptation of the proof of Theorem 2.13 yields an similar result for the smaller class  $\mathcal{K}_d^{(1)}$ :

$$\sup_{K \in \mathcal{K}_d^{(1)}} \mathbb{P}_K \left[ n \left( |K \setminus \hat{K}_n| - c_4 n^{-2/(d+1)} \right) > x \right] \leq c_1 e^{-x/\beta_d}, \quad \forall x > 0, \quad (2.10)$$

where  $c_4 = c_2 \beta_d / (d^d)$ ,  $c_2$  being the same constant as in Theorem 2.13, and  $c_1$  is the same positive constant as in Theorem 2.13 as well. This deviation inequality, uniform on  $\mathcal{K}_d^{(1)}$ , is useful for adaptation in Theorem 2.12.

Theorem 2.13 allows one to derive upper bounds for all the moments of the missing volume. Indeed, applying Fubini's theorem leads to the following corollary.

**Corollary 2.2.** *For every positive number  $q$ , there exists some positive constant  $c(d, q)$ , which*

depends on  $d$  and  $q$  only, such that, for all  $n \geq 1$ ,

$$\mathbb{E}_K \left[ |K \setminus \hat{K}_n|^q \right] \leq c(d, q) |K|^q n^{-2q/(d+1)}, \quad \forall K \in \mathcal{K}_d.$$

Note that no restriction is made on  $K$  except for its compactness and convexity. In particular, its boundary may not be smooth, and  $K$  may be located anywhere in the space, not necessarily in some given compact set. In this sense, the exponential deviation inequality (2.13) and the inequality on the moments in the previous corollary are universal. Combining this corollary with the lower bound for the minimax risk given in [MT95] in a statistical framework yields the following result.

**Corollary 2.3.** *For every positive number  $q$ , there exist some positive constants  $c_1(d, q)$  and  $c_2(d, q)$ , which depend on  $d$  and  $q$  only, such that, for all  $n \geq 1$ ,*

$$c_1(d, q) n^{-2q/(d+1)} \leq \sup_{K \in \mathcal{K}_d} \mathbb{E}_K \left[ \left( \frac{|K \setminus \hat{K}_n|}{|K|} \right)^q \right] \leq c_2(d, q) n^{-2q/(d+1)}.$$

This corollary shows that the upper bound in Corollary 2.2 is sharp, up to constants. In next Section, we prove an upper bound of the moments of the number of vertices of  $\hat{K}_n$ , and we show that this upper bound is tight.

### 2.3.2 Moments of the number vertices of the convex hull estimator

Let  $q \geq 1$  be a positive integer. Let us reproduce the same scheme of the proof of (2.1). Denote by  $\mathcal{V}_{n+q}$  the set of vertices of  $\hat{K}_{n+q}$ , and by  $V_{n+q}$  its cardinality, i.e., the number of vertices of  $\hat{K}_{n+q}$ . For  $j = 1, \dots, q$ , it is clear that if  $X_{n+j}$  is a vertex of  $\hat{K}_{n+q}$ , then  $X_{n+j} \notin \hat{K}_n$ . Therefore,

$$\begin{aligned} \mathbb{E}_K^{\otimes n} [|K \setminus \hat{K}_n|^q] &= \mathbb{E}_K^{\otimes n} \left[ \int_{K^q} \mathbb{1}(x_{n+1} \notin \hat{K}_n, \dots, x_{n+q} \notin \hat{K}_n) dx_{n+1} \dots dx_{n+q} \right] \\ &= |K|^q \mathbb{P}_K^{\otimes n+q} [X_{n+j} \notin \hat{K}_n, \forall j = 1, \dots, q] \\ &\geq |K|^q \mathbb{P}_K^{\otimes n+q} [X_{n+j} \in \mathcal{V}_{n+q}, \forall j = 1, \dots, q] \\ &= \frac{|K|^q}{\binom{n+q}{q}} \sum_{1 \leq i_1 < \dots < i_q \leq n+q} \mathbb{P}_K^{\otimes n+q} [X_{i_j} \in \mathcal{V}_{n+q}, \forall j = 1, \dots, q] \\ &= \frac{|K|^q}{\binom{n+q}{q}} \mathbb{E}_K^{\otimes n+q} \left[ \sum_{1 \leq i_1 < \dots < i_q \leq n+q} \mathbb{1}(X_{i_j} \in \mathcal{V}_{n+q}, \forall j = 1, \dots, q) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{|K|^q}{\binom{n+q}{q}} \mathbb{E}_K^{\otimes n+q} \left[ \binom{V_{n+q}}{q} \right] \\
&= \frac{|K|^q \mathbb{E}_K^{\otimes n+q} [V_{n+q}(V_{n+q}-1) \dots (V_{n+q}-q+1)]}{(n+q)(n+q-1) \dots (n+1)}.
\end{aligned}$$

This yields the following:

$$\begin{aligned}
&\mathbb{E}_K^{\otimes n+q} [V_{n+q}(V_{n+q}-1) \dots (V_{n+q}-q+1)] \\
&\leq \mathbb{E}_K^{\otimes n} \left[ \left( \frac{|K \setminus \hat{K}_n|}{|K|} \right)^q \right] (n+q)(n+q-1) \dots (n+1). \tag{2.11}
\end{aligned}$$

Combined with Corollary 2.2, (2.11) implies:

$$\sup_{K \in \mathcal{K}_d} \mathbb{E}_K [V_n(V_n-1) \dots (V_n-q+1)] \leq cn^{\frac{q(d-1)}{d+1}},$$

where  $c$  is a positive constant, which depends on  $d$  and  $q$  only. Since the polynomial  $x^q$  is a linear combination of the polynomials  $x(x-1) \dots (x-k+1)$ ,  $0 \leq k \leq q$ , the following inequality:

$$\sup_{K \in \mathcal{K}_d} \mathbb{E}_K [V_n^q] \leq B_{d,q} n^{\frac{q(d-1)}{d+1}}, \tag{2.12}$$

for some positive constant  $B_{d,q}$  which depends on  $d$  and  $q$  only. Note that (2.12) can be extended to non integer values of  $q$ , by Hölder inequality.

Let  $K \in \mathcal{K}_d$ ,  $n$  a positive integer, and  $q \geq 1$ . By Hölder inequality,  $\mathbb{E}_K [V_n^q] \geq \mathbb{E}_K [V_n]^q$  and, by Efron's identity (2.1),

$$\mathbb{E}_K [V_n^q] \geq n^q \mathbb{E}_K \left[ \frac{|K \setminus \hat{K}_{n-1}|}{|K|} \right]^q.$$

If the affine surface area of  $K$  is positive, which occurs, for instance, when the boundary of  $K$  is smooth with positive Gauss curvature, then it is known that  $\mathbb{E}_K \left[ \frac{|K \setminus \hat{K}_{n-1}|}{|K|} \right]$  is exactly of the order of  $n^{-2/(d+1)}$  (see [Sch93b]). Therefore, for such convex bodies  $K$ , the rate of the upper bound in (2.12) is tight.

**Theorem 2.14.** *Let  $n$  be a positive integer, and  $q$  any real number greater or equal to one. Then, for some positive constants  $b_{d,q}$  and  $B_{d,q}$  which depend on  $d$  and  $q$  only,*

$$b_{d,q} n^{\frac{q(d-1)}{d+1}} \leq \sup_{K \in \mathcal{K}_d} \mathbb{E}_K [V_n^q] \leq B_{d,q} n^{\frac{q(d-1)}{d+1}}.$$

### 2.3.3 Efron's identity revisited

#### An extension of Efron's identity

We have already mentioned Efron's identity (2.1) several times. It connects the expected volume of the convex hull  $\hat{K}_n$  of  $n$  i.i.d. random points uniformly distributed in some convex body, and the expected number of vertices of  $\hat{K}_{n+1}$ . We used a similar idea in Section 2.3.2. As in the proof of (2.1), cf. Section 2.5.2, it is clear that the vertices of  $\hat{K}_{n+1}$  are those points of the sample of size  $n+1$ , which are not included in the convex hull of all the other  $n$  points of that sample. Let now  $\mathcal{C}$  be a given class of compact and measurable sets of positive volume in  $\mathbb{R}^d$ , and let  $G \in \mathcal{C}$ . Let  $X_1, \dots, X_{n+1}$  be i.i.d. random variables, uniformly distributed in  $G$ . Denote by  $\hat{G}_n$  the maximum likelihood estimator on the class  $\mathcal{C}$ , based on  $X_1, \dots, X_n$  and defined as

$$\hat{G}_n \in \underset{G' \in \mathcal{C}, X_i \in G', i=1, \dots, n}{\operatorname{argmin}} |G'|. \quad (2.13)$$

This is the maximum likelihood estimator, when the candidates are all members of  $\mathcal{C}$ . The random set  $\hat{G}_n$  is a set of minimal volume among all sets in  $\mathcal{C}$ , which contain the whole sample  $X_1, \dots, X_n$ . We assume there exists such a set. By definition,  $|\hat{G}_n| \leq |G|$  and therefore  $|\hat{G}_n \triangle G| \leq 2|G \setminus \hat{G}_n|$ . For  $i = 1, \dots, n+1$ , consider

$$\hat{G}_{n+1, -i} \in \underset{G' \in \mathcal{C}, X_j \in G', j=1, \dots, n+1, j \neq i}{\operatorname{argmin}} |G'|.$$

This is the maximum likelihood estimator on the class  $\mathcal{C}$ , based on the sample of size  $n+1$  excluding the  $i$ -th observation. Assume that for each  $i$ ,  $\hat{G}_{n+1, -i}$  exists and that it is unique  $\mathbb{P}_G$ -almost surely. Note that  $\hat{G}_n = \hat{G}_{n+1, -(n+1)}$ . Then, similarly to the argument in Section 2.2.1, it is clear that

$$\mathbb{E}_G[|\hat{G}_n \triangle G|] \leq \frac{2|G|\mathbb{E}_G[V'_{n+1}]}{n+1}, \quad (2.14)$$

where  $V'_{n+1}$  is the number of such  $X_i$ , that do not belong to  $\hat{G}_{n+1, -i}$ , for  $i = 1, \dots, n+1$ . If  $\mathcal{C} = \mathcal{K}_d$ , it is clear that  $V'_{n+1} = V_{n+1}$ , which is the number of vertices of  $\hat{K}_{n+1}$ . We saw, in Section 2.2.1, that the properties of the random variable  $V'_{n+1}$  are more intricate if  $\mathcal{C} = \mathcal{P}_r$ . In particular, in that case, it is not clear that this variable is bounded from above by a constant which does not depend on  $n$ , as would suggest our intuition. However, there are some other cases for which the properties of the random variable  $V'_{n+1}$  are more simple, and we give as an example, the case of closed Euclidean balls.

## Parametric families of supports

Let us keep the notation of the previous section, and denote by  $V''_{n+1}$  the number of points of the sample  $X_1, \dots, X_{n+1}$  which are on the boundary of  $\hat{C}_{n+1}$ . The main idea of this section is to use that if  $X_i$  does not belong to  $\hat{C}_{n+1,-i}$ , then it should be on the boundary  $\partial\hat{C}_{n+1}$  of  $\hat{C}_{n+1}$ . In Section 2.2.1, we saw that this is not true if  $\mathcal{C} = \mathcal{P}_r$ . However, this holds for several classes of sets, i.e.,  $V'_{n+1} = V''_{n+1}$ . For some parametric families  $\mathcal{C}$ , the boundary of any  $C' \in \mathcal{C}$  is entirely determined by a finite number of parameters, and therefore, among points in *general position*, only a fixed number of them can be on the boundary of one and the same set  $C' \in \mathcal{C}$ . This number will be an upper bound for  $V''_{n+1}$ ,  $\mathbb{P}_G$ -almost surely. We do not give a precise definition of the notion of being in *general position*, but if  $n$  is large enough, any set of  $n$  i.i.d. random points in  $\mathbb{R}^d$ , whose probability distribution is absolutely continuous with respect to the Lebesgue measure, is in general position,  $\mathbb{P}_G$ -almost surely. We propose the following example.

**Estimation of balls** Consider the class  $\mathcal{B}_d$  of all closed Euclidean balls in  $\mathbb{R}^d$  with positive radius. In this case,  $\hat{C}_{n+1}$  is the ball of smallest radius, containing the whole sample  $X_1, \dots, X_{n+1}$ . The smallest enclosing circle problem, in the plane, was originally posed by Sylvester [Syl57], and its generalizations to higher dimensions are still challenging problems, both for theoretical and computational reasons, see for instance [Wel91, CHM06, DMM07] and the references therein. If  $S$  is a finite subset of  $\mathbb{R}^d$ , we denote by  $\mathcal{B}(S)$  the smallest closed Euclidean ball which contains  $S$ . It is called the smallest enclosing ball of  $S$ . We give two useful lemmas.

**Lemma 1** (Existence and Uniqueness of the Enclosing Ball). *Let  $S$  be a set of  $n$  distinct points in  $\mathbb{R}^d$ , for  $n \geq 2$ . The smallest enclosing ball  $\mathcal{B}(S)$  exists and is unique.*

The uniqueness is very simple and intuitive. Let  $B_1$  and  $B_2$  be two closed balls of minimal radius, both containing  $S$ , and assume  $B_1 \neq B_2$ . Let  $a_1$  and  $a_2$  the respective centers of  $B_1$  and  $B_2$ , and let  $r > 0$  be their common radius. It is clear that  $B_1 \cap B_2 \neq \emptyset$ , since both  $B_1$  and  $B_2$  contain  $S$ . Thus,  $r \geq \frac{\rho(a_1, a_2)}{2}$ . Let  $a = \frac{a_1 + a_2}{2}$ . The Euclidean ball  $B_0 = B_2^d \left( a, \sqrt{r^2 - \frac{\rho(a_1, a_2)^2}{4}} \right)$  contains  $B_1 \cap B_2$ , and therefore  $S \subseteq B_0$ , and its volume is strictly less than  $r$ . This contradicts the minimality of  $r$ .

**Lemma 2.** *Let  $S$  be a finite set of points in  $\mathbb{R}^d$ , and  $x \in \mathbb{R}^d \setminus S$ . If  $x \notin \mathcal{B}(S)$ , then  $x \in \partial\mathcal{B}(S \cup \{x\})$ .*

The proofs of these lemmata can be found in the appendix of this chapter.



Therefore, it follows that (2.14) can be applied, with  $V'_{n+1}$  being the number of  $X_i, i = 1, \dots, n+1$  which belong to the boundary of  $\hat{C}_{n+1}$ , by Lemma 2. The weighted minimax risk on the class  $\mathcal{B}_d$ , for Model (DS), is at most of the order  $1/n$ :

**Theorem 2.15.** *For all  $n \in \mathbb{N}^*$ ,*

$$\mathcal{Q}_n(\mathcal{B}_d) \leq \frac{2d+2}{n}.$$

Proving a lower bound of the order of  $1/n$  for  $\mathcal{Q}_n(\mathcal{B}_d)$  is straightforward, using usual arguments for the lower bound.

### Non parametric families of supports

In this section, our main purpose is to present general ideas, and to show that they can lead to optimal results.

Let  $\mathcal{C}$  be a class of compact and measurable subsets of  $[0, 1]^d$  with positive volume. Let  $(\mathcal{C}_M)_{M \geq 1}$  be a nested family of parametric classes of subsets of  $\mathbb{R}^d$  :  $\mathcal{C}_M \subseteq \mathcal{C}_{M+1}, \forall M \geq 1$ . Assume that there exists a positive number  $\alpha$ , which may depend on  $d$ , such that for all  $M \geq 1$ , the dimension of  $\mathcal{C}_M$  is less or equal to  $\alpha M$ , i.e., there is an surjective function from a subset of  $\mathbb{R}^k$  onto  $\mathcal{C}_M$ , for some  $k \leq \alpha M$ . This means that any  $G \in \mathcal{C}_M$  can be described by a parameter  $\theta \in \mathbb{R}^k$ . Assume that each  $G \in \mathcal{C}$  can be well approximated by some  $G_M^* \in \mathcal{C}_M$ , containing  $G$ . Namely, assume that there exists a parameter  $\beta > 0$  and a positive constant  $c$ , such that for all  $G \in \mathcal{C}$  and all  $M \geq 1$ , there exists  $G_M^* \in \mathcal{C}_M$  satisfying  $G \subseteq G_M^*$ , and such that:

$$|G_M^* \setminus G| \leq cM^{-\beta}.$$

Note that the classes  $\mathcal{C}_M, M \geq 1$  need not be included in  $\mathcal{C}$ . In particular, we do not assume that the sets which belong to  $\mathcal{C}_M$  are included in  $[0, 1]^d$ .

Assume Model (DS), with  $G \in \mathcal{C}$ . For  $M \geq 1$ , let  $\hat{G}_n^{(M)}$  be the maximum likelihood estimator of  $G$ , selected from the class of candidates  $\mathcal{C}_M$ :

$$\hat{G}_n^{(M)} \in \underset{G' \in \mathcal{C}_M, X_i \in G', i=1, \dots, n}{\operatorname{argmin}} |G'|.$$

For  $i = 1, \dots, n+1$ , define also

$$\hat{G}_{n+1, -i}^{(M)} \in \underset{G' \in \mathcal{C}_M, X_j \in G', j=1, \dots, n+1, j \neq i}{\operatorname{argmin}} |G'|.$$

Assume that for each  $i$ ,  $\hat{G}_{n+1, -i}^{(M)}$  exists and is unique,  $\mathbb{P}_G$ -almost surely. By definition,  $|\hat{G}_n^{(M)}| \leq$

$|G_M^*|$  and therefore,  $|\hat{G}_n^{(M)} \triangle G_M^*| \leq 2|G_M^* \setminus \hat{G}_n^{(M)}|$ . Let us bound the risk of  $\hat{G}_n^{(M)}$ :

$$\begin{aligned}
\mathbb{E}_G[|\hat{G}_n^{(M)} \triangle G|] &\leq \mathbb{E}_G[|\hat{G}_n^{(M)} \triangle G_M^*|] + |G_M^* \triangle G| \quad \text{by the triangle inequality} \\
&\leq 2\mathbb{E}_G[|G_M^* \setminus \hat{G}_n^{(M)}|] + cM^{-\beta} \\
&\leq 2|G_M^* \setminus G| + 2\mathbb{E}_G[|G \setminus \hat{G}_n^{(M)}|] + cM^{-\beta} \\
&\leq 2\mathbb{E}_G[|G \setminus \hat{G}_n^{(M)}|] + 3cM^{-\beta},
\end{aligned} \tag{2.15}$$

where we used the inequality  $|A \setminus C| \leq |A \setminus B| + |B \setminus C|$ , for all measurable sets  $A, B$  and  $C$ . Following (2.14),

$$\mathbb{E}_G[|G \setminus \hat{G}_n^{(M)}|] \leq \frac{|G| \mathbb{E}_G[V'_{n+1}]}{n+1}$$

where  $V'_{n+1}$  is the number of  $X_i$  that do not belong to  $\hat{G}_{n+1,-i}^{(M)}$ , for  $i = 1, \dots, n+1$ . Then, continuing (2.15), and recalling that  $|G| \leq 1$ , since  $G \subseteq [0, 1]^d$ ,

$$\mathbb{E}_G[|\hat{G}_n^{(M)} \triangle G|] \leq \frac{2\mathbb{E}_G[V'_{n+1}]}{n+1} + 3cM^{-\beta}. \tag{2.16}$$

Assume that if  $X_i \notin \hat{G}_{n+1,-i}^{(M)}$ , then  $X_i \in \partial \hat{G}_{n+1}^{(M)}$ ,  $\mathbb{P}_G$ -almost surely. Then,  $V'_{n+1} \leq V''_{n+1}$ , where  $V''_{n+1}$  is the number of  $X_i$ 's that are on the boundary of  $\hat{G}_{n+1}^{(M)}$ ,  $i = 1, \dots, n+1$ . It is natural to assume that  $V''_{n+1} \leq \alpha M$   $\mathbb{P}_G$ -almost surely. Indeed,  $\hat{G}_{n+1}^{(M)}$  belongs to the parametric class  $\mathcal{C}_M$ , and thus, it is entirely determined by at most  $\alpha M$  real parameters. Then, (2.16) becomes:

$$\mathbb{E}_G[|\hat{G}_n^{(M)} \triangle G|] \leq \frac{2M}{n} + 3cM^{-\beta}. \tag{2.17}$$

An optimal choice of the tuning parameter  $M$  is  $M^* = \lfloor n^{1/(\beta+1)} \rfloor$ , and yields:

$$\mathbb{E}_G[|\hat{G}_n^{(M^*)} \triangle G|] \leq c' n^{-\frac{\beta}{\beta+1}}, \forall G \in \mathcal{C}, \tag{2.18}$$

where  $c'$  is a constant independent on  $n$  and  $G$ . Let us give two examples.

**Convex bodies and circumscribed polytopes** Let  $\mathcal{C} = \mathcal{K}_d^{(1)}$ , and let  $\mathcal{C}_M$  be the class of all  $d$ -dimensional polytopes with at most  $M$  facets. We recall that a facet is a  $(d-1)$ -dimensional face. The dimension of  $\mathcal{C}_M$  is  $(d+1)M$ . In addition, it is well known (see [Gru91, GG97]) that every  $G \in \mathcal{K}_d^{(1)}$  can be approximated by a circumscribed polytope  $G_M^* \in \mathcal{C}_M$  satisfying  $|G_M^* \setminus G| \leq cM^{-2/(d-1)}$ , for some positive constant  $c$  which depends on  $d$  only. Here,  $\beta = 2/(d-1)$  and  $\beta/(\beta+1) = 2/(d+1)$ , which is the exponent of the minimax rate on the class  $\mathcal{K}_d^{(1)}$ .

**Boundary fragments** The second example deals with boundary fragments. Let us introduce some notation. Let  $L^2([0, 1])$  be the set of all real valued measurable functions  $f$  defined on the interval  $[0, 1]$ , such that  $\int_0^1 f^2 < \infty$ . The Fourier basis  $(\phi_j)_{j \geq 0}$  of  $L^2([0, 1])$  is defined as follows. Let  $\phi_0(x) = 1, \forall x \in [0, 1]$ . For  $j \in \mathbb{N}^*$ , let  $\phi_{2j-1}$  and  $\phi_{2j}$  be the functions

$$\phi_{2j-1}(x) = \sqrt{2} \sin(2\pi jx) \text{ and } \phi_{2j}(x) = \sqrt{2} \cos(2\pi jx), \forall x \in [0, 1].$$

Let the sequence  $(a_j)_{j \geq 0}$  be defined as

$$a_j = \begin{cases} 2\pi(j+1) & \text{if } j \text{ is odd,} \\ 2\pi j & \text{if } j \text{ is even.} \end{cases}$$

The Fourier coefficients of a function  $f \in L^2([0, 1])$  are the projections of  $f$  on the complete orthonormal system  $\{\phi_j : j \geq 0\}$  of  $L^2([0, 1])$ , with respect to the scalar product  $\langle f_1, f_2 \rangle = \int_0^1 f_1 f_2, \forall f_1, f_2 \in L^2([0, 1])$ . Let  $\gamma$  and  $L$  be positive numbers, and let  $\Theta(\gamma, L)$  be the class of all real valued sequences  $(\theta^{(j)})_{j \geq 0}$  satisfying

$$\sum_{j=1}^{\infty} a_j^\gamma |\theta^{(j)}| \leq L.$$

Denote by  $\Sigma(\gamma, L)$  the class of all functions  $f \in L^2([0, 1])$ , whose coefficients are in  $\Theta(\gamma, L)$ , and such that  $0 \leq f \leq 1$ . If  $M$  is a positive integer, consider

$$\Theta_M(\gamma, L) = \left\{ (\tau^{(0)}, \dots, \tau^{(M-1)}) \in \mathbb{R}^M : \sum_{j=0}^{M-1} a_j^\gamma \tau^{(j)} \leq L. \right\},$$

and denote by  $\Sigma_M(\gamma, L)$  the class of all functions  $f = \sum_{j=0}^{M-1} \tau^{(j)} \phi_j$ , such that  $(\tau^{(0)}, \dots, \tau^{(M-1)}) \in \Theta_M(\gamma, L)$ . Let  $f = \sum_{j=0}^{\infty} \tau^{(j)} \phi_j \in \Sigma(\gamma, L)$ , and let  $f_M$  be the truncation of  $f$  up to the  $M$  first terms:  $f_M = \sum_{j=0}^{M-1} \tau^{(j)} \phi_j$ . For  $x \in [0, 1]$ ,

$$|f(x) - f_M(x)| \leq \sqrt{2} \sum_{j=M}^{\infty} |\tau^{(j)}|$$

$$\begin{aligned}
&\leq \frac{\sqrt{2}}{a_M^\gamma} \sum_{j=M}^{\infty} a_M^\gamma |\tau^{(j)}| \\
&\leq \frac{\sqrt{2}L}{(2\pi M)^\gamma}.
\end{aligned} \tag{2.19}$$

Thus, if we consider  $f_M^* = f_M + \frac{\sqrt{2}L}{(2\pi M)^\gamma}$ , then

$$f_M^* \geq f, \tag{2.20}$$

and  $f_M^* \in \Sigma_M(\gamma, L)$ . In addition, by (2.19),

$$\int_0^1 |f_M^* - f| \leq \frac{2\sqrt{2}L}{(2\pi M)^\gamma}. \tag{2.21}$$

Denote by  $G$  the boundary fragment associated to  $f$ , that is,

$$G = \{(x, y) \in [0, 1] \times \mathbb{R} : 0 \leq y \leq f(x)\}.$$

Similarly, denote by  $G_M^*$  the boundary fragment associated to  $f_M^*$ . Denote by  $\mathcal{C}(\gamma, L)$  the class of all sets which are boundary fragments associated to a function in  $\Sigma(\gamma, L)$ , and by  $\mathcal{C}_M(\gamma, L)$  the class of all sets which are boundary fragments associated to a function in  $\Sigma_M(\gamma, L)$ , for  $M \geq 1$ . For  $M \geq 1$ , the class  $\Sigma_M(\gamma, L)$  has dimension  $M$ . In addition, by (2.20) and (2.21),

$$G \subseteq G_M^*$$

and

$$|G_M^* \setminus G| \leq \frac{2\sqrt{2}L}{(2\pi M)^\gamma}.$$

Here,  $\beta = \gamma$ , and  $\beta/(\beta + 1) = \gamma/(\gamma + 1)$ , which is a usual exponent for the minimax rate of convergence.

### 2.3.4 Non uniform distributions

#### The density is separated from zero

Model (DS) consists in observing realizations of a uniform distribution on some unknown subset of  $\mathbb{R}^d$ . The assumption that the underlying distribution is uniform is very strong, and may be too restrictive. An easy relaxation consists in assuming that the density is supported on an unknown convex body, and that it is separated from zero on its support. Let  $a \in (0, 1]$  be a given positive number. For  $K \in \mathcal{K}_d$ , denote by  $\mathcal{F}_K(a)$  the class of all densities  $f$  on  $\mathbb{R}^d$ ,

such that  $\forall x \in K, f(x) \geq \frac{a}{|K|}$  and  $\forall x \in \mathbb{R}^d \setminus K, f(x) = 0$ . Assume that  $X_1, \dots, X_n$  are i.i.d. random variables, with density  $f \in \mathcal{F}_K(a)$ , for some  $K \in \mathcal{K}_d$ . As for Model (DS), the maximum likelihood estimator of  $K$  is still  $\hat{K}_n = CH(X_1, \dots, X_n)$ , or  $\hat{P}_n^{(r)}$  if  $K$  is known to belong to the class  $\mathcal{P}_r^{(1)}$ .

It is easy to adapt the proofs of Theorem 2.13 and 2.9 to show the following.

**Theorem 2.16.** *Let  $a \in (0, 1]$ . There exist a positive constant  $c_1$  which depends on  $d$  only and a positive constant  $c_3$ , which depends on  $d$  and  $a$  only, such that:*

$$\sup_{K \in \mathcal{K}_d} \sup_{f \in \mathcal{F}_K(a)} \mathbb{P}_f \left[ n \left( \frac{|K \setminus \hat{K}_n|}{|K|} - c_3 n^{-2/(d+1)} \right) > x \right] \leq c_1 e^{-ax/(d^d)}, \forall x > 0.$$

The constant  $c_1$  is the same as in Theorem 2.13, and  $c_3 = c_2/a$ , where  $c_2$  is the same as in Theorem 2.13.

In the polytopal case, the following theorem holds:

**Theorem 2.17.** *Let  $a \in (0, 1]$ . Let  $r \geq d + 1$  be an integer, and  $n \geq 2$ . Then, there exists a positive constant  $c_1$ , which depends on  $d$  and  $a$  only, such that:*

$$\sup_{P \in \mathcal{P}_r^{(1)}} \sup_{f \in \mathcal{F}_P(a)} \mathbb{P}_f \left[ n \left( |\hat{P}_n^{(r)} \triangle P| - \frac{4dr \ln n}{an} \right) \geq x \right] \leq c_1 e^{-ax/2}, \forall x > 0.$$

From these two theorems, it follows that  $K \in \mathcal{K}_d$  and  $P \in \mathcal{P}_r^{(1)}$  can be estimated at the same speed of convergence in the non uniform case considered in this section, as in the uniform case. It is clear that the lower bounds proved in the uniform case still hold here, since the uniform density on a convex body  $K$  belongs to the class  $\mathcal{F}_K(a)$ , for all  $a \in (0, 1]$ .

### The density decreases slowly to zero

Estimation of the compact and convex support  $K$  of a density in  $\mathbb{R}^d$  seems still possible if that density does not decrease too fast to zero near the boundary of  $K$ . Let  $\alpha, L$  be positive numbers. Assume that  $f$  is a density supported on some  $K \in \mathcal{K}_d$ . For  $\epsilon \in (0, 1)$ , consider  $K_\epsilon = \left\{ x \in K : f(x) \leq \frac{\epsilon}{|K|} \right\}$ . Suppose that  $f$  satisfies a *margin condition* of this kind:

$$|K_\epsilon| \leq L|K|\epsilon^\alpha, \forall \epsilon \in (0, 1), \quad (2.22)$$

Similar conditions appear in classification, e.g. in smooth discrimination analysis, see [MT99], and in inference on level sets [Pol95, Tsy97]. In [CF97], such an assumption is made on the density whose support is to be estimated, using a plug-in estimator. It allows to separate

well but smoothly the two sets we are interested in here, namely  $K$  and its complement. Other conditions of slow decrease of  $f$  near the boundary of  $K$  can be considered. See [Tsy97] for example:  $f(x)$  can be assumed to decrease at most as  $r^{1/\alpha}$ , where  $r$  is the distance from  $x$  to  $\partial K$ .

Let  $X_1, \dots, X_n$  be i.i.d. random points with density  $f$  in  $\mathbb{R}^d$ . Again, the maximum likelihood estimator of  $K \in \mathcal{K}_d$  is  $\hat{K}_n = CH(X_1, \dots, X_n)$ . Let us adapt the proof of Efron's identity (2.1) to the present case. For any  $\epsilon \in (0, 1)$ , we have :

$$\begin{aligned}
\mathbb{E}_f^{\otimes n} [ |K \setminus \hat{K}_n| ] &= \mathbb{E}_f^{\otimes n} \left[ \int_K \mathbb{1}(x \notin \hat{K}_n) dx \right] \\
&= \mathbb{E}_f^{\otimes n} \left[ \int_{K_\epsilon} \mathbb{1}(x \notin \hat{K}_n) dx + \int_{K \setminus K_\epsilon} \mathbb{1}(x \notin \hat{K}_n) dx \right] \\
&\leq |K_\epsilon| + \frac{|K|}{\epsilon} \mathbb{E}_f^{\otimes n} \left[ \int_{K \setminus K_\epsilon} \mathbb{1}(x \notin \hat{K}_n) f(x) dx \right] \\
&\leq L|K|\epsilon^\alpha + \frac{|K|}{\epsilon} \mathbb{P}_f^{\otimes n+1} [X_{n+1} \notin \hat{K}_n] \\
&\leq L|K|\epsilon^\alpha + \frac{|K| \mathbb{E}_f^{\otimes n+1} [V_{n+1}]}{(n+1)\epsilon},
\end{aligned} \tag{2.23}$$

where  $V_{n+1}$  is the number of vertices of  $\hat{K}_{n+1}$ .

The choice  $\epsilon = \left( \mathbb{E}_f^{\otimes n+1} [V_{n+1}] / (n+1) \right)^{1/(\alpha+1)}$  yields:

**Theorem 2.18.** *Let  $K \in \mathcal{K}_d$ , and  $f$  a density supported on  $K$  and satisfying (2.22). The risk of the convex hull estimator is bounded from above by:*

$$\mathbb{E}_f^{\otimes n} \left[ \frac{|K \setminus \hat{K}_n|}{|K|} \right] \leq (L+1) \left( \frac{\mathbb{E}_f^{\otimes n+1} [V_{n+1}]}{(n+1)} \right)^{\frac{\alpha}{\alpha+1}}.$$

The uniform distribution on  $K$  corresponds to the limiting case,  $L = 0$ ,  $\alpha = \infty$ , and this theorem gives exactly Efron's identity (2.1). In general,  $\mathbb{E}_f^{\otimes n+1} [V_{n+1}]$  is not known, and its computation remains an open problem. It can take a broad range of values, see for example the work [HRW08] in the two-dimensional case, taken from game theory. Note that if  $f$  is a concave function and  $\alpha = 2$ , then  $f$  is the density of the random variable  $\pi(\tilde{X})$ , where  $\pi$  is the orthogonal projection of  $\mathbb{R}^{d+1}$  on  $\mathbb{R}^d$  and  $\tilde{X}$  is a random variable with uniform distribution on a convex body  $\tilde{K} \in \mathcal{K}_{d+1}$  satisfying  $\pi(\tilde{K}) = K$ . Therefore, in that case,  $V_{n+1}$  is the number of vertices of the orthogonal projection in  $\mathbb{R}^d$  of the convex hull of  $n+1$  i.i.d. random variables uniformly distributed in  $\tilde{K}$ .

In the one-dimensional case,  $d = 1$ ,  $V_{n+1} = 2$ ,  $\hat{K}_n = [X_{(1)}, X_{(n)}]$ , where  $X_{(i)}$  is the  $i$ -th order statistic of  $X_1, \dots, X_n$ , and Theorem 2.18 yields:

$$\mathbb{E}_f^{\otimes n} \left[ \frac{|K \setminus \hat{K}_n|}{|K|} \right] \leq 2^{\frac{1}{\alpha+1}} (L+1) n^{-\frac{\alpha}{\alpha+1}}.$$

### 2.3.5 Estimation of log-concave densities

Model (DS) consists of estimating the convex support of a uniform density in  $\mathbb{R}^d$ . Such densities are log-concave, i.e. their logarithm, taking values in  $[-\infty, \infty)$ , is a concave function. Estimation of log-concave densities in  $\mathbb{R}^d$  is a particular case of density estimation under shape constraints. It is a challenging domain, and log-concave densities are of particular interest, especially for some of their properties, which are useful in many applications. First of all, some very important and useful densities are log-concave, such as Gaussian densities, logistic, Gumbel, Laplace, etc. Their level sets are convex. They are unimodal, and their convolution with any unimodal density is unimodal. In addition, the class of log-concave densities is closed under weak convergence.

In some recent leading works [CS10, CSS10, DW13], the maximum likelihood estimator of a log-concave density has been studied. In particular, this estimator is proved to exist and to be unique, and it is a tent function: it is supported on the convex hull of the sample, and the subgraph of its logarithm on its support is a convex polytope. Its consistency and its rate of convergence, measured with the Hellinger distance, have been studied, mostly in small dimensions,  $d = 1, 2$ . In case of misspecification, i.e., when the true density is not log-concave, under mild assumptions, the maximum likelihood estimator converges to the projection of that density - with respect to the Kullback-Leiber divergence - on the class of log-concave densities.

Some questions, related to our work, remain open. In particular, how many vertices does the graph of the maximum likelihood estimator have? More generally, how many  $k$ -faces does it have, for  $k = 1, \dots, d-1$ ? Is it possible, similarly to Efron's identity, to connect the risk of this estimator to its geometrical properties? Does this estimator achieve the minimax rate of convergence, if measured using the Hellinger distance between densities? Or using a  $L^p$  distance?

## 2.4 Proofs

### Proof of Theorem 2.9

Let  $r \geq d + 1$  be an integer, and  $n \geq 2$ . Let  $P_0 \in \mathcal{P}_r^{(1)}$  and consider a sample  $X_1, \dots, X_n$  of i.i.d. random variables with uniform distribution on  $P_0$ . For simplicity's sake, we will denote  $\hat{P}_n$  instead of  $\hat{P}_n^{(r)}$  in this Section.

Let  $\hat{P}_n$  be the estimator defined in Theorem 2.9. Recall that  $\mathcal{P}_{r,n}^{(1)}$  is the class of all convex polytopes of  $\mathcal{P}_r^{(1)}$  whose vertices lay on the grid  $(\frac{1}{n}\mathbb{Z})^d$ , i.e. have as coordinates integer multiples of  $1/n$ . We first apply Lemma 6, in Section 2.5.1. In particular, taking  $P = P_0$  or  $P = \hat{P}_n$  in Lemma 6, we can find two polytopes  $P^*$  and  $\tilde{P}_n$  in  $\mathcal{P}_{r,n}^{(1)}$  such that

$$\begin{cases} |P^* \triangle P_0| \leq \frac{K_1}{n} \\ P^* \subseteq P_0^{\sqrt{d}/n}, \quad P_0 \subseteq (P^*)^{\sqrt{d}/n} \end{cases}$$

and

$$\begin{cases} |\tilde{P}_n \triangle \hat{P}_n| \leq \frac{K_1}{n} \\ \tilde{P}_n \subseteq \hat{P}_n^{\sqrt{d}/n}, \quad \hat{P}_n \subseteq \tilde{P}_n^{\sqrt{d}/n}. \end{cases}$$

Note that  $\tilde{P}_n$  is random. Let  $\epsilon > 0$ . By construction,  $|\hat{P}_n| \leq |P_0|$ , so  $|\hat{P}_n \triangle P_0| \leq 2|P_0 \setminus \hat{P}_n|$ . Besides, if  $G_1, G_2$  and  $G_3$  are three measurable subsets of  $\mathbb{R}^d$ , the following triangle inequality holds :

$$|G_1 \setminus G_3| \leq |G_1 \setminus G_2| + |G_2 \setminus G_3|. \quad (2.24)$$

Let us now write the following inclusions between the events.

$$\begin{aligned} \left\{ |\hat{P}_n \triangle P_0| > \epsilon \right\} &\subseteq \left\{ |P_0 \setminus \hat{P}_n| > \epsilon/2 \right\} \\ &\subseteq \left\{ |P^* \setminus \tilde{P}_n| > \epsilon/2 - \frac{2K_1}{n} \right\} \\ &\subseteq \bigcup_P \left\{ \tilde{P}_n = P \right\}, \end{aligned} \quad (2.25)$$

where the last union is over the class of all  $P \in \mathcal{P}_{r,n}^{(1)}$  that satisfy the inequality  $|P^* \setminus P| > \epsilon/2 - \frac{2K_1}{n}$ . Let  $P$  be such a polytope, then if  $\tilde{P}_n = P$ , then necessarily the sample  $\{X_1, \dots, X_n\}$



is included in  $P^{\frac{\sqrt{d}}{n}}$ , by definition of  $\tilde{P}_n$ , and (2.25) becomes

$$\begin{aligned}
\mathbb{P}_{P_0} [\tilde{P}_n = P] &\leq \mathbb{P}_{P_0} [X_i \in P^{\frac{\sqrt{d}}{n}}, i = 1, \dots, n] \\
&\leq \left(1 - \frac{|P_0 \setminus P^{\frac{\sqrt{d}}{n}}|}{|P_0|}\right)^n \\
&\leq \left(1 - |P_0 \setminus P^{\frac{\sqrt{d}}{n}}|\right)^n, \quad \text{since } |P_0| \leq 1 \\
&\leq \left(1 - |P^* \setminus P| + |P^* \setminus P_0| + |P^{\frac{\sqrt{d}}{n}} \setminus P|\right)^n, \quad \text{using (2.24) twice} \\
&\leq \left(1 - \epsilon/2 + \frac{4K_1}{n}\right)^n \\
&\leq C_1 \exp(-n\epsilon/2),
\end{aligned} \tag{2.26}$$

where  $C_1 = e^{4K_1}$ . Therefore, using (2.25) and (2.26) and denoting by  $\#\mathcal{P}_{r,n}^{(1)}$  the cardinality of the finite class  $\mathcal{P}_{r,n}^{(1)}$ ,

$$\begin{aligned}
P_{P_0} [|\hat{P}_n \triangle P_0| > \epsilon] &\leq \#\mathcal{P}_{r,n}^{(1)} C_1 \exp(-n\epsilon/2) \\
&\leq (n+1)^{dr} C_1 \exp(-n\epsilon/2) \\
&\leq C_1 \exp(-n\epsilon/2 + 2dr \ln n).
\end{aligned} \tag{2.27}$$

It turns out that if we take  $\epsilon$  of the form  $\frac{4dr \ln n}{n} + \frac{x}{n}$ , (2.27) becomes

$$\mathbb{P}_{P_0} \left[ n \left( |\hat{P}_n \triangle P_0| - \frac{4dr \ln n}{n} \right) \geq x \right] \leq C_1 e^{-x/2}, \tag{2.28}$$

which holds for any  $x > 0$  and any  $P_0 \in \mathcal{P}_r^{(1)}$ . Theorem 2.9 is proved.  $\blacksquare$

## Proof of Theorem 2.10

Let  $r \geq 9$  be an integer, supposed to be even, without loss of generality and assume  $n \geq r$ . Consider a regular convex polytope  $P^*$  in  $[0,1]^2$  with center  $C = (1/2, 1/2)$  and with  $r/2$  vertices, denoted by  $A_0, A_2, \dots, A_{r-2}$ , such that for all  $k = 0, \dots, r/2 - 1$ , the distance between  $A_{2k}$  and the center  $C$  is  $1/2$ . Let  $A_1, A_3, \dots, A_{r-1}$  be  $r/2$  points built as in Figure 2.2: for  $k = 0, \dots, r/2 - 1$ ,  $A_{2k+1}$  is on the mediator of the segment  $[A_{2k}, A_{2k+2}]$ , outside  $P^*$ , at a distance  $\delta = h/2 \cos(2\pi/r) \tan(4\pi/r)$  of  $P^*$ , with  $h \in (0, 1)$  to be chosen. Note that by our construction,  $A_{2k}$  and  $A_{2k+2}$  are vertices of the convex hull of  $A_0, A_2, \dots, A_{r-2}$  and  $A_{2k+1}$ .

Let us denote by  $D_k$  the smallest convex cone with apex  $C$ , containing the points  $A_{2k}, A_{2k+1}$  and  $A_{2k+2}$ , as drawn in Figure 2.2. Note that the cones  $D_k$  have pairwise a null Lebesgue

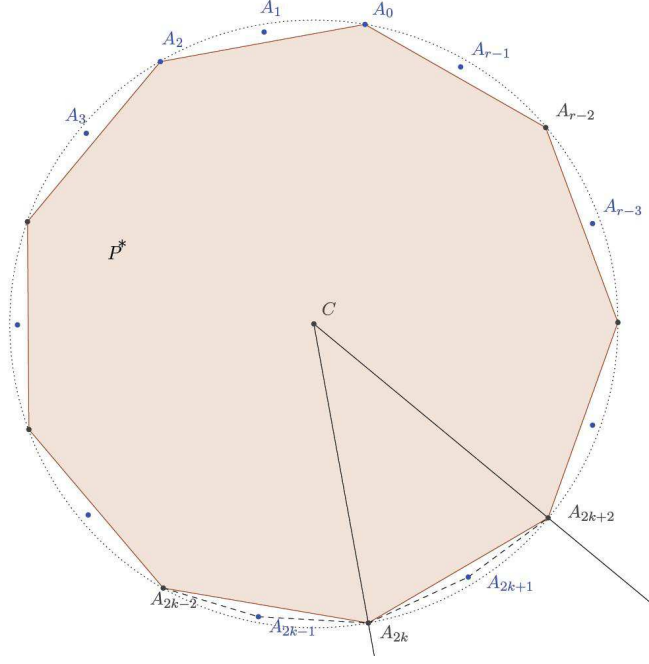


Figure 2.2: Construction of hypotheses for the lower bound

measure intersection. For  $\omega = (\omega_0, \dots, \omega_{r/2-1}) \in \{0, 1\}^{r/2}$ , we denote by  $P_\omega$  the convex hull of  $P^*$  and the points  $A_{2k+1}, k = 0, \dots, r/2 - 1$  such that  $\omega_k = 1$ .

The proof is inspired by Assouad's lemma, see [Tsy09, Section 2.7.2].

For  $k = 0, \dots, r/2 - 1$ , and  $(\omega_0, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_{r/2-1}) \in \{0, 1\}^{r/2-1}$ , we denote by

$$\begin{aligned}\omega^{(k,0)} &= (\omega_0, \dots, \omega_{k-1}, 0, \omega_{k+1}, \dots, \omega_{r/2-1}) \text{ and by} \\ \omega^{(k,1)} &= (\omega_0, \dots, \omega_{k-1}, 1, \omega_{k+1}, \dots, \omega_{r/2-1}).\end{aligned}$$

Note that for  $k = 0, \dots, r/2 - 1$ , and  $(\omega_0, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_{r/2-1}) \in \{0, 1\}^{r/2-1}$ ,

$$|P_{\omega^{(k,0)}} \triangle P_{\omega^{(k,1)}}| = \frac{\delta}{2} \cos(2\pi/r).$$

By a simple computation,

$$\begin{aligned}1 - \frac{H(P_{\omega^{(k,0)}}, P_{\omega^{(k,1)}})^2}{2} &= \sqrt{1 - \frac{|P_{\omega^{(k,1)}} \setminus P_{\omega^{(k,0)}}|}{|P_{\omega^{(k,1)}}|}} \\ &= \sqrt{1 - \frac{\delta/2 \cos(2\pi/r)}{|P_{\omega^{(k,1)}}|}} \\ &\geq \sqrt{1 - \frac{\delta \cos(2\pi/r)}{4}}\end{aligned}\tag{2.29}$$

since  $|P_{\omega^{(k,1)}}| \geq |P^*| \geq 1/2$ . Now, let  $\hat{P}$  be any estimator of  $P^*$ , based on a sample of  $n$  i.i.d. random variables.

We have the following inequalities.

$$\begin{aligned}
& \sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{E}_P \left[ |P \Delta \hat{P}| \right] \\
& \geq \frac{1}{2^{r/2}} \sum_{\omega \in \{0,1\}^{r/2}} \mathbb{E}_{P_\omega} \left[ |P_\omega \Delta \hat{P}| \right] \\
& \geq \frac{1}{2^{r/2}} \sum_{\omega \in \{0,1\}^{r/2}} \sum_{k=0}^{r/2-1} \mathbb{E}_{P_\omega} \left[ |(P_\omega \cap D_k) \Delta (\hat{P} \cap D_k)| \right] \\
& = \frac{1}{2^{r/2}} \sum_{k=0}^{r/2-1} \sum_{\omega \in \{0,1\}^{r/2}} \mathbb{E}_{P_\omega} \left[ |(P_\omega \cap D_k) \Delta (\hat{P} \cap D_k)| \right] \\
& = \frac{1}{2^{r/2}} \sum_{k=0}^{r/2} \sum_{\omega_1, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_{r/2-1}} \left( \mathbb{E}_{P_\omega^{(k,0)}} \left[ |(P_\omega^{(k,0)} \cap D_k) \Delta (\hat{P} \cap D_k)| \right] \right. \\
& \quad \left. + \mathbb{E}_{P_\omega^{(k,1)}} \left[ |(P_\omega^{(k,1)} \cap D_k) \Delta (\hat{P} \cap D_k)| \right] \right). \tag{2.30}
\end{aligned}$$

For  $k = 1, \dots, r/2 - 1$ , and  $(\omega_0, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_{r/2-1}) \in \{0,1\}^{r/2-2}$  we have

$$\begin{aligned}
& \mathbb{E}_{P_\omega^{(k,0)}} \left[ |(P_\omega^{(k,0)} \cap D_k) \Delta (\hat{P} \cap D_k)| \right] + \mathbb{E}_{P_\omega^{(k,1)}} \left[ |(P_\omega^{(k,1)} \cap D_k) \Delta (\hat{P} \cap D_k)| \right] \\
& = \int_{(\mathbb{R}^2)^n} |(P_\omega^{(k,0)} \cap D_k) \Delta (\hat{P} \cap D_k)| d\mathbb{P}_{P_\omega^{(k,0)}}^{\otimes n} \\
& \quad + \int_{(\mathbb{R}^2)^n} |(P_\omega^{(k,1)} \cap D_k) \Delta (\hat{P} \cap D_k)| d\mathbb{P}_{P_\omega^{(k,1)}}^{\otimes n} \\
& \geq \int_{(\mathbb{R}^2)^n} \left( |(P_\omega^{(k,0)} \cap D_k) \Delta (\hat{P} \cap D_k)| + |(P_\omega^{(k,1)} \cap D_k) \Delta (\hat{P} \cap D_k)| \right) \times \\
& \quad \min(d\mathbb{P}_{P_\omega^{(k,0)}}^{\otimes n}, d\mathbb{P}_{P_\omega^{(k,1)}}^{\otimes n}) \\
& \geq \int_{(\mathbb{R}^2)^n} \left( |(P_\omega^{(k,0)} \cap D_k) \Delta (P_\omega^{(k,1)} \cap D_k)| \right) \min(d\mathbb{P}_{P_\omega^{(k,0)}}^{\otimes n}, d\mathbb{P}_{P_\omega^{(k,1)}}^{\otimes n}), \\
& \quad \text{by the triangle inequality,} \\
& = \frac{\delta \cos(2\pi/r)}{2} \int_{(\mathbb{R}^2)^n} \min(d\mathbb{P}_{P_\omega^{(k,0)}}^{\otimes n}, d\mathbb{P}_{P_\omega^{(k,1)}}^{\otimes n}) \\
& \geq \frac{\delta \cos(2\pi/r)}{4} \left( 1 - \frac{H^2(\mathbb{P}_{P_\omega^{(k,0)}}^{\otimes n}, \mathbb{P}_{P_\omega^{(k,1)}}^{\otimes n})}{2} \right)^2
\end{aligned}$$

$$= \frac{\delta \cos(2\pi/r)}{4} \left( 1 - \frac{H^2(\mathbb{P}_{P_\omega^{(k,0)}}, \mathbb{P}_{P_\omega^{(k,1)}})}{2} \right)^{2n}, \quad (2.31)$$

using properties of the Hellinger distance (cf. Lemma 7).

Finally, by (2.29), (2.30) and (2.31),

$$\begin{aligned} \sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{E}_P \left[ |P \Delta \hat{P}_n| \right] &\geq \frac{r\delta \cos(2\pi/r)}{16} \left( 1 - \frac{\delta \cos(2\pi/r)}{4} \right)^n \\ &\geq \frac{rh \cos\left(\frac{2\pi}{r}\right)^2 \tan\left(\frac{4\pi}{r}\right)}{16} \left( 1 - \frac{h \cos\left(\frac{2\pi}{r}\right)^2 \tan\left(\frac{4\pi}{r}\right)}{8} \right)^n. \end{aligned} \quad (2.32)$$

Note that if we denote by  $x = \frac{2\pi}{r} > 0$  and  $\phi(x) = \frac{1}{x} \cos(x)^2 \tan(2x)$ , then  $\phi(x) \geq c_1$  for some universal positive constant  $c_1$ , since  $r$  is supposed to be greater or equal to 9. Therefore, by the choice  $h = r/n \leq 1$  (we assumed that  $n \geq r$ ), (2.32) becomes

$$\sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{E}_P \left[ |P \Delta \hat{P}_n| \right] \geq \frac{cr}{n},$$

for some universal positive constant  $c$ , and Theorem 2.10 is proved.  $\blacksquare$

### Proof of Theorem 2.13

For the proof of Theorem 2.13, we first refer to lemmata 3 and 4, in Section 2.5.1.

The proof of Theorem 2.13 is inspired by that of Theorem 1 in [KST95], which derives an upper bound of the risk of a convex hull type estimator of a convex function. Let  $K \in \mathcal{K}_d$ . Let  $E$  be an ellipsoid which satisfies the properties of Lemma 3, and  $T$  an affine transform in  $\mathbb{R}^d$  which maps  $E$  to the unit ball  $B_2^d$ . Note that  $\beta_d = |\det T| |E|$ , so  $T$  is invertible. Let us denote  $K' = T(K)$  and  $X'_i = T(X_i), i = 1, \dots, n$ . Let  $\hat{K}'_n$  be the convex hull of  $X'_1, \dots, X'_n$ . By the definition of  $T$ , the following properties hold :

- (i)  $K' \in \mathcal{K}_d^1$ ,
- (ii)  $X'_1, \dots, X'_n$  are i.i.d. uniformly distributed in  $K'$ ,
- (iii)  $T(\hat{K}_n) = \hat{K}'_n$ .

Furthermore, one has the following:

$$\frac{|K \setminus \hat{K}_n|}{|K|} = \frac{|K' \setminus \hat{K}'_n|}{|\det T| |K|} = |K' \setminus \hat{K}'_n| \frac{|E|}{\beta_d |K|} \leq \frac{d^d}{\beta_d} |K' \setminus \hat{K}'_n|. \quad (2.33)$$

Let  $\delta = n^{-2/(d+1)}$ , and  $\{G_1, \dots, G_{N_\delta}\}$  be a  $\delta$ -net of  $\mathcal{K}_d^{(1)}$  with respect to the Hausdorff distance  $d_H$ , where  $N_\delta \leq \tau_1 \exp\left(\tau_2 \delta^{-\frac{d-1}{2}}\right)$ , cf. Lemma 5. The definition of a  $\delta$ -net is also given in Lemma 5. Let  $j^*, \hat{j} \in \{1, \dots, N_\delta\}$  be such that:

$$d_H(K', G_{j^*}) \leq \delta \quad \text{and} \quad d_H(\hat{K}'_n, G_{\hat{j}}) \leq \delta.$$

Let  $\varepsilon > 0$ . By (2.33) and (ii),

$$\mathbb{P}_K \left[ \frac{|K \setminus \hat{K}_n|}{|K|} > \varepsilon \right] \leq \mathbb{P}_{K'} \left[ |K' \setminus \hat{K}'_n| > \frac{\beta_d}{d^d} \varepsilon \right]. \quad (2.34)$$

Let us recall that if  $G, G'$  and  $G''$  are three Borel subsets of  $\mathbb{R}^d$ , then the following triangle inequality holds:

$$|G \setminus G''| \leq |G \setminus G'| + |G' \setminus G''|. \quad (2.35)$$

Thus,  $|K' \setminus \hat{K}'_n| \leq |K' \setminus G_{j^*}| + |G_{j^*} \setminus G_{\hat{j}}| + |G_{\hat{j}} \setminus \hat{K}'_n|$  and, by the definition of  $j^*$  and  $\hat{j}$ , and by Lemma 4 and (2.34),

$$\mathbb{P}_K \left[ \frac{|K \setminus \hat{K}_n|}{|K|} > \varepsilon \right] \leq \mathbb{P}_{K'} \left[ |G_{j^*} \setminus G_{\hat{j}}| > \frac{\beta_d}{d^d} \varepsilon - 2\alpha_1 \delta \right]. \quad (2.36)$$

Set  $\varepsilon' = \frac{\beta_d}{d^d} \varepsilon - 2\alpha_1 \delta$ . (2.36) implies

$$\mathbb{P}_K \left[ \frac{|K \setminus \hat{K}_n|}{|K|} > \varepsilon \right] \leq \sum_{j=1, \dots, N_\delta: |G_{j^*} \setminus G_j| > \varepsilon'} \mathbb{P}_{K'} \left[ \hat{j} = j \right]. \quad (2.37)$$

Let  $j \in \{1, \dots, N_\delta\}$  be fixed, such that  $|G_{j^*} \setminus G_j| > \varepsilon'$ . Recall that  $\hat{K}'_n \subseteq G_{\hat{j}}^\delta$ , and thus if  $\hat{j} = j$ , then  $X'_i \in G_j^\delta, i = 1, \dots, n$ . So,

$$\begin{aligned} \mathbb{P}_{K'} \left[ \hat{j} = j \right] &\leq \left( \mathbb{P}_{K'} \left[ X'_1 \in G_j^\delta \right] \right)^n \\ &\leq \left( 1 - \frac{|K' \setminus G_j^\delta|}{|K'|} \right)^n \\ &\leq \left( 1 - \frac{1}{\beta_d} (|G_{j^*} \setminus G_j| - |G_{j^*} \setminus K'| - |G_j^\delta \setminus G_j|) \right)^n, \end{aligned}$$

using the triangle inequality (2.35) and the fact that  $|K'| \leq \beta_d$ . Denote by  $I_{\varepsilon'} = 1$  if  $\varepsilon' < \beta_d$ , and 0 otherwise. Continuing (2.37), and using (2.41), one gets:

$$\begin{aligned}
\mathbb{P}_K \left[ \frac{|K \setminus \hat{K}_n|}{|K|} > \varepsilon \right] &\leq \sum_{j=1, \dots, N_\delta: |G_j^* \setminus G_j| > \varepsilon'} \left( 1 - \frac{\varepsilon'}{\beta_d} + \frac{(\alpha_1 + \alpha_2)\delta}{\beta_d} \right)^n \\
&\leq N_\delta \left( 1 - \frac{\varepsilon'}{\beta_d} + \frac{(\alpha_1 + \alpha_2)\delta}{\beta_d} \right)^n I_{\varepsilon'} \\
&\leq C_1 \exp \left( \delta^{-\frac{d-1}{2}} - \frac{\varepsilon n}{d^d} + \frac{(3\alpha_1 + \alpha_2)\delta n}{\beta_d} \right) \\
&\leq C_1 \exp \left( \alpha_3 \delta n - \frac{\varepsilon n}{d^d} \right),
\end{aligned}$$

where  $\alpha_3 = 1 + \frac{3\alpha_1 + \alpha_2}{\beta_d}$  is a positive constant which depends on  $d$  only (recall that  $\delta^{-\frac{d-1}{2}} = \delta n$ ). Finally, by choosing  $\varepsilon = \alpha_3 d^d \delta + x/n$ , for any  $x > 0$ , and by setting the constant  $c_2 = \alpha_3 d^d$ , one gets Theorem 2.13.  $\blacksquare$

## 2.5 Appendix to Chapter 2

### 2.5.1 Lemmata

**Lemma 3.** *Let  $G \in \mathcal{K}_d$ . There exists an ellipsoid  $E$  in  $\mathbb{R}^d$  such that  $G \subseteq E$  and  $|E| \leq d^d |G|$ .*

Proof of Lemma 3 can be found in [Lei59] and [HS07].

Next lemma is based on the Steiner formula for convex bodies. It shows that on  $\mathcal{K}_d^{(1)}$ , the Nikodym distance is bounded from above by the Hausdorff distance, up to a positive constant.

**Lemma 4.** *There exists some positive constant  $\alpha_1$  which depends on  $d$  only, such that*

$$|G \Delta G'| \leq \alpha_1 d_H(G, G'), \forall G, G' \in \mathcal{K}_d^{(1)}.$$

The following lemma is due to Bronshtein [Bro76], and gives an upper bound for the metric entropy of the class  $\mathcal{K}_d^{(1)}$ , with respect to the Nikodym distance. The original result is given with respect to the Hausdorff distance, but using Lemma 4, an analogous version of this result could be given with respect to the Nikodym distance. A  $\delta$ -net ( $\delta > 0$ ) of a subclass  $\mathcal{C}$  of  $\mathcal{K}_d^{(1)}$ , with respect to the Hausdorff distance, is a collection  $\mathcal{G}_\delta$  of sets of  $\mathcal{K}_d^{(1)}$ , such that for each  $G \in \mathcal{C}$ , there exists  $G^* \in \mathcal{G}_\delta$  satisfying  $d_H(G, G^*) \leq \delta$ .

**Lemma 5** ([Bro76]). *Let  $\delta > 0$ . There exists a finite  $\delta$ -net of  $\mathcal{K}_d^{(1)}$ , with respect to  $d_H$ , of cardinality  $N_\delta \leq \tau_1 \exp \left( \tau_2 \delta^{-\frac{d-1}{2}} \right)$ , for some positive constants  $\tau_1$  and  $\tau_2$ , which depend on  $d$  only.*

The value of the constants  $\tau_1$  and  $\tau_2$  can be found by investigating the proof of this lemma,

in [Bro76]. Another result on the metric entropy of  $\mathcal{K}_d^{(1)}$  was obtained by Dudley [Dud74], but in a weaker form than Bronshtein's upper bound, and could not be used in our work.

**Lemma 6.** *Let  $r \geq d+1, n \geq 2$ . There exists a positive constant  $K_1$ , which depends on  $d$  only, such that for any convex polytope  $P$  in  $\mathcal{P}_r^{(1)}$  there is a convex polytope  $P^* \in \mathcal{P}_{r,n}^{(1)}$  such that :*

$$\begin{cases} |P^* \triangle P| \leq \frac{K_1}{n} \\ P^* \subseteq P^{\sqrt{d}/n}, \quad P \subseteq (P^*)^{\sqrt{d}/n}. \end{cases}$$

**Lemma 7** ([Tsy09, Section 2.4]). *Let  $\mathbb{P}$  and  $\mathbb{Q}$  are two probability measures on the same probability space. The Hellinger distance satisfies the following properties.*

- *It is a distance between probability measures on a common probability space.*

- $1 - \frac{H(\mathbb{P}^{\otimes n}, \mathbb{Q}^{\otimes n})^2}{2} = \left(1 - \frac{H(\mathbb{P}, \mathbb{Q})^2}{2}\right)^n.$

- *Le Cam inequality:*

$$\int \min(d\mathbb{P}, d\mathbb{Q}) \geq \frac{1}{2} \left(1 - \frac{H(\mathbb{P}, \mathbb{Q})^2}{2}\right)^2,$$

where we denoted by  $\int \min(d\mathbb{P}, d\mathbb{Q}) = \int \min(p, q) d\nu$ , where  $p$  and  $q$  are the respective densities of  $\mathbb{P}$  and  $\mathbb{Q}$  with respect to a common  $\sigma$ -finite dominating measure  $\nu$ .

## 2.5.2 Proofs of Lemmata and others

**Proof of Efron's identity** (2.1) Let  $X_1, \dots, X_n$  be  $n$  i.i.d. random variables, uniformly distributed on a convex body  $K$ . Let  $\hat{K}_n = CH(X_1, \dots, X_n)$ .  $\mathbb{P}_K$ -almost surely,  $\hat{K}_n \subseteq K$ , so

$$\begin{aligned} \mathbb{E}_K^{\otimes n} [|\hat{K}_n \triangle K|] &= \mathbb{E}_K^{\otimes n} [|\hat{K}_n \setminus K|] \\ &= \mathbb{E}_K^{\otimes n} \left[ \int_K I(x \notin \hat{K}_n) dx \right] \\ &= |K| \mathbb{E}_K^{\otimes n} \left[ \frac{1}{|K|} \int_K I(x \notin \hat{K}_n) dx \right] \\ &= |K| \mathbb{E}_K^{\otimes n} \left[ \mathbb{P}_K[X \notin \hat{K}_n | X_1, \dots, X_n] \right], \end{aligned} \tag{2.38}$$

where  $X$  is a random variable with the same distribution as  $X_1$ , and independent of the sample  $X_1, \dots, X_n$ , and  $\mathbb{P}_K[\cdot | X_1, \dots, X_n]$  denotes the conditional distribution given  $X_1, \dots, X_n$ . In what follows, we set  $X_{n+1} = X$ , so that we can consider the bigger sample  $X_1, \dots, X_{n+1}$ . For  $i = 1, \dots, n+1$ , we denote by  $\hat{K}^{-i}$  the convex hull of the sample  $X_1, \dots, X_{n+1}$  from which the  $i$ -th variable  $X_i$  is withdrawn. Then  $\hat{K}_n = \hat{K}^{-(n+1)}$ , and by continuing (2.38), and by using the

symmetry of the sample,

$$\begin{aligned}
\mathbb{E}_K^{\otimes n}[|\hat{K}_n \triangle K|] &= |K| \mathbb{P}_K^{\otimes n+1}[X_{n+1} \notin \hat{K}^{-(n+1)}] \\
&= \frac{|K|}{n+1} \sum_{i=1}^{n+1} \mathbb{P}_K^{\otimes n+1}[X_i \notin \hat{K}^{-(i)}] \\
&= \frac{|K|}{n+1} \sum_{i=1}^{n+1} \mathbb{P}_K^{\otimes n+1}[X_i \in V(\hat{K}_{n+1})], \tag{2.39}
\end{aligned}$$

where  $V(\hat{K}_{n+1})$  is the set of vertices of  $\hat{K}_{n+1} = CH(X_1, \dots, X_{n+1})$ . Indeed, with probability one, the point  $X_i$  is not in the convex hull of the  $n$  other points if and only if it is a vertex of the convex hull of the whole sample. By rewriting the probability of an event as the expectation of its indicator function, one gets from (2.39),

$$\begin{aligned}
\mathbb{E}_K^{\otimes n}[|\hat{K}_n \triangle K|] &= \frac{|K|}{n+1} \sum_{i=1}^{n+1} \mathbb{E}_K^{\otimes n+1}[I(X_i \in V(\hat{K}_{n+1}))] \\
&= \frac{|K|}{n+1} \mathbb{E}_K^{\otimes n+1} \left[ \sum_{i=1}^{n+1} I(X_i \in V(\hat{K}_{n+1})) \right] \\
&= \frac{|K| \mathbb{E}_K^{\otimes n+1}[V_{n+1}]}{n+1},
\end{aligned}$$

where  $V_{n+1}$  denotes the cardinality of  $V(\hat{K}_{n+1})$ , i.e. the number of vertices of the convex hull  $\hat{K}_{n+1}$ . Efron's equality is then proved.  $\square$

**Proof of Lemma 4** Let  $G \in \mathcal{K}_d$ . Steiner formula (see Section 4.1 in [Sch93a]) states that there exist some positive numbers  $L_1(G), \dots, L_d(G)$ , such that

$$|G^\lambda \setminus G| = \sum_{j=1}^d L_j(G) \lambda^j, \lambda \geq 0. \tag{2.40}$$

Besides, the  $L_j(G), j = 1, \dots, d$  are increasing functions of  $G$ . In particular, if  $G \in \mathcal{K}_d^{(1)}$ , then  $L_j(G) \leq L_j(B_2^d)$ .

Let  $G, G' \in \mathcal{K}_d^{(1)}$ , and let  $\lambda = d_H(G, G')$ . Since  $G$  and  $G'$  are included in the unit ball,  $\lambda$  is not greater than its diameter, so  $\lambda \leq 2$ . By definition of the Hausdorff distance,  $G \subseteq G'^\lambda$  and  $G' \subseteq G^\lambda$ . Hence,

$$|G \triangle G'| = |G \setminus G'| + |G' \setminus G| \leq |G'^\lambda \setminus G'| + |G^\lambda \setminus G|$$



$$\leq 2 \sum_{j=1}^d L_j(B_2^d) \lambda^j \leq \lambda \sum_{j=1}^d L_j(B_2^d) 2^j.$$

Lemma 4 is proved by setting  $\alpha_1 = \sum_{j=1}^d L_j(B_2^d) 2^j$ .

Note that since  $\delta \leq 1$ , Steiner formula (2.40) implies, for  $G \in \mathcal{K}_d^{(1)}$ , that

$$|G^\delta \setminus G| \leq \alpha_2 \delta, \quad (2.41)$$

where  $\alpha_2 = \sum_{j=1}^d L_j(B_2^d)$ . □

**Proof of Lemma 6** Let  $r \geq d+1$ ,  $n \geq 2$  and  $P \in \mathcal{P}_r$ . The convex polytope  $P^*$  is constructed as follows. For any vertex  $x$  of  $P$ , let  $x^*$  be the closest point to  $x$  in  $[0, 1]^d$  with coordinates that are integer multiples of  $\frac{1}{n}$  (if there are several such points  $x^*$ , take any of them). The euclidean distance between  $x$  and  $x^*$  is bounded by  $\frac{\sqrt{d}}{n}$ .

Let us define  $P^*$  as the convex hull of all these resulting  $x^*$ . Then  $P^* \in \mathcal{P}_{r,n}^{(1)}$ .

It is clear that the Hausdorff distance between  $P$  and  $P^*$  is less than  $\frac{\sqrt{d}}{n}$ . Therefore if we denote  $\epsilon = \frac{\sqrt{d}}{n}$  we have  $P^* \subseteq P^\epsilon$  and  $P \subseteq (P^*)^\epsilon$ .

Using Lemma 4,  $|P^* \triangle P| \leq \alpha_1 d_H(P^*, P) \leq \frac{\alpha_1 \sqrt{d}}{n}$ . Lemma 6 is proved, by taking  $K_1 = \alpha_1 \sqrt{d}$ . □

**Proof of Lemma 1** For  $t > 0$ , let

$$C(t) = \bigcap_{p \in S} B_2^d(p, t).$$

When  $t$  is sufficiently large, the  $C(t)$  is nonempty. If  $t$  is too small,  $C(t) = \emptyset$ . In addition, note that  $C(t)$  is strictly increasing with  $t$ , i.e.,  $C(t) \subsetneq C(t'), \forall t < t'$ . Let  $t^* = \inf \{t \geq 0 : C(t) \neq \emptyset\}$ . Write  $C(t^*)$  as

$$C(t^*) = \bigcap_{k \in \mathbb{N}^*} C(t^* + 1/k).$$

For  $k \in \mathbb{N}^*$ , the set  $C(t^* + 1/k)$  is nonempty by the definition of  $t^*$ , and it is compact, as a finite intersection of compact sets. Then,  $C(t^*) \neq \emptyset$ , as the intersection of a decreasing sequence of nonempty compact sets. On the other hand, note that  $C(t) = \emptyset$  for  $t < t^*$ . Let  $c \in C(t^*)$ . Let us show that  $B_2^d(c, t^*)$  is an enclosing ball of  $S$  of minimal volume. Assume there exists a closed ball  $B$  of radius  $t < t^*$ , such that  $S \subseteq B$ . Let  $a$  be the center of  $B$ . Then  $\rho(a, p) \leq t, \forall p \in S$ . This means that  $a \in \bigcap_{p \in S} B_2^d(p, t) = C(t)$ . This is impossible, since  $C(t) = \emptyset$ ,  $t$  being strictly

smaller than  $t^*$ .

The uniqueness of the smallest enclosing ball has been already proved. Note that it implies, in particular, that the set  $C(t^*)$  contains exactly one point.  $\square$

**Proof of Lemma 2** Let  $x \in \mathbb{R}^d \setminus S$ , and assume that  $x \notin \mathcal{B}(S)$ .

Let  $r_1$  be the radius of  $\mathcal{B}(S)$ , and  $r_2$  that of  $\mathcal{B}(S \cup \{x\})$ . It is clear that  $r_1 \leq r_2$ . Note also that  $r_1 < r_2$ . Otherwise, if  $r_1$  were equal to  $r_2$ , then by the uniqueness of  $\mathcal{B}(S)$ , this would mean that  $\mathcal{B}(S) = \mathcal{B}(S \cup \{x\})$ . This is impossible, since we have assumed that  $x \notin \mathcal{B}(S)$ , although  $x \in \mathcal{B}(S \cup \{x\})$ . Thus,  $r_1 < r_2$ .

Let  $c_1$  and  $c_2$  be the respective centers of  $\mathcal{B}(S)$  and  $\mathcal{B}(S \cup \{x\})$ . We need to show that  $\rho(c_2, x) = r_2$ . Assume the opposite, i.e.,  $\rho(c_2, x) < r_2$ . Denote by  $C_1(t) = \bigcap_{p \in S} B_2^d(p, t)$  and

$C_2(t) = \bigcap_{p \in S \cup \{x\}} B_2^d(p, t)$ , for  $t \geq 0$ . Write

$$\begin{aligned} C_2(r_2) &= \left( \bigcap_{p \in S} B_2^d(p, r_2) \right) \cap B_2^d(x, r_2) \\ &= C_1(r_2) \cap B_2^d(x, r_2). \end{aligned} \tag{2.42}$$

Since  $\rho(c_2, x) < r_2$ ,  $B_2^d(c_2, \epsilon) \subseteq B_2^d(x, r_2)$ , for some small enough  $\epsilon > 0$ . On the other hand,  $C_2(t) \subseteq C_1(t)$ ,  $\forall t \geq 0$ . In particular,  $c_2 \in C_2(r_2) \subseteq C_1(r_2)$ . Since  $C_1(t)$  is strictly increasing with  $t$ , and  $C_1(r_1)$  contains one point,  $C_1(r_2)$  must contain at least one point  $c$  other than  $c_2$ . By the convexity of the set  $C_1(r_2)$  - which is an intersection of convex sets -, the whole segment  $[c, c_2]$  is contained in  $C_1(r_2)$ . Therefore, by (2.42),  $[c, c_2] \cap B_2^d(c_2, \epsilon) \subseteq C_2(r_2)$ . This is impossible, since  $C_2(r_2)$  must contain only one point, as proved in the proof of Lemma 1.  $\square$

## Chapter 3

# A nonparametric regression model

In this chapter, we are interested in statistical inference in the regression model (RM). As in the previous chapter, we investigate the polytopal case, with either known or unknown number of vertices, and the general case. In addition, we concentrate on the one-dimensional case, in order to understand what assumptions on the location, or on the size of the unknown set, yield an improvement on the minimax rate of convergence. Finally, still in the one-dimensional case, we compute the minimal size of the set so it can be detected, and we propose a decision rule for testing consistently whether it is nonempty.

Estimation of convex bodies in a regression model with multiplicative noise has been investigated by Mammen and Tsybakov [MT95]. In [MT95] Mammen and Tsybakov proposed an estimator of a convex set  $G$ , based on likelihood maximization over an  $\varepsilon$ -net, whose cardinality is bounded in terms of the metric entropy [Dud74]. They showed that if the design is i.i.d., uniformly distributed in  $[0, 1]^d$ , then the rate of their estimator is  $n^{-2/(d+1)}$ . In addition, they proved that this is the minimax rate on the class  $\mathcal{K}_d^{(1)}$  for this multiplicative model, and that if the design is arbitrary - not necessarily i.i.d. and uniformly distributed in  $[0, 1]^d$  -, this rate cannot be beaten by any estimator.

The regression model with additive noise (RM) has been studied in [KT92, KT93b], in the case where  $G$  belongs to a smooth class of boundary fragments and the errors are i.i.d. Gaussian variables with known variance. If  $\gamma$  is the smoothness parameter of the studied class, it is shown that the rate of the minimax risk on the class is  $n^{-\gamma/(\gamma+d-1)}$ . The case of convex boundary fragments is covered by the case  $\gamma = 2$ , which leads to the expected rate  $n^{-2/(d+1)}$  for the minimax risk. It is important to note that in these works the authors always assumed that the fragment, which is included in  $[0, 1]^d$ , has a boundary which is uniformly separated from 0 and 1. We do not make such an assumption in our work. Korostelev and Tsybakov [KT92, KT93b]

also considered non-gaussian noises, making more general assumptions.

One problem has not been investigated yet: What is the minimax rate of convergence if one assumes that the unknown set  $G$  in Model (RM) is a convex polytope with a bounded number of vertices ? More generally, what is the minimax rate of convergence on parametric families of supports ? Parametric families are considered in the method proposed in [KT92, KT93b], where the true fragment is first approximated by an element of a parametric family of fragments, whose dimension is chosen afterwards according to the optimal bias-variance tradeoff. The estimator targets that parametric version of the fragment. Thus, a parametric approximation of the fragment  $G$  and not directly  $G$  itself is estimated. This idea will be exploited when we estimate convex bodies. We will first approximate a convex body by a convex polytope, and then estimate that polytope. This method, using polytopal approximation, will provide an explicit estimator but its risk will be shown to be suboptimal. This is why we will propose another method, which is rather classical, using the metric entropy, and yields an estimator which achieves the minimax rate of convergence. The parametric case of polytopes, with a given number of vertices, is interesting by itself, not only as a step for estimating general convex bodies. We will discuss why the minimax rate on such classes of polytopes is not, as one may expect,  $1/n$ , but it is altered by a logarithmic factor.

## 3.1 Estimation of polytopes

### 3.1.1 Upper bound

We denote by  $P_0$  the true polytope, i.e.  $G = P_0$  in (RM) and we assume that  $P_0 \in \mathcal{P}_r^{(1)}$ . Denote by  $\mathcal{P}_{r,n}^{(1)}$  the class of all the convex polytopes in  $[0, 1]^d$ , with at most  $r$  vertices, and with coordinates that are integer multiples of  $\frac{1}{n}$ . It is clear that the set  $\mathcal{P}_{r,n}^{(1)}$  is finite and its cardinality is less than  $(n+1)^{dr}$ .

We estimate  $P_0$  by a polytope in  $\mathcal{P}_{r,n}^{(1)}$  which minimizes the sum of squared errors

$$\mathcal{A}(P, \{(X_i, Y_i)\}_{i=1, \dots, n}) = \sum_{i=1}^n (1 - 2Y_i) I(X_i \in P). \quad (3.1)$$

In what follows, we will write  $\mathcal{A}(P)$  instead of  $\mathcal{A}(P, \{(X_i, Y_i)\}_{i=1, \dots, n})$  in order to simplify the notations. Note that if the noise variables  $\xi_i$  are Gaussian, then minimization of  $\mathcal{A}(P)$  is

equivalent to maximization of the likelihood. Consider the set estimator of  $P_0$  defined as

$$\hat{P}_n^{(r)} \in \operatorname{argmin}_{P \in \mathcal{P}_{r,n}^{(1)}} \mathcal{A}(P). \quad (3.2)$$

Note that since  $\mathcal{P}_{r,n}^{(1)}$  is finite, the estimator  $\hat{P}_n^{(r)}$  exists but is not necessarily unique.

The next theorem establishes an exponential deviation inequality for the estimator  $\hat{P}_n^{(r)}$ .

**Theorem 3.1** ([Bru13]). *Let  $r \geq d + 1$  be an integer, and  $n \geq 2$ . Consider Model (RM), with  $G = P$ , where  $P \in \mathcal{P}_r^{(1)}$ . Let Assumption A be satisfied. For the estimator  $\hat{P}_n^{(r)}$ , there exist two positive constants  $c_1$  and  $c_2$ , which depend on  $d$  and  $\sigma$  only, such that*

$$\sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{P}_P \left[ n \left( |\hat{P}_n^{(r)} \triangle P| - \frac{2dr \ln n}{c_2 n} \right) \geq x \right] \leq c_1 e^{-c_2 x}, \forall x > 0.$$

The explicit expressions of the constants  $c_1$  and  $c_2$  are given in the proof. From the deviation inequality of Theorem 3.1, one can easily derive that the risk of the estimator  $\hat{P}_n^{(r)}$  on the class  $\mathcal{P}_r^{(1)}$  is of the order  $\frac{\ln n}{n}$ . Indeed, we have the following result.

**Corollary 3.1.** *Let  $n \geq 2$ . Let the assumptions of Theorem 3.1 be satisfied. Then, for any positive number  $q$ , there exists a constant  $c_1(\sigma, d, q)$  which depends on  $\sigma, d$  and  $q$  such that*

$$\sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{E}_P \left[ |\hat{P}_n^{(r)} \triangle P|^q \right] \leq c_1(\sigma, d, q) \left( \frac{r \ln n}{n} \right)^q.$$

Note that the construction of  $\hat{P}_n^{(r)}$  does not require the knowledge of  $\sigma$ .

### 3.1.2 Lower bound

Corollary 3.1 provides an upper bound of the order  $\frac{\ln n}{n}$  for the risk of  $\hat{P}_n^{(r)}$ . The next result shows that if the noise is Gaussian, then  $\frac{\ln n}{n}$  is the minimax rate of convergence on the class  $\mathcal{P}_r^{(1)}$ .

**Theorem 3.2** ([Bru13]). *Let  $r \geq d + 1$  be an integer. Consider Model (RM) and assume that the errors  $\xi_i$  are zero-mean Gaussian random variables with variance  $\sigma^2 > 0$ . For any large enough  $n$ , we have the following lower bound.*

$$\inf_{\hat{P}} \sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{E}_P \left[ |\hat{P} \triangle P| \right] \geq \frac{\alpha^2 \sigma^2 \ln n}{n},$$

where  $\alpha = \frac{1}{2} - \frac{\ln 2}{2 \ln 3} \approx 0.29...$

Corollary 3.1 together with Theorem 3.1.2 give the following bound on the class  $\mathcal{P}_r^{(1)}$ , in the case of Gaussian noise with variance  $\sigma^2$ .

$$0 < \alpha^2 \sigma^2 \leq \frac{n}{\ln n} \mathcal{R}_n(\mathcal{P}_r^{(1)}) \leq c_1(\sigma, d, 1)r < \infty,$$

for  $n$  large enough and  $r \geq d + 1$ . Note that the lower bound does not depend on the number of vertices  $r$ . This is because we prove our lower bound for the class  $\mathcal{P}_{d+1}^{(1)}$  and we use that  $\mathcal{P}_r^{(1)} \supseteq \mathcal{P}_{d+1}^{(1)}$ , for  $r \geq d+1$ . The minimax rate of convergence on any of the classes  $\mathcal{P}_r^{(1)}$ ,  $r \geq d+1$ , is therefore of the order  $(\ln n)/n$ .

An inspection of the proofs shows that these results still hold for  $d = 1, r = 2$ ; namely, in Model (RM), the minimax risk for the estimation of segments in  $[0, 1]$  is of order  $(\ln n)/n$ .

## 3.2 Estimation of convex bodies

### A first estimator

Now we aim to estimate convex bodies, not necessarily polytopes. If  $G$  is a convex body in Model (RM), an idea is to approximate  $G$  by a convex polytope. For example one can select  $r$  points on the boundary of  $G$  and take their convex hull. This will give a convex polytope  $P_r$  with  $r$  vertices inscribed in  $G$ . In Section 3.1.1 we showed how to estimate such a  $r$ -vertex convex polytope as  $P_r$ . Thus, if  $P_r$  approximates well  $G$ , an estimator of  $P_r$ , for example  $\hat{P}_n^{(r)}$ , is a candidate to be a good estimator of  $G$ . The larger is  $r$ , the better  $P_r$  should approximate  $G$  with respect to the Nikodym distance. At the same time, when  $r$  increases the upper bound of Corollary 3.1 increases as well. Therefore  $r$  should be chosen according to the bias-variance tradeoff. The bias term, due to the approximation of  $G$  by  $P_r$ , will be of the order  $1/r^{2/(d-1)}$ , by Lemma 10. The variance term, will be of the order  $r(\ln n)/n$ , by Corollary 3.1. The bias-variance tradeoff should lead to the choice of  $r$  of the order  $(n/\ln n)^{\frac{d-1}{d+1}}$ .

**Theorem 3.3** ([Bru13]). *Let  $n \geq 2$ . Consider Model (RM) where  $G$  is any convex subset of  $[0, 1]^d$ . Set  $r = \left\lfloor \left( \frac{n}{\ln n} \right)^{\frac{d-1}{d+1}} \right\rfloor$ , and let  $\hat{P}_n^{(r)}$  the estimator defined in (3.2). Let Assumption A be satisfied. Then, there exist positive constants  $c_1, c_2$  and  $c_3$ , which depend on  $d$  and  $\sigma$  only, such that*

$$\sup_{G \in \mathcal{K}_d^{(1)}} \mathbb{P}_G \left[ n \left( |\hat{P}_n^{(r)} \triangle G| - \left( \frac{c_3 \ln n}{n} \right)^{2/(d+1)} \right) \geq x \right] \leq c_1 e^{-c_2 x}, \forall x > 0.$$

The constants  $c_1$  and  $c_2$  are the same as in Theorem 3.1, and  $c_3$  is given explicitly in the proof of the theorem. From Theorem 3.3 we get the next corollary.

**Corollary 3.2.** *Let the assumptions of Theorem 3.3 be satisfied. Then, for any positive number  $q$  there exists a positive constant  $c_2(\sigma, d, q)$  which depends on  $\sigma, d$  and  $q$  such that*

$$\sup_{G \in \mathcal{K}_d^{(1)}} \mathbb{E}_G \left[ |\hat{P}_n^{(r)} \triangle G|^q \right] \leq c_2(\sigma, d, q) \left( \frac{\ln n}{n} \right)^{\frac{2q}{d+1}}.$$

Note again that the construction of our estimator does not require the knowledge of  $\sigma$ .

Corollary 3.2 shows that the estimator given in Theorem 3.3 achieves the rate  $\left( \frac{\ln n}{n} \right)^{\frac{2}{d+1}}$ . This estimator has an advantage: it is computable and its definition comes from an intuitive geometrical argument, polytopal approximation of convex sets. However, as we will show next, there exists an estimator which achieves the same rate without the logarithmic factor. That estimator is based on the metric entropy of the class  $\mathcal{K}_d^{(1)}$ , and is mainly of theoretical interest. We recall that a  $\delta$ -net of a class  $\mathcal{C} \subseteq \mathcal{K}_d^{(1)}$ , with respect to the Nikodym distance, is a family  $\mathcal{G}_\delta$  of sets such that, for all  $G \in \mathcal{C}$ , there exists  $G^* \in \mathcal{G}_\delta$  satisfying  $|G \triangle G^*| \leq \delta$ . If there exists such a family, that is finite, it provides a discrete and finite version of the class  $\mathcal{C}$ , and the smaller that family can be chosen, the more simple the class  $\mathcal{C}$  is.

### Improvement of the upper bound

We propose an estimator whose construction is similar to [MT95], where the multiplicative model was considered. First of all, let us recall that Lemma 5, in Section 2.5.1, gives an upper bound of the cardinality of a  $\delta$ -net of  $\mathcal{K}_d^{(1)}$ , with respect to the Hausdorff distance, for all  $\delta > 0$ .

Let  $\delta = n^{-2/(d+1)}$ . Let  $\{G_1, \dots, G_N\}$  be a  $\delta$ -net of  $\mathcal{K}_d^{(1)}$  with respect to the Hausdorff distance  $d_H$ , with  $N \leq \tau_1 e^{\tau_2 \delta^{-(d-1)/2}}$ . Let  $G \in \mathcal{K}_d^{(1)}$  be the true set in Model (RM). We define the estimator  $\tilde{G}_n = G_{\hat{j}}$ , where  $\hat{j}$  minimizes the sum of squared errors, as in Section 3.1.1:

$$\hat{j} \in \operatorname{argmin}_{j=1, \dots, N} \mathcal{A}(G_j), \tag{3.3}$$

where  $\mathcal{A}$  is defined in (3.1). Note again that  $\hat{j}$  may not be unique. We have the following result.

**Theorem 3.4** ([Bru13]). *Let  $n \geq 1$ . Consider Model (RM) with  $G \in \mathcal{K}_d^{(1)}$ . Set  $\tilde{G}_n = G_{\hat{j}}$ , where  $\hat{j}$  is defined in (3.3). Let Assumption A be satisfied. Then, there exist a positive integer  $n_0(d)$  which depends on  $d$  only and positive constants  $c_0$  and  $c_2$ , which depend on  $d$  and  $\sigma$  only, such that*

$$\sup_{G \in \mathcal{K}_d^{(1)}} \mathbb{P}_G \left[ |\tilde{G}_n \triangle G| \geq c_0 n^{-2/(d+1)} + \frac{x}{n} \right] \leq \tau_1 e^{-c_2 x}, \forall x > 0.$$

Here,  $c_0 = \frac{\tilde{c}_1 \alpha_1 + \tau_2}{c_1}$ , where the constants  $\tilde{c}_1$ ,  $c_1$  and  $c_2$  are given in the proof of Theorem 3.1 and  $\alpha_1$ , which depends on  $d$  only, comes from Lemma 4 in Section 2.5.1. Note again that the construction of the estimator  $\tilde{G}_n$  does not require the knowledge of the noise level  $\sigma$ .

As for the estimator of the previous section, we derive from Theorem 3.4 an upper bound of the risk of the estimator  $\tilde{G}_n$ , and we have the following result.

**Corollary 3.3.** *Let the assumptions of Theorem 3.4 be satisfied. Then, for any positive number  $q$  there exists a positive constant  $c_3(\sigma, d, q)$  such that*

$$\sup_{G \in \mathcal{K}_d^{(1)}} \mathbb{E}_G \left[ |\tilde{G}_n \triangle G|^q \right] \leq c_3(\sigma, d, q) n^{-\frac{2q}{d+1}}.$$

### Lower bound

In this section we give a lower bound of the minimax risk on the class  $\mathcal{K}_d^{(1)}$ .

**Theorem 3.5** ([Bru13]). *Let  $n \geq 125$ . Consider the Model (RM) and assume that the errors  $\xi_i$  are zero-mean Gaussian random variables, with variance  $\sigma^2 > 0$ . There exist a positive constant  $c_4$  which depends only on the dimension  $d$  and on  $\sigma$ , such that for any estimator  $\hat{C}_n$ , based on  $n$  observations,*

$$\sup_{C \in \mathcal{K}_d^{(1)}} \mathbb{E}_C \left[ |C \triangle \hat{C}_n| \right] \geq c_4 n^{-2/(d+1)}.$$

The explicit form of the constant  $c_4$  can be found in the proof of the theorem.

From Theorem 3.5 and Corollary 3.3, one gets, for  $n \geq 125$  and in the case of Gaussian noise,

$$0 < c_4 \leq n^{\frac{2}{d+1}} \mathcal{R}_n(\mathcal{K}_d^{(1)}) \leq c_3(\sigma, d, q) < \infty,$$

and therefore the minimax risk on the class  $\mathcal{K}_d^{(1)}$  is of the order  $n^{-2/(d+1)}$ .

### 3.3 Adaptation and misspecification

In Section 3.1.1, we proposed an estimator that depends on the parameter  $r$ . A natural question is to find an estimator that is adaptive to  $r$ , i.e. that does not depend on  $r$ , but achieves the optimal rate on the class  $\mathcal{P}_r^{(1)}$ . The idea of the following comes from Lepski's method for adaptation (see [Lep91], or [Chi12], Section 1.5, for a nice overview). Assume that the true number of vertices, denoted by  $r^*$ , is unknown, but is bounded from above by a given integer  $R_n \geq d + 1$  that may depend on  $n$  and be arbitrarily large. Theorem 3.1 would provide the estimator  $\hat{P}_n^{(R_n)}$ , but it is clearly suboptimal if  $r^*$  is small and  $R_n$  is large. Indeed the rate of convergence of  $\hat{P}_n^{(R_n)}$  is  $\frac{R_n \ln n}{n}$ , although the rate  $\frac{r^* \ln n}{n}$  can be achieved according to Theorem



3.1, when  $r^*$  is known. The procedure that we propose selects an integer  $\hat{r}$  based on the observations, and the resulting estimator is  $\hat{P}_n^{(\hat{r})}$ .

Note that  $R_n$  should not be of order larger than  $(\ln n)^{-1}n^{\frac{d-1}{d+1}}$ , since for larger values of  $r$ , Corollaries 1 and 3 show that the estimation rate is better when one considers the class  $\mathcal{K}_d^{(1)}$  instead of  $\mathcal{P}_r^{(1)}$ . Let us denote, for  $r = d+1, \dots, R_n - 1$ ,  $\hat{Q}_n^{(r)} = \hat{P}_n^{(r)}$ , and  $\hat{Q}_n^{(R_n)} = \tilde{G}_n$ , the estimator defined in Section 3.2. Let  $c_a = \frac{1}{c_2} + \max\left(\frac{2d}{c_2}, c_0\right)$ , where the constants  $c_0$  and  $c_2$  are given in theorems 3.1 and 3.4 respectively. The explicit form of the constant  $c_a$  is known and can be deduced from the proofs of theorems 3.1 and 3.4. This constant depends on  $d$  and  $\sigma$ .

$$\hat{r} = \min \left\{ r \in \{d+1, \dots, R_n\} : |\hat{Q}_n^{(r)} \triangle \hat{Q}_n^{(r')}| \leq \frac{2c_a r' \ln n}{n}, \forall r' = r, \dots, R_n \right\}.$$

The integer  $\hat{r}$  is well defined ; indeed, the set in the brackets in the last display is not empty, since the inequality is satisfied for  $r = R_n$ .

The adaptive estimator is defined as  $\hat{P}_n^{adapt} = \hat{Q}_n^{(\hat{r})}$ . Note that the construction of  $\hat{P}_n^{adapt}$  requires the knowledge of  $\sigma$  through the definition of  $\hat{r}$  ; it depends on the constant  $c_2$  of Theorem 3.1, which depends itself on  $\sigma$ . Recall that  $\mathcal{P}_\infty^{(1)} = \mathcal{K}_d^{(1)}$ . We have the following theorem.

**Theorem 3.6** ([Bru13]). *Let  $n \geq 2$ . Let Assumption A be satisfied. Let  $R_n = \lfloor (\ln n)^{-1}n^{\frac{d-1}{d+1}} \rfloor$  and  $\phi_{n,r} = \min\left(\frac{r \ln n}{n}, n^{-\frac{2}{d+1}}\right)$ , for all integers  $r \geq d+1$  and  $r = \infty$ . There exists a positive constant  $c_5$  that depends on  $d$  and  $\sigma$  only, such that the adaptive estimator  $\hat{P}_n^{adapt}$  satisfies the following inequality.*

$$\sup_{d+1 \leq r \leq \infty} \sup_{P \in \mathcal{P}_r^{(1)}} \mathbb{E}_P \left[ \phi_{n,r}^{-1} |\hat{P}_n^{adapt} \triangle P| \right] \leq c_5.$$

Thus, we show that one and the same estimator  $\hat{P}_n^{adapt}$  attains the optimal rate simultaneously on all the classes  $\mathcal{P}_r^{(1)}$ ,  $d+1 \leq r$ , and on the class  $\mathcal{K}_d^{(1)}$  of all convex subsets of  $[0, 1]^d$ . The explicit form of the constant  $c_5$  can be easily derived from the proof of the theorem.

### 3.4 Discussion

Theorem 3.4 shows that the logarithmic factor in Corollary 3.3 can be dropped and that the minimax rate of convergence on the class  $\mathcal{K}_d^{(1)}$  is  $n^{-2/(d+1)}$ . However, Theorems 3.1 and 3.4 show that the logarithmic factor is significant in the case of convex polytopes. Let us try to understand what brings this logarithmic factor in one case and not in the other.

Let us first ask the following question: What makes the estimation of sets on a given class  $\mathcal{C} \subseteq \mathcal{K}_d^{(1)}$  difficult in Model (RM) ? First, it is the complexity of the class, which can

be expressed, for example, in terms of the metric entropy. We worked with this notion of complexity in Theorem 3.4, using  $\delta$ -nets. The second issue is how detectable the individual sets of the given class are, in the model. If the unknown set  $G$  is too small, then, with high probability, it contains no points of the design. Conditionally to this event, all the data have the same distribution and no information in the sample can be used in order to detect  $G$ . A set  $G$  has to be large enough in order to be detectable by some procedure. The threshold on the volume beyond which a subset cannot be detected by any procedure should give a lower bound for the rate of the minimax risk. This comparison between the minimax risk and the detection bound motivates Section 3.5. In [Jan87], Janson studied asymptotic properties of the maximal volume of holes with a given shape. A hole is a subset of  $[0, 1]^d$  that contains no point of the design  $(X_1, \dots, X_n)$ . Janson showed that with high probability, there are convex and polytopal holes that have a volume of order  $(\ln n)/n$ . This result suggests that a lower bound of the minimax risk in Theorem 3.5 should be of the order  $(\ln n)/n$ . Our lower bound is attained on the polytopes with very small volumes. We do not use the specific structure of these polytopes to derive the lower bound ; we only use the fact that some of them cannot be distinguished from the empty set, no matter what the shape of their boundary is, when we choose their volume of order no larger than  $\frac{\ln n}{n}$ . This shows that the rate  $1/n$ , which could be expected on the parametric class  $\mathcal{P}_r^{(1)}$ , is not the right minimax rate of convergence: the order  $(\ln n)/n$  is dominating. On the other hand, the proof of the lower bound of the order  $n^{-2/(d+1)}$  for general convex bodies uses only the structure and regularity of the boundaries ; we do not deal especially with small hypotheses. The order  $n^{-2/(d+1)}$  is much larger than the detectability bound  $(\ln n)/n$ , and therefore seems to determine the minimax risk on the class  $\mathcal{K}_d^{(1)}$ .

Let us add two remarks in this discussion. First, if  $d = 2$  it is easy to prove a better lower bound for the minimax risk on the class  $\mathcal{P}_r^{(1)}$ , for any integer  $r \geq 3$ , using the scheme of the proof of Theorem 3.5 in the case  $d = 2$ :

$$\mathcal{R}_n(\hat{P}_n^{(r)}; \mathcal{P}_r^{(1)}) \geq \max \left( \frac{\lambda_1 \ln n}{n}, \frac{\lambda_2 r}{n} \right), \quad (3.4)$$

for some positive constants  $\lambda_1$  and  $\lambda_2$ , which do not depend on  $r$ . It seems to us that this lower bound should remain true for any value of  $d$ .

In (3.4), if  $\ln n$  is larger than  $r$ , then the minimax risk is controlled from below by the rate  $\frac{\ln n}{n}$ . On the opposite, if the number of vertices of the unknown convex polytope can be arbitrarily large, then the dominating term in (3.4) is of the order  $\frac{r}{n}$ .

Our second remark is the following. Let  $\mu_0$  be a fixed positive number. If one considers the subclass  $\mathcal{P}_r^{(1)}(\mu_0) = \{P \in \mathcal{P}_r^{(1)} : |P| \geq \mu_0\}$ , then subsets of  $[0, 1]^d$  with too small volume are excluded. The proof of the lower bound of Theorem 3.2 is based on hypotheses testing, with very small hypotheses, i.e., candidates  $P$  of small volume, depending on  $n$ . Therefore, the construction used in the proof of Theorem 3.4 is no more valid and we expect the minimax rate of convergence on this class to be of the order  $r/n$ , i.e., without a logarithmic factor. This will be discussed later. In addition, after a discussion with Lazlo Györfi, we believe that  $1/n$  is an individual lower rate of convergence (1.6) on the class  $\mathcal{P}_r^{(1)}$ , although  $(\ln n)/n$  is not. The right rate of convergence, in that setup, seems to be  $1/n$ . This is not proved yet.

### 3.5 The one-dimensional case and the change point problem

Consider Model (RM), with dimension  $d$  equal to one. The unknown set  $G$  is a segment in  $[0, 1]$ . In this section, we no longer assume that the design points  $X_i$  are necessarily i.i.d. uniformly distributed in  $[0, 1]$ . Instead, we distinguish two types of design:

(DD) Deterministic, and regular design:  $X_i = i/n, i = 1, \dots, n$ ;

(RD) Random, uniform design: the variables  $X_i, i = 1, \dots, n$ , are i.i.d. uniform on  $[0, 1]$ .

In the sequel, the design will be denoted by  $\mathcal{X}$ . In this one-dimensional case, we will not only focus on estimation of  $G$ , but also on its detection. The motivation is mainly given in the discussion of Section 3.4. As we already said before, Model (RM) can be interpreted as a partial and noisy observation of an image,  $[0, 1]$ , in which there is an unknown object  $G$ . From this observation, one may like to determine whether it is true that there is an object - the unknown set  $G$  might be empty -, and/or to recover that object, i.e., to estimate  $G$ . Another framework for this model is the noisy observation of some signal, here  $\mathbb{1}(X_i \in G)$ , and one would like to determine if the observation comes from a pure noise, or if there actually is some signal, and/or to recover that signal.

We would like to understand under which assumptions detection is not an obstacle for estimating  $G$ . In particular, the two following assumptions will be of interest for us :

**Assumption 1.** *The set  $G$  is of the form  $[0, \theta]$ , for some unknown number  $0 \leq \theta \leq 1$ .*

**Assumption 2.**  $|G| \geq \mu$ .

Here and in the rest of the chapter,  $\mu \in (0, 1)$  is a given positive number.

Let us discuss these two assumptions in order to understand, intuitively, why they yield easier detection of  $G$ . The first one gives some information on the location of  $G$ . In other words, it tells that the set  $G$  starts from the left side of the frame. The second one tells that  $G$  is not too small, and thus should not be unnoticed by the statistician. This was discussed in Section 3.4. Model (RM), together with Assumption 1, is well known under the name of the change point problem. It can be rewritten as :

$$Y_i = \mathbf{1}(X_i \leq \theta) + \xi_i, i = 1, \dots, n, \quad (3.5)$$

for some number  $\theta \in [0, 1]$ . The change point problem was studied in [KT93b, Sec. 1.9], and a continuous-time version of this model is addressed in [Kor06]. The aim is to estimate the breakpoint  $\theta$ . In the continuous time version, Korostelev [Kor06] proposed a more general framework. Instead of the indicator function in the regression equation (3.5), he considered a function with a jump at the point  $\theta$ , and satisfying a Lipschitz condition both on the left and on the right sides of  $\theta$ . In these two works [Kor06, KT93b], the change point  $\theta$  is estimated with a precision, in expectation, of the order of  $1/n$ , if  $\theta$  is assumed to be separated from 0 and 1 :  $h \leq \theta \leq 1 - h$ , for some  $h \in (0, 1/2)$ . Ibragimov and Khasminskii [IK84], under, among others, the same assumption of separation from 0 and 1, also proposed a consistent estimator of the discontinuity point of a regression function, with precision of order  $1/n$  as well. As we have already mentioned, this separation hypothesis is common in this kind of estimation problems. For instance, it is made in [KT92] and [KT93b, Chap 3], where Tsybakov and Korostelev propose an estimator for boundary fragments. The authors study a similar model to Model (RM), where  $G$  is a boundary fragment instead of a segment. This is one possible generalization in higher dimensions of our problem. They assume that the true  $g$  belongs to some Hölder class, and is separated from 0 and 1:

$$h \leq g(x) \leq 1 - h, \forall x \in [0, 1]^{d-1},$$

for a given parameter  $h \in (0, 1/2)$ . In [KT92], the same authors estimate the support of a uniform density, assuming it is a boundary fragment. Again, they make the hypothesis that the underlying function  $g$  is separated from 0 and 1. In both models, they build a piecewise polynomial estimator of the function  $g$ , and prove that it is optimal in a minimax sense - see more details below -. Our concern is to understand whether this separation hypothesis is necessary, in the simple case of dimension 1.

In [CW13], the detection question is addressed in a slightly different framework. In Model (RM), we assume that the strength of the signal is given, equal to 1. It is also interesting to deal with the case of a signal of unknown strength, i.e.

$$Y_i = \delta \mathbf{1}(X_i \in G) + \xi_i, i = 1, \dots, n, \quad (3.6)$$

where  $\delta$  is a positive number. For the signal to be detectable, there should be a tradeoff between its length  $|G|$  and its strength  $\delta$ . Naturally, if  $\delta$  is small, then the set  $G$  should be big enough and conversely, if  $\delta$  is large, the set  $G$  is allowed to be small, so that the signal can be detected. Testing the presence of a signal, i.e. whether  $\delta = 0$  or not, is considered in [CW13]. This work is concerned with the power of two tests: the scan - or maximum - likelihood ratio, and the average likelihood ratio. The two tests are compared in two regimes: signals of small scales, i.e.  $|G| \xrightarrow{n \rightarrow \infty} 0$ , and signals of large scales, i.e.  $\liminf_{n \rightarrow \infty} |G| > 0$ . The design is (DD), and it is proved that if  $\delta \sqrt{n|G|} \geq \sqrt{2 \ln \frac{1}{|G|}} + b_n$ , for some sequence  $b_n$  such that  $b_n \rightarrow \infty$ , then there is a test with asymptotic power 1. Note that here,  $\delta \sqrt{|G|}$  is exactly the  $L^2$ -norm of the signal,  $\|\delta \mathbf{1}(\cdot \in G)\|_2 = \delta \sqrt{|G|}$ . In [LT00], signals of unknown shape but known smoothness are considered. Exact minimax separation rates, in terms of the  $L^\infty$ -norm of the signal, for distinguishing the null hypothesis, under which observations are pure noise, and the alternative one, under which there is a signal, are given. Detection is harder in that framework, because unlike in Model (RM) or (3.6), where the shape of the signal is known - it is piecewise constant -, only its smoothness is known, and the separation rates are larger than those of models (RM) and (3.6), in the sense that they allow less freedom for the size of the signal. However, this problem is different from ours, since we are concerned with the location of the signal, not the signal itself.

Model (RM) deals with change points in the mean of the observations, conditionally to the design. Under Assumption 1, there is only one change point, and under Assumption 2, there are two. A problem of interest, in time series analysis, is that of detecting change points in the mean, the covariance function, or other characteristics of the series. We refer to [SZ10] and the references therein. In [FMSnt], a sample of  $n$  independent observations  $Y_1, \dots, Y_n$  is given, and one assumes that  $Y_i$  admits a density  $f(\cdot, \vartheta(\frac{i}{n}))$  with respect to a given measure, for  $i = 1, \dots, n$ , where  $f$  belongs to an exponential parametric class of densities. The real valued function  $\vartheta$  is assumed to be piecewise constant on  $[0, 1]$ , with a finite number  $K$  of jumps, and  $K$  is not necessarily known. Under a similar condition on  $\delta$  and  $|G|$  to that of [CW13], it is

shown that at least one change point is consistently detectable: the proposed estimator of  $K$  is positive with probability that goes to 1, as  $n \rightarrow \infty$ . When  $f(\cdot, \theta)$  is the density of a Gaussian distribution with mean  $\theta$  and given variance  $\sigma^2 > 0$ , the problem was addressed in [Leb03]. This model includes models (RM) and (3.6), when the design is (DD). However, Assumption 2 is not considered in that work, and optimality of the estimator of the change point is not treated, in the case when it is assumed to be unique - which corresponds to our Assumption 1 - or when it is known that there are only two of them.

A test consists in deciding whether to reject or not a given hypothesis, called the null hypothesis, when it is compared to an alternative one. Let  $h \in (0, 1)$ . In the whole paper, we will consider the following null hypothesis:

$$H_0 : G = \emptyset,$$

and the alternative hypothesis:

$$H_1 : |G| \geq h.$$

Testing  $H_0$  against  $H_1$  is equivalent to deciding whether the signal is pure noise. A test is a random variable  $\tau_n$ , which is built from the data set, and whose possible values are 0 and 1. The decision associated to the test  $\tau_n$  is to reject  $H_0$  if and only if  $\tau_n = 1$ . We measure the performance of a test  $\tau_n$  on a class  $\mathcal{C}$  using

$$\gamma_n(\tau_n, \mathcal{C}) = \mathbb{P}_\emptyset [\tau_n = 1] + \sup_{G \in \mathcal{C}, |G| \geq h} \mathbb{P}_G [\tau_n = 0].$$

This quantity is the sum of the errors of the first and the second kinds of the test  $\tau_n$ . This way of measuring the performance gives the same importance to the two kinds of errors. We say that  $\tau_n$  is consistent on the class  $\mathcal{C}$  if and only if  $\gamma_n(\tau_n, \mathcal{C}) \rightarrow 0$ , when  $n \rightarrow \infty$ . Let us allow the number  $h$  to depend on  $n$ . We call the separation rate on the class  $\mathcal{C}$  any sequence of positive numbers  $r_n$  such that:

- if  $\frac{h}{r_n} \xrightarrow{n \rightarrow \infty} \infty$ , then there exists a consistent test on  $\mathcal{C}$ , and
- if  $\frac{h}{r_n} \xrightarrow{n \rightarrow \infty} 0$ , then no test is consistent test on  $\mathcal{C}$ .

With regard to both Assumptions 1 and 2, we focus on three different classes of sets, which are defined as bellow:

- $\mathcal{S} = \{[a, b] : 0 \leq a \leq b \leq 1\}$  is the class of all segments on  $[0, 1]$ ,

- $\mathcal{S}_0 = \{[0, \theta] : 0 \leq \theta \leq 1\}$  is the class of all segments on  $[0, 1]$ , satisfying Assumption 1,
- $\mathcal{S}(\mu) = \{G \in \mathcal{S} : |G| \geq \mu\}$  is the class of all segments on  $[0, 1]$ , satisfying Assumption 2.

Since the design and the noise are assumed to be independent, reordering the  $X_i$ 's does not modify the model. Indeed, there exists a reordering  $\{i_1, \dots, i_n\}$  of  $\{1, \dots, n\}$ , such that  $X_{i_1} \leq \dots \leq X_{i_n}$ . The random indexes  $i_1, \dots, i_n$  are independent of the noise, and therefore the new noise vector  $(\xi_{i_1}, \dots, \xi_{i_n})$  has the same distribution as  $(\xi_1, \dots, \xi_n)$ . Thus, we assume from now on that  $\mathcal{X}$  is the reordering of a preliminary design, and therefore  $X_1 \leq \dots \leq X_n$  almost surely, without loss of generality.

### 3.5.1 Detection of a segment

We consider the problem of testing the null hypothesis  $H_0 : G = \emptyset$ , against the alternative hypothesis  $H_1 : |G| \geq h$ . We find the minimal magnitude of  $h$ , as a function of  $n$ , so the null and the alternative hypothesis are sufficiently well separated, and there exists a consistent test. Intuitively, the *a priori* knowledge that the unknown set  $G$  belongs to the class  $\mathcal{S}_0$  gives useful information about the location of this set, and therefore it makes it detect easier. Actually, the following theorem confirms this intuition, by showing that the separation rate is smaller - by a logarithmic factor - for the subclass  $\mathcal{S}_0$  than for the whole class  $\mathcal{S}$ .

The idea, for the class  $\mathcal{S}_0$ , is the following. Under  $H_1$ ,  $[0, h] \subseteq G$ . Therefore, we check among those observations  $(X_i, Y_i)$  for which  $X_i \leq h$  if there is a sufficiently large number of  $Y_i$ 's that are large, e.g. larger than  $1/2$ . Let  $N = \max\{i = 1, \dots, n : X_i \leq h\} = \#(\mathcal{X} \cap [0, h])$ . Let  $S$  be the following test statistics:

$$S = \#\{i = 1, \dots, N : Y_i \leq \frac{1}{2}\}.$$

If the alternative hypothesis holds, i.e. if  $|G| \geq h$ , all the  $X_i, i = 1, \dots, N$  fall inside the set  $G$ , and the corresponding  $Y_i$  should not be too small. The test statistic  $S$  counts how many of these  $Y_i$ 's are suspiciously small. This is how we build the test  $T_n^0$ :

$$T_n^0 = \mathbb{1}(S \leq cN),$$

where  $c \in (\mathbb{P}[\xi_1 \leq -1/2], \mathbb{P}[\xi_1 \leq 1/2])$ , assuming this interval nonempty.

For the class  $\mathcal{S}$ , we propose a scan test, i.e., a procedure which scans the whole frame  $[0, 1]$  and seeks for a large enough quantity of successive observations for which  $Y_i$  is large. If  $G \in \mathcal{S}$ , let  $R(G) = \sum_{i=1}^n Y_i \mathbb{1}(X_i \in G) - \frac{\#(\mathcal{X} \cap G)}{2}$ , and  $R = \sup_{|G| \geq h} R(G)$ . Under the alternative hypothesis,  $R$  should be quite large, and we define the test  $T_n^1 = \mathbb{1}(R \geq 0)$ .

**Theorem 3.7** ([Bru14b]). *Let Model (RM) hold.*

1. *Assume that the design is (DD) or (RD), and that the noise satisfies:*

$$\mathbb{P}[\xi_1 \leq -1/2] < \mathbb{P}[\xi_1 \leq 1/2].$$

*Then, if  $nh \rightarrow \infty$ , the test  $T_n^0$  is consistent, i.e.  $\gamma_n(T_n^0, \mathcal{S}_0) \rightarrow 0$ . If, in addition, the noise is Gaussian, then a separation rate on the class  $\mathcal{S}_0$  is  $r_n = 1/n$ .*

2. *Assume that the design is (DD) or (RD). Then, if  $nh/\ln n \rightarrow \infty$ , the test  $T_n^1$  is consistent, i.e.  $\gamma_n(T_n^1, \mathcal{S}) \rightarrow 0$ . If, in addition, the noise is Gaussian, then a separation rate on the class  $\mathcal{S}$  is  $r_n = \ln(n)/n$ .*

Note that the construction of the test  $T_n^0$  requires the knowledge of the law of the noise.

In the next section, we show that the separation rates given in Theorem 3.7 are the minimax rates of convergence on the corresponding classes.

### 3.5.2 Estimation of a segment

#### Least square estimators

Let Model (RM) hold. For  $G' \in \mathcal{S}$ , let  $A_0(G') = \sum_{i=1}^n (Y_i - \mathbb{1}(X_i \in G'))^2$  be the sum of squared errors. A way to estimate  $G$  is to find a random set  $\hat{G}_n$  which minimizes  $A_0(G')$ , among all possible candidates  $G'$ . Note that minimizing  $A_0(G')$  is equivalent to maximizing

$$A(G') = \sum_{i=1}^n (2Y_i - 1)\mathbb{1}(X_i \in G'). \quad (3.7)$$

Denote by  $S = \{i = 1, \dots, n : X_i \in G\}$  and by  $S' = \{i = 1, \dots, n : X_i \in G'\}$ , for some  $G' \in \mathcal{S}$ . Denote by  $\#(\cdot)$  the cardinality, for finite sets. The criterion  $A(G')$  becomes, if denoted as a function of  $S'$ ,

$$\begin{aligned} A(S') &= \sum_{i \in S'} (2Y_i - 1) \\ &= \sum_{i \in S'} (2\mathbb{1}(X_i \in G) + 2\xi_i - 1) \\ &= 2\#(S \cap S') - \#S' + 2 \sum_{i \in S'} \xi_i, \end{aligned}$$

so,

$$A(S') - A(S) = -\#(S \triangle S') + 2 \left( \sum_{i \in S' \setminus S} \xi_i - \sum_{i \in S \setminus S'} \xi_i \right). \quad (3.8)$$



A subset  $S'$  of  $\{1, \dots, n\}$  is called convex if and only if it is of the form  $\{i, \dots, j\}$ , for some  $1 \leq i \leq j \leq n$ . It is clear that if a convex subset  $S'$  of  $\{1, \dots, n\}$  maximizes  $A(S') - A(S)$ , then by defining  $G' = [X_{\min S'}, X_{\max S'}]$ , the segment  $G'$  maximizes  $A(G')$  (cf. (3.7)).

### Estimation of a unique change point

Let Model (RM) hold, with design (DD). Assume that  $G$  belongs to  $\mathcal{S}_0$ . This is the change point problem. For some  $\theta \in [0, 1]$ ,  $G$  can be written as  $G = [0, \theta]$ . Let us make one preliminary remark. For any estimator  $\hat{G}_n$  of  $G$ , the random segment  $\tilde{G}_n = [0, \sup \hat{G}_n]$  performs better than  $\hat{G}_n$ , since  $|\tilde{G}_n \triangle G| \leq |\hat{G}_n \triangle G|$   $\mathbb{P}_G$ -almost surely. Therefore, it is sufficient to consider only estimators of the form  $\hat{G}_n = [0, \hat{\theta}_n]$ , where  $\hat{\theta}_n$  is a random variable. Then,  $|\hat{G}_n \triangle G| = |\hat{\theta}_n - \theta|$ , and the performance of the estimator  $\hat{G}_n$  of  $G$  is that of the estimator  $\hat{\theta}_n$  of  $\theta$ . Let us build a least square estimator (LSE) of  $\theta$ . For  $M = 1, \dots, n$ , let

$$\begin{aligned} F(M) &= A(\{1, \dots, M\}) \\ &= \sum_{i=1}^M (2Y_i - 1). \end{aligned}$$

Let  $\hat{M}_n \in \operatorname{ArgMax}_{M=1, \dots, n} F(M)$ , and  $\hat{\theta}_n = X_{\hat{M}_n}$ . The following theorem follows.

**Theorem 3.8** ([Bru14b]). *Let  $n \geq 1$ . Let Model (RM) hold, with design (DD). Let  $\hat{G}_n = [0, \hat{\theta}_n]$ . Then,*

$$\sup_{G \in \mathcal{S}_0} \mathbb{P}_G \left[ |\hat{G}_n \triangle G| \geq \frac{x}{n} \right] \leq c_0 e^{-x/(8\sigma^2)}, \forall x > 0,$$

where  $c_0$  is a positive constant which depends on  $\sigma$  only.

A simple application of Fubini's theorem leads to the following result.

**Corollary 3.4.** *Let the assumptions of Theorem 3.8 be satisfied. Then, for all  $q > 0$ , there exists a positive constant  $c(\sigma, q)$  which depends on  $\sigma$  and  $q$  only, such that*

$$\sup_{G \in \mathcal{S}_0} \mathbb{E}_G \left[ |\hat{G}_n \triangle G|^q \right] \leq \frac{c(\sigma, q)}{n^q}.$$

This corollary shows that the minimax risk on the class  $\mathcal{S}_0$  is bounded from above by  $1/n$ , up to multiplicative constants. Next theorem proves that up to a positive constant,  $1/n$  is also a lower bound of the minimax risk, if the noise is Gaussian.

**Theorem 3.9** ([Bru14b]). *Consider Model (RM), with design (DD), and assume the noise is*

Gaussian. Then for all integer  $n \geq 1$ ,

$$\mathcal{R}_n(\mathcal{S}_0) \geq \frac{1}{2n}.$$

Combining Theorems 3.8 and 3.9 yields

**Theorem 3.10** ([Bru14b]). *Consider Model (RM), with design (DD) and Gaussian noise. Then, the minimax risk on the class  $\mathcal{S}$  satisfies*

$$\frac{c_1}{n} \leq \mathcal{R}_n(\mathcal{S}_0) \leq \frac{c_2}{n}, \forall n \geq 1,$$

where the constants  $c_1$  and  $c_2$  depend on  $\sigma$ .

### Recovering any segment

Let us now assume that the unknown set  $G$  does not necessarily contain 0. We shall prove that whether to assume that  $G$  belongs to the class  $\mathcal{S}(\mu)$  or not does not lead to the same minimax rate. As we saw in Section 3.5.2, an estimator of  $G$  in Model (RM) can be obtained by maximizing the Gaussian process (3.8) over all segments of  $\{1, \dots, n\}$ . This is not the track that we will borrow, but it would be interesting to work precisely on this process. This would probably be the first step to extensions of our results in higher dimensions. However, this problem remains open for now. The methods that we develop in this section are quite different. If  $G$  is only assumed to belong to the biggest class  $\mathcal{S}$ , the proposed estimator is the LSE, which was already detailed in [Bru13] for convex polytopes, in higher dimension. If  $|G|$  is *a priori* known to be greater or equal to  $\mu$ , then we first build a preliminary estimator of  $G$  - the LSE -, using one half of the observed sample. This estimator is not optimal, but it is close to  $G$  with high probability. We show that the middle point  $\hat{m}_n$  of this estimator is in  $G$  with high probability. This brings us back to the change point problem, where 0 is now replaced by  $\hat{m}_n$ , and we use the second half of the observed sample to estimate two change points.

Let us first state the following theorem, which is, for the design (RD), a particular case of Theorem 3.1, for  $d = 1$ .

**Theorem 3.11** ([Bru14b]). *Let  $n \geq 2$ . Let Model (1.3) hold, with design (DD) or (RD). Let  $\hat{G}_n \in \underset{G' \in \mathcal{S}}{\text{ArgMax}} A(G')$  be a LSE estimator of  $G$ . Then, there exist two positive constants  $c_1$  and  $c_2$  which depend on  $\sigma$  only, such that*

$$\sup_{G \in \mathcal{S}} \mathbb{P}_G \left[ n \left( |\hat{G}_n \triangle G| - \frac{4 \ln n}{c_2 n} \right) \geq x \right] \leq c_1 e^{-c_2 x}, \forall x > 0.$$

The expressions of  $c_1$  and  $c_2$  are given in the proof of Theorem 3.1, for the design (RD). For the design (DD), we do not give a proof of this theorem here, but it can be easily adapted from that of the case of the design (RD). The next corollary is immediate.

**Corollary 3.5.** *Let the assumptions of Theorem 3.11 be satisfied. Then, for all  $q > 0$ , there exists a positive constant  $B_q$  which depends on  $q$  and  $\sigma$  only, such that*

$$\sup_{G \in \mathcal{S}} \mathbb{E}_G \left[ |\hat{G}_n \triangle G|^q \right] \leq B_q \left( \frac{\ln n}{n} \right)^q.$$

This corollary shows that the minimax risk on the class  $\mathcal{S}$  is bounded from above by  $\ln(n)/n$ , up to a multiplicative constant. The following theorem establishes a lower bound, if the noise is supposed to be Gaussian.

**Theorem 3.12** ([Bru14b]). *Consider Model (RM), with design (DD) or (RD). Assume that the noise terms  $\xi_i$  are i.i.d. Gaussian random variables, with variance  $\sigma^2 > 0$ . For any large enough  $n$ ,*

$$\mathcal{R}_n(\mathcal{S}) \geq \frac{\alpha^2 \sigma^2 \ln n}{n},$$

where  $\alpha$  is a universal positive constant.

This lower bound comes from [Bru13, Theorem 2] in the case of the design (RD), and the proof is easily adapted for the design (DD). Eventually, the minimax risk on the class  $\mathcal{S}$  is of the order  $\ln(n)/n$ :

**Theorem 3.13** ([Bru14b]). *Consider Model (RM), with design (DD) or (RD). Assume that the noise terms  $\xi_i$  are i.i.d. Gaussian random variables, with variance  $\sigma^2 > 0$ . The minimax risk on the class  $\mathcal{S}$  satisfies:*

$$\frac{c_2 \ln n}{n} \leq \mathcal{R}_n(\mathcal{S}) \leq \frac{c_1 \ln n}{n}, \forall n \geq n_0,$$

where  $n_0$  is a positive integer which depends on  $\sigma$ , as for the positive constants  $c_1$  and  $c_2$ .

### Recovering a segment in $\mathcal{S}(\mu)$

For the design (DD), we combine both Theorems 3.8 and 3.11 to find the minimax rate on the class  $\mathcal{S}(\mu)$ . Let Model (RM) hold, and let  $G \in \mathcal{S}(\mu)$ . First, we split the sample into two equal parts. Let  $\mathcal{D}_0$  be the set of sample points with even indexes, and  $\mathcal{D}_1$  the set of sample points with odd indexes. Note that  $\mathcal{D}_0 \cup \mathcal{D}_1$  is exactly the initial sample, that these two subsample are independent, and that each of them is made of at least  $(n-1)/2$  data points. Let  $\hat{G}_n$  be the LSE estimator of  $G$  given in Theorem 3.11, built from the subsample  $\mathcal{D}_0$ . Let  $\hat{m}_n$  be the

middle of  $\hat{G}_n$ . As it will be shown in the proof of the next theorem,  $\hat{m}_n$  satisfies both following properties, with high probability:

1.  $\hat{m}_n \in G$ ,
2.  $\mu/2 \leq \hat{m}_n \leq 1 - \mu/2$ .

This brings us to estimating two change points - the endpoints of  $G$  - , using the second subsample  $\mathcal{D}_1$ . From Theorem 3.8, we know that this can be done at the speed  $1/n$ , up to multiplicative constants.

**Theorem 3.14** ([Bru14b]). *Consider Model (RM), with design (DD). There exists an estimator  $\tilde{G}_n$  of  $G$ , such that*

$$\sup_{G \in \mathcal{S}(\mu)} \mathbb{P}_G \left[ |\tilde{G}_n \triangle G| \geq \frac{x}{n} \right] \leq 2c_0 e^{-\mu x / (256\sigma^2)} + c_1 n^4 e^{-c_2 \mu n / 2}, \forall x > 0,$$

for  $n$  large enough. The positive constants  $c_0$  and  $c_2$  appeared in Theorems 3.8 and 3.11.

Naturally, Theorem 3.14 leads to the next corollary.

**Corollary 3.6.** *Let the assumptions of Theorem 3.14 be satisfied. Then, for all  $q > 0$ , there exists a positive constant  $c(q, \mu, \sigma)$  which depends on  $q$ ,  $\mu$  and  $\sigma$  only, such that*

$$\sup_{G \in \mathcal{S}(\mu)} \mathbb{E}_G \left[ |\tilde{G}_n \triangle G|^q \right] \leq \frac{c(q, \mu, \sigma)}{n^q}, \forall n \geq 1.$$

This corollary, for  $q = 1$ , shows that the minimax risk on the class  $\mathcal{S}(\mu)$  is bounded from above by  $1/n$ , up to a multiplicative constant. A very similar proof to that of Theorem 3.9 yields a lower bound for this minimax risk, which yields next Theorem.

**Theorem 3.15** ([Bru14b]). *Consider Model (RM), with design (DD) and Gaussian noise. The minimax risk on the class  $\mathcal{S}(\mu)$  satisfies:*

$$\frac{c_1}{n} \leq \mathcal{R}_n(\mathcal{S}(\mu)) \leq \frac{c_2}{n}, \forall n \geq n_0,$$

for some positive integer  $n_0$  depending on  $\mu$  and  $\sigma$ , and positive constants  $c_1$  and  $c_2$  depending on  $\mu$  and  $\sigma$  as well.

**Remark 3.1.** *Note that, in Theorem 3.14, the upper bound contains one residual term which does not depend on  $x$ . This term, in order to be sufficiently small, requires that  $\mu$  - if allowed to depend on  $n$  - is of larger order than  $\ln(n)/n$ . This reminds Theorem 3.7, in which we showed that the smallest set which can be detected has measure of this order exactly. In addition, if  $\mu$  is of the order of  $\ln(n)/n$ , then the proof of the lower bound of Theorem 3.12 can be applied, and the minimax risk on the class  $\mathcal{S}(\mu)$  will be of the order  $\frac{\ln n}{n}$ .*

### 3.5.3 Conclusion and discussion

We summarize our results in Table 3.1. The rates that are written in this table hold for Gaussian noise, which is the most important case. In each case we indicate the design for which the rate has been proved.

	$\mathcal{S}$	$\mathcal{S}_0$	$\mathcal{S}(\mu)$
Minimax rate	$\ln(n)/n$ (DD,RD)	$1/n$ (DD)	$1/n$ (DD)
Separation rate	$\ln(n)/n$ (DD,RD)	$1/n$ (DD,RD)	$\cdot$

Table 3.1: Minimax risks and separation rates for the classes  $\mathcal{S}$ ,  $\mathcal{S}_0$  and  $\mathcal{S}(\mu)$ .

Note that in two cases, only the design (DD) has been considered. This is for technical reasons, and we believe that the rates are still the same for the design (RD).

It comes out that asymptotically, a segment can be estimated infinitely faster when it is *a priori* supposed either to contain a given point (e.g., 0), or to be large enough. The main question that remains is: does this phenomenon still hold for two - or higher - dimensional sets ? An important - if not essential - assumption which has been done is the convexity of the unknown set. In dimension 1, the class of convex subsets of  $[0, 1]$  is simple, and parametric. In dimension  $d \geq 2$ , the class  $\mathcal{K}_d^{(1)}$  is much more complex. In particular, its metric entropy is much larger than that of a parametric family [Bro76], and it seems to us that it is the complexity of this class that makes it harder to estimate a set, than detectability.

In Model (RM), if  $G$  belongs to  $\mathcal{K}_d^{(1)}$ , estimation of  $G$  can be done at the minimax rate  $n^{-2/(d+1)}$  [Bru13]. However, an adaptation of the proof of 3.2 shows the following: on any subclass of  $\mathcal{K}_d$  invariant under translations and invertible affine transformations - which keep a set  $G$  inside the frame  $[0, 1]^d$  -, the minimax rate is at least of order  $\ln(n)/n$ . In addition, we believe this is the separation rate for the detection problem, on any such subclass of  $\mathcal{K}_d^{(1)}$ .

For the whole class  $\mathcal{K}_d^{(1)}$ , since  $\ln(n)/n$  is much smaller than  $n^{-2/(d+1)}$ , the minimax rate is of the order of  $n^{-2/(d+1)}$ . However, for a parametric subclass, such as that of all convex polytopes with a given number of vertices, we believe that the complexity of the class leads to a term of order  $1/n$  in the minimax risk, which is dominated by the term  $\ln(n)/n$  that comes from detectability. In our opinion, this is what explains that, as shown in [Bru13], the minimax risk on the class of all convex polytopes of  $[0, 1]^d$ , with a given number of vertices, is of order  $\ln(n)/n$ .

Motivated by Theorem 3.14, we also conjecture that  $1/n$  is the minimax rate on the class of all convex polytopes of  $[0, 1]^d$ , of volume greater than a given  $\mu > 0$ . However, note that

it is not possible to extend Theorem 3.8 to higher dimensions. An adaptation of the proof of Theorem 3.2, by taking, as the  $M$  hypotheses used in the proof, sets which contain the origin and have pairwise zero measure intersections, would show that  $(\ln n)/n$  remains a lower bound under the assumption that  $0 \in G$ . Yet, we believe that if the unknown set is assumed to contain a given section of positive  $(d-1)$ -dimensional Lebesgue measure, of a given hyperplane in  $\mathbb{R}^d$ , e.g.  $[0, 1]^{d-1} \times \{0\}$  then an analog of Theorem 3.8 should hold, and the minimax rate should be of order  $1/n$ .

## 3.6 Proofs

### Proof of Theorem 3.1

Let  $P_0 \in \mathcal{P}_r$  be the true polytope. We have the following lemma, which is a direct consequence of Lemma 6, see Section 2.5.1.

**Lemma 8.** *Let  $r \geq d+1, n \geq 2$ . For any convex polytope  $P$  in  $\mathcal{P}_r^{(1)}$  there exists a convex polytope  $P^* \in \mathcal{P}_{r,n}^{(1)}$  such that*

$$|P^* \triangle P| \leq \frac{2d^{d+1}(3/2)^d \beta_d}{n}. \quad (3.9)$$

Let  $P^* \in \mathcal{P}_{r,n}^{(1)}$  such that  $|P^* \triangle P_0| \leq \frac{2d^{d+1}(3/2)^d \beta_d}{n}$ . Note that for all  $\epsilon > 0$ ,

$$\mathbb{P}_{P_0} \left[ |\hat{P}_n^{(r)} \triangle P_0| \geq \epsilon \right] = \mathbb{P}_{P_0} \left[ \exists P \in \mathcal{P}_{r,1}^{(n)} : \mathcal{A}(P) \leq \mathcal{A}(P^*), |P \triangle P_0| \geq \epsilon \right], \quad (3.10)$$

where  $P^*$  is a convex polytope chosen in  $\mathcal{P}_{r,n}^{(1)}$  satisfying  $|P^* \setminus P_0| \leq \frac{2d^{d+1}(3/2)^d \beta_d}{n}$ , cf. (3.9). For any  $P$  we have, by a simple algebra,

$$\mathcal{A}(P) - \mathcal{A}(P^*) = \sum_{i=1}^n Z_i, \quad (3.11)$$

where

$$Z_i = I(X_i \in P) - I(X_i \in P^*) - 2I(X_i \in P_0) [I(X_i \in P) - I(X_i \in P^*)] - 2\xi_i [I(X_i \in P) - I(X_i \in P^*)], \quad i = 1, \dots, n.$$

The random variables  $Z_i$  depend on  $P$  but we omit this dependence in the notation. Therefore, (3.10) implies that

$$\begin{aligned}
\mathbb{P}_{P_0} \left[ |\hat{P}_n^{(r)} \triangle P_0| \geq \epsilon \right] &\leq \sum_{P \in \mathcal{P}_{r,n}^{(1)}: |P \triangle P_0| \geq \epsilon} \mathbb{P}_{P_0} \left[ \sum_{i=1}^n Z_i \leq 0 \right] \\
&\leq \sum_{P \in \mathcal{P}_{r,n}^{(1)}: |P \triangle P_0| \geq \epsilon} \mathbb{E}_{P_0} \left[ \exp \left( -u \sum_{i=1}^n Z_i \right) \right], \tag{3.12}
\end{aligned}$$

for all positive number  $u$ , by Markov's inequality. Since  $Z_i$ 's are mutually independent, we obtain

$$\mathbb{P}_{P_0} \left[ |\hat{P}_n^{(r)} \triangle P_0| \geq \epsilon \right] \leq \sum_{P \in \mathcal{P}_{r,n}^{(1)}: |P \triangle P_0| \geq \epsilon} \prod_{i=1}^n \mathbb{E}_{P_0} [\exp(-uZ_i)]. \tag{3.13}$$

By conditioning on  $X_1$  and denoting by  $W = I(X_1 \in P) - I(X_1 \in P^*)$  we have

$$\begin{aligned}
\mathbb{E}_{P_0} [\exp(-uZ_1)] &= \mathbb{E}_{P_0} [\mathbb{E}_{P_0} [\exp(-uZ_1) | X_1]] \\
&= \mathbb{E}_{P_0} [\exp(-uW + 2uI(X_1 \in P_0)W) \mathbb{E}_{P_0} [\exp(2u\xi_1 W) | X_1]] \\
&= \mathbb{E}_{P_0} [\exp(-uW + 2uI(X_1 \in P_0)W) \exp(2\sigma^2 u^2 I(X_1 \in P \triangle P^*))] \\
&= \mathbb{E}_{P_0} [\exp(2\sigma^2 u^2 I(X_1 \in P \triangle P^*) - uW + 2uI(X_1 \in P_0)W)]. \tag{3.14}
\end{aligned}$$

We will now reduce the last expression in (3.14). It is convenient to use Table 3.2: the first three columns represent the values that can be taken by the binary variables  $I(X_1 \in P)$ ,  $I(X_1 \in P^*)$  and  $I(X_1 \in P_0)$  respectively, and the last column gives the resulting value of the term under the expectation in (3.14), that is  $\exp(2\sigma^2 u^2 I(X_1 \in P \triangle P^*) - uW + 2uI(X_1 \in P_0)W)$ .

$P$	$P^*$	$P_0$	Value
1	1	1	1
1	1	0	1
1	0	1	$\exp(2\sigma^2 u^2 + u)$
1	0	0	$\exp(2\sigma^2 u^2 - u)$
0	1	1	$\exp(2\sigma^2 u^2 - u)$
0	1	0	$\exp(2\sigma^2 u^2 + u)$
0	0	1	1
0	0	0	1

Table 3.2: Values of  $\exp(2\sigma^2 u^2 I(X_1 \in P \triangle P^*) - uW + 2uI(X_1 \in P_0)W)$

Hence one can write

$$\begin{aligned}\mathbb{E}_{P_0} [\exp(-uZ_1)] &= 1 - |P \triangle P^*| + e^{2\sigma^2 u^2 + u} (|(P \cap P_0) \setminus P^*| + |P^* \setminus (P \cup P_0)|) \\ &\quad + e^{2\sigma^2 u^2 - u} (|(P^* \cap P_0) \setminus P| + |P \setminus (P^* \cup P_0)|).\end{aligned}$$

Besides by the triangle inequality,

$$|P \triangle P_0| \leq |P \triangle P^*| + |P^* \triangle P_0|,$$

which implies

$$\begin{aligned}\mathbb{E}_{P_0} [\exp(-uZ_1)] &\leq 1 - |P \triangle P_0| + |P^* \triangle P_0| + e^{2\sigma^2 u^2 + u} (|P_0 \setminus P^*| + |P^* \setminus P_0|) \\ &\quad + e^{2\sigma^2 u^2 - u} (|P_0 \setminus P| + |P \setminus P_0|) \\ &\leq 1 - |P \triangle P_0| + |P^* \triangle P_0| + e^{2\sigma^2 u^2 + u} |P^* \triangle P_0| + e^{2\sigma^2 u^2 - u} |P \triangle P_0| \\ &\leq 1 - |P \triangle P_0| \left(1 - e^{2\sigma^2 u^2 - u}\right) + \frac{2d^{d+1}(3/2)^d \beta_d}{n} \left(1 + e^{2\sigma^2 u^2 + u}\right).\end{aligned}\tag{3.15}$$

Choose  $u = \frac{1}{4\sigma^2}$ . Then the quantity  $1 - e^{2\sigma^2 u^2 - u}$  is positive and if  $|P \triangle P_0| \geq \epsilon$ , then

$$\mathbb{E}_{P_0} [\exp(-uZ_1)] \leq 1 - \epsilon \left(1 - e^{-\frac{1}{4\sigma^2}}\right) + \frac{2d^{d+1}(3/2)^d \beta_d}{n} \left(1 + e^{\frac{3}{8\sigma^2}}\right).\tag{3.16}$$

We set  $\tilde{c}_1 = 1 + e^{\frac{3}{8\sigma^2}}$  and  $c_2 = 1 - e^{-\frac{1}{4\sigma^2}}$ . These are positive constants that do not depend on  $n$  or  $P_0$ . Recall that  $\mathcal{P}_{r,n}^{(1)}$  has cardinality less than  $(n+1)^{dr}$ . From (3.13) and (3.16), and by the independence of  $Z_i$ 's we have

$$\mathbb{P}_{P_0} \left[ |\hat{P}_n^{(r)} \triangle P_0| \geq \epsilon \right] \leq \sum_{P \in \mathcal{P}_{r,n}^{(1)} : |P \triangle P_0| \geq \epsilon} \left(1 - c_2 \epsilon + \frac{2d^{d+1}(3/2)^d \beta_d \tilde{c}_1}{n}\right)^n\tag{3.17}$$

$$\begin{aligned}&\leq (n+1)^{dr} \left(1 - c_2 \epsilon + \frac{2d^{d+1}(3/2)^d \beta_d \tilde{c}_1}{n}\right)^n \\ &\leq \exp \left( dr \ln(n+1) - c_2 \epsilon n + 2d^{d+1}(3/2)^d \beta_d \tilde{c}_1 \right) \\ &\leq \exp \left( 2dr \ln n - c_2 \epsilon n + 2d^{d+1}(3/2)^d \beta_d \tilde{c}_1 \right),\end{aligned}\tag{3.18}$$

where  $c_1 = \exp(2d^{d+1}(3/2)^d \beta_d \tilde{c}_1)$ , noting that  $n+1 \leq n^2$ . Therefore if we set  $\epsilon = \frac{2dr \ln n}{c_2 n} + \frac{x}{n}$  for a positive number  $x$ , we get the following deviation inequality

$$\mathbb{P}_{P_0} \left[ n \left( |\hat{P}_n^{(r)} \triangle P_0| - \frac{2dr \ln n}{c_2 n} \right) \geq x \right] \leq c_1 e^{-c_2 x}.$$



■

### Proof of Theorem 3.2

This proof is a simple application of Fano's method, see Corollary 2.6 in [Tsy09] or, for a more general setting, [Gun11]. Let  $M$  be a positive integer, and  $h = \frac{1}{M+1}$ . Let  $T_k, k = 0, \dots, M$  be  $M$  disjoint convex polytopes in  $\mathcal{P}_{d+1}^{(1)}$  and with same volume:  $|T_0| = \dots = |T_M| = h/2$ . Such a finite family of  $M+1$  disjoint convex polytopes can be constructed by dividing the hypercube  $[0, 1]^d$  into the  $M+1$  subsets  $[k/(M+1), (k+1)/(M+1)] \times [0, 1]^{d-1}$ , which have volume  $h = 1/(M+1)$ , and by constructing a convex polytope of  $\mathcal{P}_{d+1}$ , of volume  $h/2$ , in each of them.

For  $k = 1, \dots, M$ , we use the notation  $\mathbb{P}_k$  and  $\mathbb{E}_k$  instead of  $\mathbb{P}_{T_k}$  and  $\mathbb{E}_{T_k}$ , respectively, for simplicity's sake. A simple computation shows that the Kullback-Leibler divergence  $K(\mathbb{P}_k, \mathbb{P}_l)$  between  $\mathbb{P}_k$  and  $\mathbb{P}_l$ , for  $k \neq l$ , is equal to  $\frac{nh}{4\sigma^2}$ . On the other hand, the distance between  $T_k$  and  $T_l$ , for  $k \neq l$ , is  $|T_k \triangle T_l| = |T_k| + |T_l| = h$ . Then

$$\frac{1}{M+1} \sum_{j=1}^M K(\mathbb{P}_j, \mathbb{P}_0) = \frac{Mnh}{4(M+1)\sigma^2} \leq \frac{n}{4M\sigma^2}.$$

Let  $\alpha \in (0, 1)$ , and  $\gamma = \frac{1}{2\sigma^2\alpha}$ . Then, if  $M = \frac{\gamma n}{\ln n}$ , supposed without loss of generality to be an integer, we have

$$4\sigma^2\alpha M \ln M = 2n - 2n \frac{\ln \ln n}{\ln n} + 2n \frac{\ln \gamma}{\ln n} \geq n$$

for  $n$  large enough, so that

$$\frac{1}{M+1} \sum_{j=1}^M K(\mathbb{P}_j, \mathbb{P}_0) \leq \alpha \ln M.$$

Therefore, applying Corollary 2.6 in [Tsy09] with the Nikodym distance, we set, for  $r \geq d+1$ , the following inequality

$$\mathcal{R}_n(\mathcal{P}_r^{(1)}) \geq \frac{1}{M+1} \left( \frac{\ln(M+1) - \ln 2}{\ln M} - \alpha \right).$$

For  $n$  great enough, we have  $M \geq 3$  and  $\frac{\ln(M+1) - \ln 2}{\ln M} \geq 1 - \frac{\ln 2}{\ln 3}$ . We choose  $\alpha = \frac{1}{2} - \frac{\ln 2}{2\ln 3} \in (0, 1)$ .

So, we get

$$\mathcal{R}_n(\mathcal{P}_r^{(1)}) \geq \frac{\alpha}{M+1} \geq \frac{\alpha}{2M} \geq \frac{\alpha \ln n}{\gamma n} \geq \frac{\alpha^2 \sigma^2 \ln n}{n}.$$

This immediately implies Theorem 3.2. ■

### Proof of Theorem 3.3

The idea of the proof is very similar to that of Theorem 3.1. Here, we need to control an extra bias term, due to the approximation of  $G$  by a convex polytope. We first refer to Lemma 10 (cf. [GMR95]), for polytopal approximation of convex bodies: let  $r \geq d + 1$ , and  $P_r \in \mathcal{P}_r$  satisfying  $P_r \subseteq G$  and:

$$|G \triangle P_r| \leq cd \frac{|G|}{r^{2/(d-1)}}.$$

Note that since  $P_r \subseteq G$ ,  $P_r \in \mathcal{P}_r^{(1)}$ .

Let  $P^*$  be a polytope chosen in  $\mathcal{P}_{r,n}^{(1)}$  such that  $|P^* \triangle P_r| \leq \frac{(4d)^{d+1} \beta_d}{n}$ , like in the proof of Theorem 3.1. Thus by the triangle inequality,

$$|P^* \triangle G| \leq |P^* \triangle P_r| + |P_r \triangle G| \leq \frac{cd}{r^{2/(d-1)}} + \frac{(4d)^{d+1} \beta_d}{n}.$$

We now bound from above the probability  $\mathbb{P}_G \left[ |\hat{P}_n^{(r)} \triangle G| \geq \epsilon \right]$  for any  $\epsilon > 0$ . As in (3.10) and (3.12) we have

$$\begin{aligned} \mathbb{P}_G \left[ |\hat{P}_n^{(r)} \triangle G| \geq \epsilon \right] &\leq \mathbb{P}_G \left[ \exists P \in \mathcal{P}_{r,n}^{(1)}, \mathcal{A}(P) \leq \mathcal{A}(P^*), |P \triangle G| \geq \epsilon \right] \\ &\leq \sum_{P \in \mathcal{P}_{r,n}^{(1)}: |P \triangle G| \geq \epsilon} \mathbb{P}_G [\mathcal{A}(P) \leq \mathcal{A}(P^*)]. \end{aligned}$$

Repeating the argument in (3.11) with  $G$  instead of  $P_0$  we set

$$\mathcal{A}(P) - \mathcal{A}(P^*) = \sum_{i=1}^n Z_i,$$

where

$$\begin{aligned} Z_i &= I(X_i \in P) - I(X_i \in P^*) - 2I(X_i \in G) [I(X_i \in P) - I(X_i \in P^*)] \\ &\quad - 2\xi_i [I(X_i \in P) - I(X_i \in P^*)], \quad i = 1, \dots, n. \end{aligned}$$

The rest of the proof is very similar to the one of Theorem 3.1. Indeed, replacing  $P_0$  by  $G$  in that proof between (3.10) and (3.15), and  $\frac{2d^{d+1}(3/2)^d \beta_d}{n}$  by  $\frac{2d^{d+1}(3/2)^d \beta_d}{n} + \frac{cd}{r^{2/(d-1)}}$  in (3.16) and (3.18) one gets

$$\begin{aligned}
& \mathbb{P}_G \left[ |\hat{P}_n^{(r)} \triangle G| \geq \epsilon \right] \\
& \leq \sum_{P \in \mathcal{P}_{r,n}^{(1)} : |P \triangle G| \geq \epsilon} \left( 1 - c_2 \epsilon + \tilde{c}_1 \left( \frac{cd}{r^{2/(d-1)}} + \frac{2d^{d+1}(3/2)^d \beta_d}{n} \right) \right)^n \\
& \leq (n+1)^{dr} \left( 1 - c_2 \epsilon + \tilde{c}_1 \left( \frac{cd}{r^{2/(d-1)}} + \frac{2d^{d+1}(3/2)^d \beta_d}{n} \right) \right)^n \\
& \leq \exp \left( 2dr \ln n - c_2 \epsilon n + \tilde{c}_1 \left( \frac{cdn}{r^{2/(d-1)}} + 2d^{d+1}(3/2)^d \beta_d \right) \right).
\end{aligned}$$

Therefore if we set  $\epsilon = \frac{2dr \ln n}{c_2 n} + \frac{\tilde{c}_1 cd}{c_2 r^{2/(d-1)}} + \frac{x}{n}$  for a positive number  $x$ , we get the following deviation inequality

$$\mathbb{P}_G \left[ n \left( |\hat{P}_n^{(r)} \triangle G| - \frac{2dr \ln n}{c_2 n} - \frac{\tilde{c}_1 cd}{c_2 r^{2/(d-1)}} \right) \geq x \right] \leq c_1 e^{-c_2 x},$$

where the constants are defined as in the previous section. That ends the proof of Theorem 3.3, by choosing  $r = \lfloor \left( \frac{n}{\ln n} \right)^{\frac{d-1}{d+1}} \rfloor$ , and the constant  $c_3$  is given by

$$c_3 = (1 + \tilde{c}_1 c) \frac{d}{c_2} = \left( 1 + (1 + e^{3/(8\sigma^2)})c \right) \frac{d}{1 - e^{-1/(4\sigma^2)}}.$$

■

### Proof of Theorem 3.4

The proof is similar to the proof of Theorem 3.1. The difference is that we now use a  $\delta$ -net instead of a grid. If  $G$  is the true set, let  $i^*$ ,  $1 \leq i^* \leq N$ , be the index of a set of the  $\delta$ -net whose Hausdorff distance to  $G$  is not greater than  $\delta$ . By Lemma 4 in Section 2.5.1,

$$|G \triangle G_{i^*}| \leq \alpha_1 \delta.$$

It follows, from the definition of the estimator, that

$$\mathbb{P}_G \left[ |\tilde{G} \triangle G| \geq \epsilon \right] \leq \sum_{i \in \{1, \dots, N\} : |G_i \triangle G| \geq \epsilon} \mathbb{P}_G [\mathcal{A}(G_i) \leq \mathcal{A}(G_{i^*})]$$

This leads to the same inequality as (3.17) where the sum is now over  $i = 1, \dots, N$ , for which  $|G_i \triangle G| \geq \epsilon$ , and the term  $\frac{2d^{d+1}(3/2)^d \beta_d \tilde{c}_1}{n}$  should be replaced by  $\tilde{c}_1 \alpha_1 \delta$ :

$$\begin{aligned}
\mathbb{P}_G \left[ |\tilde{G} \triangle G| \geq \epsilon \right] &\leq \sum_{i \in \{1, \dots, N\}: |G_i \triangle G| \geq \epsilon} (1 - c_2 \epsilon + \tilde{c}_1 \alpha_1 \delta)^n \\
&\leq N \exp(-c_2 \epsilon n + \tilde{c}_1 \alpha_1 \delta n) \\
&\leq \tau_1 \exp\left(-c_2 \epsilon n + \tilde{c}_1 \alpha_1 \delta n + \tau_2 \delta^{-(d-1)/2}\right) \\
&\leq \tau_1 \exp(-c_2 \epsilon n + (\tilde{c}_1 \alpha_1 + \tau_2) \delta n),
\end{aligned}$$

since our choice of  $\delta$  guarantees that  $n\delta = \delta^{-(d-1)/2}$ . Hence, by choosing  $\epsilon = \frac{x}{n} + \frac{\tilde{c}_1 \alpha_1 + \tau_2}{c_2}$ , we get Theorem 3.4.  $\blacksquare$

### Proof of Theorem 3.5

We first prove this theorem in the case  $d = 2$  and then generalize the proof for  $d \geq 3$ .

We more or less follow the lines of the proof of the lower bound in [KT94] (which is similar to the proof of Assouad's lemma, see [Tsy09, Section 2.7.2]). Let  $G$  be the disk centered in  $(1/2, 1/2)$  of radius  $1/2$ , and  $P$  be a regular convex polygon with  $M$  vertices, all of them lying on the edge of  $G$ . Each edge of  $P$  cuts a cap off  $G$ , of area  $h$ , with  $\pi^3/(12M^3) \leq h \leq \pi^3/M^3$  as soon as  $M \geq 6$ , which we will assume in the sequel. We denote these caps by  $D_1, \dots, D_M$ , and for any  $\omega = (\omega_1, \dots, \omega_M) \in \{0, 1\}^M$  we denote by  $G_\omega$  the set made of  $G$  out of which we took all the caps  $D_j$  for which  $\omega_j = 0$ ,  $j = 1, \dots, M$ .

For  $j = 1, \dots, M$ , and  $(\omega_1, \dots, \omega_{j-1}, \omega_{j+1}, \dots, \omega_M) \in \{0, 1\}^{M-1}$  we denote by

$$\omega^{(j,0)} = (\omega_1, \dots, \omega_{j-1}, 0, \omega_{j+1}, \dots, \omega_M) \text{ and by}$$

$$\omega^{(j,1)} = (\omega_1, \dots, \omega_{j-1}, 1, \omega_{j+1}, \dots, \omega_M).$$

Note that for  $j = 1, \dots, M$ , and  $(\omega_1, \dots, \omega_{j-1}, \omega_{j+1}, \dots, \omega_M) \in \{0, 1\}^{M-1}$ ,

$$|G_{\omega^{(j,0)}} \triangle G_{\omega^{(j,1)}}| = h.$$

Now, let us consider any estimator  $\hat{G}$ . For  $j = 1, \dots, M$  we denote by  $A_j$  the smallest convex cone with origin at  $(1/2, 1/2)$  and which contains the cap  $D_j$ . Note that the cones  $A_j, j = 1, \dots, M$  have pairwise a null Lebesgue measure intersection. Then, we have the following inequalities.

$$\begin{aligned}
& \sup_{G \in \mathcal{K}_2} \mathbb{E}_G \left[ |G \Delta \hat{G}| \right] \\
& \geq \frac{1}{2^M} \sum_{\omega \in \{0,1\}^M} \mathbb{E}_{G_\omega} \left[ |G_\omega \Delta \hat{G}| \right] \\
& \geq \frac{1}{2^M} \sum_{\omega \in \{0,1\}^M} \sum_{j=1}^M \mathbb{E}_{G_\omega} \left[ |(G_\omega \cap A_j) \Delta (\hat{G} \cap A_j)| \right] \\
& = \frac{1}{2^M} \sum_{j=1}^M \sum_{\omega \in \{0,1\}^M} \mathbb{E}_{G_\omega} \left[ |(G_\omega \cap A_j) \Delta (\hat{G} \cap A_j)| \right] \\
& = \frac{1}{2^M} \sum_{j=1}^M \sum_{\omega_1, \dots, \omega_{j-1}, \omega_{j+1}, \dots, \omega_M} \left( \mathbb{E}_{G_\omega^{(j,0)}} \left[ |(G_\omega^{(j,0)} \cap A_j) \Delta (\hat{G} \cap A_j)| \right] \right. \\
& \quad \left. + \mathbb{E}_{G_\omega^{(j,1)}} \left[ |(G_\omega^{(j,1)} \cap A_j) \Delta (\hat{G} \cap A_j)| \right] \right). \tag{3.19}
\end{aligned}$$

Besides for  $j = 1, \dots, M$  and  $(\omega_1, \dots, \omega_{j-1}, \omega_{j+1}, \dots, \omega_M) \in \{0, 1\}^{M-1}$  we have

$$\begin{aligned}
& \mathbb{E}_{G_\omega^{(j,0)}} \left[ |(G_\omega^{(j,0)} \cap A_j) \Delta (\hat{G} \cap A_j)| \right] + \mathbb{E}_{G_\omega^{(j,1)}} \left[ |(G_\omega^{(j,1)} \cap A_j) \Delta (\hat{G} \cap A_j)| \right] \\
& = \int_{([0,1]^2 \times \mathbb{R})^n} |(G_\omega^{(j,0)} \cap A_j) \Delta (\hat{G} \cap A_j)| d\mathbb{P}_{G_\omega^{(j,0)}}^{\otimes n} \\
& \quad + \int_{([0,1]^2 \times \mathbb{R})^n} |(G_\omega^{(j,1)} \cap A_j) \Delta (\hat{G} \cap A_j)| d\mathbb{P}_{G_\omega^{(j,1)}}^{\otimes n} \\
& \geq \int_{([0,1]^2 \times \mathbb{R})^n} \left( |(G_\omega^{(j,0)} \cap A_j) \Delta (\hat{G} \cap A_j)| + |(G_\omega^{(j,1)} \cap A_j) \Delta (\hat{G} \cap A_j)| \right) \times \\
& \quad \min(d\mathbb{P}_{G_\omega^{(j,0)}}^{\otimes n}, d\mathbb{P}_{G_\omega^{(j,1)}}^{\otimes n}) \\
& \geq \int_{([0,1]^2 \times \mathbb{R})^n} \left( |(G_\omega^{(j,0)} \cap A_j) \Delta (\hat{G} \cap A_j)| \right) \min(d\mathbb{P}_{G_\omega^{(j,0)}}^{\otimes n}, d\mathbb{P}_{G_\omega^{(j,1)}}^{\otimes n}), \\
& \quad \text{by triangle inequality,} \\
& = h \int_{([0,1]^2 \times \mathbb{R})^n} \min(d\mathbb{P}_{G_\omega^{(j,0)}}^{\otimes n}, d\mathbb{P}_{G_\omega^{(j,1)}}^{\otimes n}) \\
& \geq \frac{h}{2} \left( 1 - \frac{H^2(\mathbb{P}_{G_\omega^{(j,0)}}^{\otimes n}, \mathbb{P}_{G_\omega^{(j,1)}}^{\otimes n})}{2} \right)^2 \\
& = \frac{h}{2} \left( 1 - \frac{H^2(\mathbb{P}_{G_\omega^{(j,0)}}^{\otimes n}, \mathbb{P}_{G_\omega^{(j,1)}}^{\otimes n})}{2} \right)^{2n}, \tag{3.20}
\end{aligned}$$

using properties of the Hellinger distance (cf. Lemma 7). To compute the Hellinger distance between  $\mathbb{P}_{G_\omega^{(j,0)}}$  and  $\mathbb{P}_{G_\omega^{(j,1)}}$  we use Lemma 11, in Section 2.5.1.

Then if we denote by  $c_9 = 1 - e^{-\frac{1}{8\sigma^2}}$ , it follows from (3.19) and (3.20) that

$$\begin{aligned} \sup_{G \in \mathcal{K}_2} \mathbb{E}_G \left[ |G \triangle \hat{G}| \right] &\geq \frac{1}{2^M} M 2^{M-1} \frac{h}{2} (1 - c_9 h)^{2n} \\ &\geq \frac{Mh}{4} (1 - c_9 h)^{2n} \\ &\geq \frac{\pi^3}{12M^2} (1 - \pi^3 c_9 / M^3)^{2n}. \end{aligned}$$

Besides, since we assumed that  $M \geq 6$ , we have that

$$\pi^3 c_9 / M^3 \leq \pi^3 c_9 / 6^3 = \frac{\pi^3}{6^3} \left( 1 - \exp\left(-\frac{1}{8\sigma^2}\right) \right) \leq \frac{\pi^3}{6^3} < 1,$$

and if we take  $M = \lfloor n^{1/3} \rfloor$ , we get by concavity of the logarithm

$$\begin{aligned} \sup_{G \in \mathcal{K}_2} \mathbb{E}_G \left[ |G \triangle \hat{G}| \right] &\geq \frac{\pi^3}{12M^2} \exp \left( \frac{432 \ln(1 - \pi^3/216) \left( 1 - e^{-\frac{1}{8\sigma^2}} \right) n M^{-3}}{\pi^3} \right) \\ &\geq c_{14} n^{-2/3}, \end{aligned}$$

where  $c_{14} = \frac{\pi^3}{12} \exp \left( \frac{432 \ln(1 - \pi^3/216) \left( 1 - e^{-\frac{1}{8\sigma^2}} \right)}{\pi^3} \right)$  is a positive constant that depends on  $\sigma$  only. This inequality holds for  $n \geq 216$ , so that  $M \geq 6$ .

We now deal with the case  $d \geq 3$ . Let us first recall some definitions and resulting properties, that can also be found in [KT59].

**Definition 3.1.** Let  $(S, \rho)$  be a metric space and  $\eta$  a positive number. A family  $\mathcal{Y} \subseteq S$  is called an  $\eta$ -packing family if and only if  $\rho(y, y') \geq \eta$ , for  $(y, y') \in \mathcal{Y}$  with  $y \neq y'$ . An  $\eta$ -packing family is called maximal if and only if it is not strictly included in any other  $\eta$ -packing family. A family  $\mathcal{Z}$  is called an  $\eta$ -net if and only if for all  $x \in S$ , there is an element  $z \in \mathcal{Z}$  which satisfies  $\rho(x, z) \leq \eta$ .

The construction of the hypotheses used for the lower bound in the case  $d = 2$  requires a little more work in the general dimension case, since it is not always possible to construct a regular convex polytope with a fixed number of vertices or facets, and inscribed in a given ball. For the following geometrical construction, we refer to Figure 3.1.

Let  $G_0$  be the closed ball in  $\mathbb{R}^d$ , with center  $a_0 = (1/2, \dots, 1/2)$  and radius  $1/2$ , so that  $G_0 \subseteq [0, 1]^d$ . Let  $\eta \in (0, 1)$  which will be chosen precisely later, and  $\{y_1, \dots, y_{M_\eta}\}$  a maximal  $\eta$ -packing family of  $S = \partial G_0$ . The integer  $M_\eta$  satisfies (3.46) by Lemma 12. For  $j \in \{1, \dots, M_\eta\}$ ,

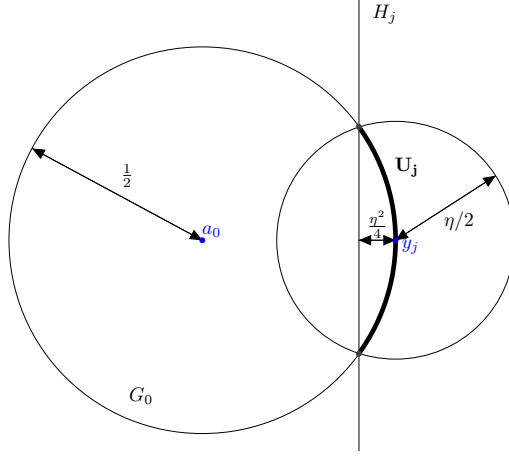


Figure 3.1: Construction of the hypotheses

we set by  $U_j = S \cap B_d(y_j, \eta/2)$ , and denote by  $W_j$  the  $(d-2)$ -dimensional sphere  $S \cap \partial B_d(y_j, \eta/2)$ . Let  $H_j$  be affine hull of  $W_j$ , i.e. its supporting hyperplane. This hyperplane dissects the space  $\mathbb{R}^d$  into two halfspaces. Let  $H_j^-$  be the one that contains the point  $y_j$ . For  $\omega = (\omega_1, \dots, \omega_{M_\eta}) \in \{0, 1\}^{M_\eta}$ , we set

$$G_\omega = G_0 \setminus \left( \bigcap_{j=1, \dots, M_\eta; \omega_j=0} H_j^- \right).$$

The set  $G_\omega$  is made of  $G_0$  from which we remove all the caps cut off by the hyperplanes  $H_j$ , for all the indices  $j$  such that  $\omega_j = 0$ .

For each  $j \in \{1, \dots, M_\eta\}$ , let  $A_j$  be the smallest closed convex cone with vertex  $a_0 = (1/2, \dots, 1/2)$  that contains  $U_j$ . Note that the cones  $A_j, j = 1, \dots, M_\eta$  have pairwise empty intersection, since  $G_0$  is convex and the sets  $U_j$  are disjoint. We are now all set to reproduce the proof written in the case  $d = 2$ . Note that

$$|G_{\omega(j,0)} \triangle G_{\omega(j,1)}| = |(G_{\omega(j,0)} \cap A_j) \triangle (G_{\omega(j,1)} \cap A_j)|,$$

for all  $\omega \in \{0, 1\}^{M_\eta}$  and  $j \in \{1, \dots, M_\eta\}$ , and this quantity is equal to

$$\int_0^{\eta^2/4} |B_{d-1}(0, \sqrt{r - r^2})|_{d-1} dr,$$

since, as mentioned before,  $\eta^2/4$  is the height of the cap cut off by  $H_j$ , which is equal to the distance between  $y_j$  and the hyperplane  $H_j$ , and which is independent of the index  $j$ .

Therefore,

$$\begin{aligned}
|G_{\omega(j,0)} \triangle G_{\omega(j,1)}| &= \int_0^{\frac{\eta^2}{4}} |B_{d-1}(0, \sqrt{r-r^2})|_{d-1} dr \\
&= \int_0^{\frac{\eta^2}{4}} \beta_{d-1} (r-r^2)^{(d-1)/2} dr \\
&= \beta_{d-1} \int_0^{\frac{\eta^2}{4}} (r-r^2)^{(d-1)/2} dr \\
&= \frac{\beta_{d-1} \eta^{d+1}}{4^{d+1}} \int_0^1 u^{(d-1)/2} \left(1 - \frac{\eta^2 u}{4}\right)^{(d-1)/2} du.
\end{aligned}$$

Since  $0 < \eta^2/4 < 1/4$ , we then get

$$\frac{3^{(d-1)/2} \eta^{d+1} \beta_{d-1}}{2^{3d}(d+1)} \leq |G_{\omega(j,0)} \triangle G_{\omega(j,1)}| \leq \frac{\eta^{d+1} \beta_{d-1}}{2^{2d+1}(d+1)}. \quad (3.21)$$

Now, continuing (3.19) and (3.20), replacing  $M$  by  $M_\eta$  and  $h$  by the lower bound in (3.21) and using lemmas 11 and 12, we get that

$$\sup_{G \in \mathcal{K}_d} \mathbb{E}_G \left[ |G \triangle \hat{G}| \right] \geq c_8 \eta^2 \left(1 - c_9 \eta^{d+1}\right)^{2n}, \quad (3.22)$$

where

$$c_8 = \frac{3^{(d-1)/2} \beta_{d-1} d}{2^{4d+1}(d+1)\sqrt{d+2}}$$

and

$$c_9 = \frac{(1 - e^{-\frac{1}{8\sigma^2}}) \beta_{d-1}}{2^{2d+1}(d+1)}.$$

Note that since the ball  $B_{d-1}(0, 1/2)$  is included in the  $(d-1)$ -dimensional hypercube centered at the origin, with sides of length 1, the following inequality holds

$$|B_{d-1}(0, \frac{1}{2})| = \frac{\beta_{d-1}}{2^{d-1}} \leq 1,$$

and this shows that  $c_9 < 1$ . Therefore, since  $\eta < 1$  as well, the concavity of the logarithm leads (3.22) to

$$\sup_{G \in \mathcal{K}_d} \mathbb{E}_G \left[ |G \triangle \hat{G}| \right] \geq c_8 \eta^2 \exp \left( 2n \ln(1 - c_9) \eta^{d+1} \right).$$

Let us choose  $\eta = n^{-1/(d+1)}$ , so that (3.22) becomes

$$\sup_{G \in \mathcal{K}_d} \mathbb{E}_G \left[ |G \triangle \hat{G}| \right] \geq c_{10} n^{-\frac{2}{d+1}},$$



where  $c_{10} = c_8(1 - c_9)^2 > 0$ . ■

### Proof of Theorem 3.6

Let  $r^*$  be a given and finite integer such that  $d + 1 \leq r^* \leq R_n - 1$ . Recall that by definition,  $\hat{Q}_n^{(r^*)} = \hat{P}_n^{(r^*)}$ . Note that if  $r^* \leq r \leq r'$ , then  $\mathcal{P}_{r^*}^{(1)} \subseteq \mathcal{P}_r^{(1)} \subseteq \mathcal{P}_{r'}^{(1)}$ . Therefore if  $P \in \mathcal{P}_{r^*}^{(1)}$  and  $G = P$  in model (1.3), by Theorem 3.1 it is likely that with high probability we have, using the triangle inequality,

$$|\hat{P}_n^{(r)} \triangle \hat{P}_n^{(r')}| \leq \frac{Cdr' \ln n}{n}, \quad (3.23)$$

for any  $r^* \leq r \leq r'$ , where  $C$  is a constant. Therefore it is reasonable to select  $\hat{r}$  as the minimal integer that satisfies (3.23).

Let  $\hat{r}$  be chosen as in Theorem 3.6. For  $r = d + 1, \dots, R_n$ , let us denote by  $A_r$  the event following event.

$$A_r = \left\{ \forall r' = r, \dots, R_n, |\hat{Q}_n^{(r)} \triangle \hat{Q}_n^{(r')}| \leq \frac{2c_a r' \ln n}{n} \right\},$$

where  $c_2$  is the same constant as in Theorem 3.1. Then  $\hat{r}$  is the smallest integer  $r \leq R_n$  such that  $A_r$  holds.

Let  $P \in \mathcal{P}_{r^*}^{(1)}$ . We write the following.

$$\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|] = \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|I(\hat{r} \leq r^*)] + \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|I(\hat{r} > r^*)], \quad (3.24)$$

and we bound separately the two terms in the right side. Note that if  $\hat{r} \leq r^*$ , then, since the event  $A_{\hat{r}}$  holds by definition,

$$|\hat{Q}_n^{(r^*)} \triangle \hat{Q}_n^{(\hat{r})}| \leq \frac{2c_a r^* \ln n}{n}.$$

Therefore, using the triangle inequality,

$$\begin{aligned} \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|I(\hat{r} \leq r^*)] &\leq \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle \hat{Q}_n^{(r^*)}|I(\hat{r} \leq r^*)] + \mathbb{E}_P[|\hat{Q}_n^{(r^*)} \triangle P|I(\hat{r} \leq r^*)] \\ &\leq \frac{2c_a r^* \ln n}{n} + \frac{c_1(\sigma, d, q)dr^* \ln n}{n} \text{ by Corollary 3.1, since } \hat{Q}_n^{(r^*)} = \hat{P}_n^{(r^*)} \\ &\leq \frac{c_{11}r^* \ln n}{n}, \end{aligned} \quad (3.25)$$

where  $c_{11}$  depends only on  $d$  and  $\sigma$ . The second term of (3.24) is bounded differently. First note that for all  $r = d + 1, \dots, R_n$ ,  $\hat{Q}_n^{(r)} \subseteq [0, 1]^d$ , so  $|\hat{Q}_n^{(r)}| \leq 1$ . Thus, if  $\overline{A_{r^*}}$  stands for the complement of the event  $A_{r^*}$ , we have the following inequalities.

$$\begin{aligned}
& \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} > r^*)] \\
& \leq 2\mathbb{P}_P[\hat{r} > r^*] \\
& \leq 2\mathbb{P}_P[\overline{A_{r^*}}] \\
& \leq 2 \sum_{r=r^*}^{R_n} \mathbb{P}_P \left[ |\hat{Q}_n^{(r^*)} \triangle \hat{Q}_n^{(r)}| > \frac{2c_a r \ln n}{n} \right] \\
& \leq 2 \sum_{r=r^*}^{R_n} \mathbb{P}_P \left[ |\hat{Q}_n^{(r^*)} \triangle P| + |\hat{Q}_n^{(r)} \triangle P| > \frac{2c_a r \ln n}{n} \right] \\
& \leq 2 \sum_{r=r^*}^{R_n} \left( \mathbb{P}_P \left[ |\hat{Q}_n^{(r^*)} \triangle P| > \frac{c_a r \ln n}{n} \right] + \mathbb{P}_P \left[ |\hat{Q}_n^{(r)} \triangle P| > \frac{c_a r \ln n}{n} \right] \right) \\
& \leq 2 \sum_{r=r^*}^{R_n-1} \left( \mathbb{P}_P \left[ |\hat{P}_n^{(r^*)} \triangle P| > \frac{c_a r^* \ln n}{n} \right] + \mathbb{P}_P \left[ |\hat{P}_n^{(r)} \triangle P| > \frac{c_a r \ln n}{n} \right] \right) \\
& \quad + 2\mathbb{P}_P \left[ |\hat{P}_n^{(r^*)} \triangle P| > \frac{c_a r^* \ln n}{n} \right] + 2\mathbb{P}_P \left[ |\tilde{G} \triangle P| > \frac{c_a R_n \ln n}{n} \right] \tag{3.26}
\end{aligned}$$

Note that since  $P \in \mathcal{P}_{r^*}^{(1)}$ , it is also true that  $P \in \mathcal{P}_r^{(1)}, \forall r \geq r^*$ . Therefore, if  $r^* \leq r \leq R^*-1$ , we have, using Theorem 3.1, with  $x = (c_a - 2d/c_2)r \ln n \geq r \ln n/c_2$ ,

$$\mathbb{P}_P \left[ |\hat{P}_n^{(r)} \triangle P| > \frac{c_a r \ln n}{n} \right] \leq c_1 e^{-r \ln n} \leq c_1 n^{-(d+1)}.$$

In addition, by Theorem 3.4, with  $x = (c_a - c_0)R_n \ln n \geq R_n \ln n/c_2$ ,

$$\mathbb{P}_P \left[ |\tilde{G} \triangle P| > \frac{c_a R_n \ln n}{n} \right] \leq \tau_1 e^{-R_n \ln n} \leq \tau_1 n^{-(d+1)}.$$

It comes from (3.26) that

$$\begin{aligned}
\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} > r^*)] & \leq 2 \sum_{r=r^*}^{R_n-1} 2c_1 n^{-(d+1)} + 2c_1 n^{-(d+1)} + 2\tau_1 n^{-(d+1)} \\
& \leq 4 \max(c_1, \tau_1) R_n n^{-(d+1)}. \tag{3.27}
\end{aligned}$$

Finally, using (3.25) and (3.27),

$$\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|] \leq \frac{c_{12} r^* \ln n}{n},$$

where  $c_{12}$  is a positive constant that depends on  $d$  and  $\sigma$ . Let us now assume that  $r^*$  is a given integer larger than  $R_n$ , possibly infinite, and that  $P \in \mathcal{P}_{r^*}^{(1)}$ . Recall that if  $r^* = \infty$  we denote by

$\mathcal{P}_\infty^{(1)}$  the class  $\mathcal{K}_d^{(1)}$ . Then with probability one,  $\hat{r} \leq r^*$ . First of all, note that obviously, since by definition,  $\hat{r} \leq R_n$ ,

$$|\hat{Q}_n^{(R_n)} \triangle \hat{Q}_n^{(\hat{r})}| \leq \frac{2c_a R_n \ln n}{n} \leq 2c_a n^{-\frac{2}{d+1}}$$

with probability one. Then, by the triangle inequality,

$$\begin{aligned} \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|] &\leq 2c_a n^{-\frac{2}{d+1}} + \mathbb{E}_P[|\hat{Q}_n^{(R_n)} \triangle P|] \\ &\leq 2c_a n^{-\frac{2}{d+1}} + c_3(\sigma, d, q) n^{-\frac{2}{d+1}}, \end{aligned}$$

by Corollary 3.3, since  $P \in \mathcal{P}_{r^*}^{(1)} \subseteq \mathcal{P}_\infty^{(1)}$  and  $\hat{Q}_n^{(R_n)}$  is the estimator of Theorem 3.4. Theorem 3.6 is then proven.  $\blacksquare$

### Proof of Theorem 3.7

**On the class  $\mathcal{S}_0$**

**Upper bound** Let us first prove the upper bound, i.e. assume that  $nh \rightarrow \infty$ , and prove that there exists a consistent test. Recall that  $N = \max\{i = 1, \dots, n : X_i \leq h\} = \#(\mathcal{X} \cap [0, h])$ . If the design is (DD), then  $N$  is just equal to the integer part of  $nh$ . If the design (RD), then  $N$  is a binomial random variable, with parameters  $n$  and  $h$ . Let us show first that the error of the first kind of the test  $T_n^0$  goes to zero, when  $n \rightarrow \infty$ .

$$\begin{aligned} \mathbb{P}_\emptyset[S \leq cN] &= \mathbb{P}_\emptyset\left[\#\left\{i = 1, \dots, N : Y_i > \frac{1}{2}\right\} \geq (1-c)N\right] \\ &\leq \mathbb{E}\left[\mathbb{P}_\emptyset\left[\#\left\{i = 1, \dots, N : \xi_i > \frac{1}{2}\right\} \geq (1-c)N \middle| \mathcal{X}\right]\right]. \end{aligned}$$

Since the  $\xi_i$ 's are independent of  $\mathcal{X}$ , the distribution of  $\#\{i = 1, \dots, N : \xi_i > \frac{1}{2}\}$  conditionally to  $\mathcal{X}$  is binomial, with parameters  $N$  and  $\beta$ , where  $\beta = \mathbb{P}[\xi_1 > 1/2] \in [0, 1)$ . Thus, by Bernstein's inequality for binomial random variables, by defining  $\gamma = \frac{(1-c-\beta)^2}{2\beta(1-\beta) + (1-c-\beta)/3} > 0$ ,

$$\mathbb{P}_\emptyset[S \leq cN] \leq \mathbb{E}[\exp(-\gamma N)].$$

If  $\mathcal{X}$  satisfies (DD), then  $N \geq nh - 1$  and it is clear that  $\mathbb{P}_\emptyset[S \leq cN] \rightarrow 0$ . If  $\mathcal{X}$  satisfies (RD), then

$$\mathbb{E}[\exp(-\gamma N)] = \exp(-nh(1 - e^{-\gamma})),$$

so  $\mathbb{P}_\emptyset[S \leq cN] \rightarrow 0$ .

Let us show, now, that the error of the second kind goes to zero as well. Let  $G \in \mathcal{S}_0$  satisfying the alternative hypothesis, i.e.  $|G| \geq h$ . Denote by  $\beta' = \mathbb{P}[\xi_1 \leq -1/2]$ , and by  $\gamma' = \frac{(c - \beta')^2}{2\beta'(1 - \beta') + (c - \beta')/3} > 0$

$$\begin{aligned} \mathbb{P}_G[S > cN] &= \mathbb{P}_G \left[ \# \left\{ i = 1, \dots, N : Y_i \leq \frac{1}{2} \right\} > cN \right] \\ &\leq \mathbb{E} \left[ \mathbb{P}_\emptyset \left[ \# \left\{ i = 1, \dots, N : \xi_i \leq -\frac{1}{2} \right\} > cN \middle| \mathcal{X} \right] \right] \\ &\leq \mathbb{E} [\exp(-\gamma'N)], \end{aligned}$$

by a similar computation to that for the error of the first kind. Since the right-side of the last inequality does not depend on  $G$ ,

$$\sup_{|G| \geq h} \mathbb{P}_G[S > cN] \leq \mathbb{E} [\exp(-\gamma'N)]$$

and therefore, by the same argument as for the error of the first kind, goes to zero when  $n \rightarrow \infty$ , for both designs (DD) and (RD).

**Lower bound** Assume, now, that  $nh \rightarrow 0$ . Let  $\tau_n$  be any test. Let  $G_1 = [0, h]$ . We denote by  $\mathcal{H}$  the Hellinger distance between probability measures. The following computation uses properties of this distance, which can be found in [Tsy09].

$$\begin{aligned} \gamma_n(\tau_n, \mathcal{C}) &\geq \mathbb{E}_\emptyset[\tau_n] + \mathbb{E}_{G_1}[1 - \tau_n] \\ &\geq \int \min(d\mathbb{P}_\emptyset, d\mathbb{P}_{G_1}) \\ &\geq \frac{1}{2} \left( 1 - \frac{H(\mathbb{P}_\emptyset, \mathbb{P}_{G_1})}{2} \right)^2. \end{aligned} \tag{3.28}$$

Let  $G, G' \in \mathcal{S}$ . A simple computation shows that, for the design (DD),

$$1 - \frac{H(\mathbb{P}_G, \mathbb{P}_{G'})}{2} = \exp \left( -\frac{\#(\mathcal{X} \cap (G \triangle G'))}{8\sigma^2} \right), \tag{3.29}$$

and for the design (RD),

$$1 - \frac{H(\mathbb{P}_G, \mathbb{P}_{G'})}{2} = \left( 1 - \left( 1 - e^{-\frac{1}{8\sigma^2}} \right) |G \triangle G'| \right)^n. \tag{3.30}$$

In particular, for the design (DD),

$$1 - \frac{H(\mathbb{P}_\emptyset, \mathbb{P}_{G_1})}{2} \geq \exp\left(-\frac{nh}{8\sigma^2}\right),$$

and for the design (RD),

$$1 - \frac{H(\mathbb{P}_\emptyset, \mathbb{P}_{G_1}^{\otimes n})}{2} = \left(1 - \left(1 - e^{-\frac{1}{8\sigma^2}}\right)h\right)^n.$$

In both cases, we showed that the right side of (3.28) tends to  $1/2$ , when  $n \rightarrow \infty$ . Therefore the test  $\tau_n$  is not consistent.

### On the class $\mathcal{S}$

**Upper bound** Assume that  $\frac{nh}{\ln n} \rightarrow \infty$ . Let us first show that the error of the first kind of  $T_n^1$  goes to zero, when  $n \rightarrow \infty$ . Recall that  $T_n^1 = \mathbb{1}(R \geq 0)$ , where  $R = \sup_{|G| \geq h} R(G)$  and  $R(G) = \sum_{i=1}^n Y_i \mathbb{1}(X_i \in G) - \frac{\#(\mathcal{X} \cap G)}{2}$ , for all  $G \in \mathcal{S}$ . Note that  $R(G)$  is piecewise constant, and can only take a finite number of values. It is clear that

$$\{R(G) : G \in \mathcal{S}, |G| \geq h\} = \{R([X_k, X_l]) : 1 \leq k < l \leq n, X_l - X_k > h\}.$$

Recall that for  $1 \leq k < l \leq n$ ,  $R([X_k, X_l]) = \frac{1}{2} \sum_{i=k}^{l-1} (2Y_i - 1)$ . Therefore,

$$\begin{aligned} \mathbb{P}_\emptyset[R \geq 0] &= \mathbb{P}_\emptyset \left[ \max_{\substack{1 \leq k < l \leq n \\ X_l - X_k > h}} R([X_k, X_l]) > 0 \right] \\ &\leq \mathbb{P}_\emptyset \left[ \bigcup_{1 \leq k < l \leq n} \{R([X_k, X_l]) > 0\} \cap \{X_l - X_k > h\} \right] \\ &\leq \sum_{1 \leq k < l \leq n} \mathbb{P}_\emptyset [R([X_k, X_l]) > 0, X_l - X_k > h] \\ &\leq \sum_{1 \leq k < l \leq n} \mathbb{P}_\emptyset \left[ \sum_{i=k}^{l-1} (2\xi_i - 1) > 0 \right] \mathbb{P}[X_l - X_k > h]. \end{aligned} \tag{3.31}$$

For  $1 \leq k < l \leq n$ ,

$$\mathbb{P}_\emptyset \left[ \sum_{i=k}^{l-1} (2\xi_i - 1) > 0 \right] \leq \exp\left(-\frac{(l-k)\sigma^2}{8}\right), \tag{3.32}$$

using Markov's inequality and Assumption A.

If the design is (DD), then  $\mathbb{P}[X_l - X_k > h]$  is 1 if and only if  $l - k > nh$ , 0 otherwise, so

from (3.31) and (3.32) we get that

$$\begin{aligned}\mathbb{P}_\emptyset[R \geq 0] &\leq \sum_{l-k > nh} \exp\left(-\frac{(l-k)\sigma^2}{8}\right) \\ &\leq \sum_{l-k > nh} \exp\left(\frac{(-nh)\sigma^2}{8}\right) \\ &\leq \frac{n^2}{2} \exp\left(\frac{(-nh)\sigma^2}{8}\right) \rightarrow 0,\end{aligned}$$

when  $n \rightarrow \infty$ , which proves that the error of the first kind goes to zero.

If the design is (RD), let us use the following Lemma.

**Lemma 9.** *Let  $X_1, \dots, X_n$  be the (RD) design. Then, for any  $1 \leq k < l \leq n$ , and  $h > 0$ ,*

$$\mathbb{P}[X_l - X_k > h] \leq n \exp(-nh(1 - e^{-u}) + u(l - k)), \forall u > 0.$$

Therefore, by (3.31), (3.32) and (3.47), and Lemma 9 with  $u = \sigma^2/8$ ,

$$\begin{aligned}\mathbb{P}_\emptyset[R \geq 0] &\leq \sum_{1 \leq k < l \leq n} \exp\left(-\frac{(l-k)\sigma^2}{8} - nh(1 - e^{-u}) + u(l - k)\right) \\ &\leq \frac{n^2}{2} \exp\left(-nh(1 - e^{-\sigma^2/8})\right) \rightarrow 0,\end{aligned}$$

when  $n \rightarrow \infty$ , which proves that the error of the first kind goes to zero.

Let us bound, now, the error of the second kind. Let  $G \in \mathcal{S}$  satisfying  $|G| \geq h$ . For this  $G$ , denote by  $N_G = \#(\mathcal{X} \cap G)$ . Then,

$$\begin{aligned}\mathbb{P}_G[R < 0] &\leq \mathbb{P}_G[R(G) \leq N_G/2] \\ &\leq \mathbb{P}\left[\sum_{i=1}^n \xi_i \mathbb{1}(X_i \in G) \leq -N_G/2\right].\end{aligned}\tag{3.33}$$

For the design (DD),  $N_G$  is the integer part of  $n|G|$ , so  $N_G \geq nh$ . Therefore, by Markov's inequality, and by Assumption A, (3.33) becomes

$$\mathbb{P}_G[R < 0] \leq \exp\left(-\frac{N_G}{8\sigma^2}\right) \leq \exp\left(-\frac{nh}{8\sigma^2}\right).\tag{3.34}$$

For the design (RD),  $N_G$  is a random binomial variable, with parameters  $n$  and  $|G|$ . By

conditioning to the design and using Markov's inequality, (3.33) becomes

$$\begin{aligned}
\mathbb{P}_G[R < 0] &\leq \mathbb{P}\left[\sum_{i=1}^n -\xi_i \mathbf{1}(X_i \in G) \geq N_G/2\right] \\
&\leq \mathbb{E}\left[\exp\left(-\frac{N_G}{8\sigma^2}\right)\right] \\
&\leq \exp(-Cn|G|) \leq \exp(-Cnh),
\end{aligned} \tag{3.35}$$

where  $C = 1 - e^{-\frac{1}{8\sigma^2}}$ .

In both cases (3.34) and (3.35), the right side does not depend on  $G$ , and goes to zero as  $n \rightarrow \infty$ . We conclude that, for both designs (DD) and (RD),

$$\sup_{|G| \geq h} \mathbb{P}_G[R < 0] \rightarrow 0,$$

which ends the proof of the upper bound.

**Lower bound** We more or less reproduce the proof of [Gay01], Theorem 3.1. Here, the noise is supposed to be Gaussian, with variance  $\sigma^2$ . Let us assume that  $\frac{nh}{\ln n} \rightarrow 0$ . Let  $M = 1/h$ , assumed to be an integer, without loss of generality. For  $q = 0, \dots, M$ , let  $G_q = [qh, (q+1)h]$ . For  $q = 1, \dots, M$ , let  $Z_q = \frac{d\mathbb{P}_{G_q}}{d\mathbb{P}_\emptyset}(X_1, Y_1, \dots, X_n, Y_n)$ , and denote by  $\bar{Z} = \frac{1}{M} \sum_{q=1}^M Z_q$ . Let  $\tau_n$  be any test. Then,

$$\begin{aligned}
\gamma_n(\tau_n, \mathcal{S}) &\geq \mathbb{P}_\emptyset[\tau_n = 1] + \frac{1}{M} \sum_{q=1}^M \mathbb{P}_{G_q}[\tau_n = 0] \\
&\geq \frac{1}{M} \sum_{q=1}^M (\mathbb{P}_\emptyset[\tau_n = 1] + \mathbb{P}_{G_q}[\tau_n = 0]) \\
&\geq \frac{1}{M} \sum_{q=1}^M (\mathbb{E}_\emptyset[\tau_n] + \mathbb{E}_{G_q}[1 - \tau_n]) \\
&\geq \frac{1}{M} \sum_{q=1}^M \mathbb{E}_\emptyset[\tau_n + (1 - \tau_n)Z_q] \\
&\geq \mathbb{E}_\emptyset[\tau_n + (1 - \tau_n)\bar{Z}] \\
&\geq \mathbb{E}_\emptyset[(\tau_n + (1 - \tau_n)\bar{Z}) \mathbf{1}(\bar{Z} \geq 1/2)] \\
&\geq \frac{1}{2} \mathbb{P}_\emptyset[\bar{Z} \geq 1/2].
\end{aligned} \tag{3.36}$$

Let us prove that  $\mathbb{E}_\emptyset[\bar{Z}] = 1$ , and that  $\mathbb{V}_\emptyset[\bar{Z}] \rightarrow 0$ . This will imply that the right side term of (3.36) goes to zero, when  $n \rightarrow \infty$ .

For  $q = 1, \dots, M$ , under the null hypothesis,

$$\begin{aligned} Z_q &= \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n ((Y_i - \mathbb{1}(X_i \in G_q))^2 - Y_i^2) \right) \\ &= \exp \left( \frac{1}{2\sigma^2} \sum_{i=1}^n (2\xi_i - 1) \mathbb{1}(X_i \in G_q) \right). \end{aligned} \quad (3.37)$$

By its definition,  $Z_q$  has expectation 1 under  $\mathbb{P}_\emptyset$ :

$$\mathbb{E}_\emptyset[\bar{Z}] = 1. \quad (3.38)$$

Since, almost surely, no design point falls in two  $G_q$ 's at the time, a simple computation shows that the random variables  $Z_q, q = 1, \dots, M$ , are not correlated. Thus,

$$\mathbb{V}_\emptyset[\bar{Z}] = \frac{1}{M^2} \sum_{q=1}^M \mathbb{V}_\emptyset[Z_q].$$

Let us bound from above  $\mathbb{V}_\emptyset[Z_q]$ , for  $q = 1, \dots, M$ :

$$\begin{aligned} \mathbb{V}_\emptyset[Z_q] &\leq \mathbb{E}_\emptyset[Z_q^2] \\ &= \mathbb{E} \left[ \exp \left( -\frac{\#(\mathcal{X} \cap G_q)}{\sigma^2} \right) \mathbb{E}_\emptyset \left[ \exp \left( \frac{2}{\sigma^2} \sum_{i=1}^n \xi_i \mathbb{1}(X_i \in G_q) \right) | \mathcal{X} \right] \right] \\ &= \mathbb{E} \left[ \exp \left( \frac{\#(\mathcal{X} \cap G_q)}{\sigma^2} \right) \right]. \end{aligned} \quad (3.39)$$

If the design is (DD), then we get that

$$\mathbb{V}_\emptyset[Z_q] \leq \exp \left( \frac{nh + 1}{\sigma^2} \right),$$

and the variance of  $\bar{Z}$  is then bounded from above:

$$\mathbb{V}_\emptyset[\bar{Z}] \leq h \exp \left( \frac{nh + 1}{\sigma^2} \right). \quad (3.40)$$

If the design is (RD), then  $\#(\mathcal{X} \cap G_q)$  is a binomial random variable with parameters  $n$  and  $h$ , so from (3.39), we get that

$$\begin{aligned} \mathbb{V}_\emptyset[Z_q] &\leq \left( 1 + \left( e^{1/\sigma^2} - 1 \right) h \right)^n \\ &\leq \exp(Cnh), \end{aligned}$$



where  $C = e^{1/\sigma^2} - 1$ , and the variance of  $\bar{Z}$  is then bounded from above:

$$\mathbb{V}_\emptyset[\bar{Z}] \leq h \exp(Cnh). \quad (3.41)$$

Since we assumed that  $nh/\ln n \rightarrow 0$ , the right side terms of (3.40) and (3.41) go to zero, and therefore, for both designs (DD) and (RD),

$$\mathbb{V}_\emptyset[\bar{Z}] \rightarrow 0. \quad (3.42)$$

Finally, we get from (3.36), (3.38) and (3.42), that

$$\liminf_{n \rightarrow \infty} \gamma_n(\tau_n, \mathcal{S}) \geq \frac{1}{2}.$$

This concludes the proof. ■

### Proof of Theorem 3.8

The beginning of this proof holds for any design  $\{X_1, \dots, X_n\}$ , independent of the noise  $\xi_i, i = 1, \dots, n$ . Let  $G \in \mathcal{S}_0$ . Let  $M = \max\{i = 1, \dots, n : X_i \in G\}$  - set  $M = 0$  if the set is empty -.

Then,  $\hat{M}_n \in \underset{M'=1, \dots, n}{\text{ArgMax}} (F(M') - F(M))$ , and, by (3.8),

$$F(M') - F(M) = -|M' - M| + \begin{cases} 2 \sum_{i=M'+1}^M \xi_i & \text{if } M > M', \\ 0 & \text{if } M' = M, \\ -2 \sum_{i=M+1}^{M'} \xi_i & \text{if } M < M'. \end{cases}$$

Let us complete the i.i.d. sequence  $\xi_1, \dots, \xi_n$  to obtain an infinite double sided i.i.d. sequence  $(\xi_i)_{i \in \mathbb{Z}}$ , independent of the design. Let  $k \in \mathbb{N}^*$  be any positive integer. Define, for  $i \in \mathbb{Z}$ ,  $\tilde{\xi}_i = \xi_{i+M}$ . Since  $M$  depends on the design only, it is independent of the  $\xi_i, i \in \mathbb{Z}$ , and therefore, the  $\tilde{\xi}_i, i \in \mathbb{Z}$  are i.i.d., with same distribution as  $\xi_1$ . Let  $E_k$  be the event  $\{\hat{M}_n \geq M + k\}$ . If  $E_k$  holds, then :

$$0 \leq F(\hat{M}_n) - F(M) = M - \hat{M}_n - 2 \sum_{i=M+1}^{\hat{M}_n} \xi_i,$$

and it follows that

$$\begin{aligned}
0 &\leq \max_{M+k \leq j \leq n} \left( M - j - 2 \sum_{i=M+1}^j \xi_i \right) \\
&\leq \max_{M+k \leq j \leq n} \left( M - j - 2 \sum_{i=1}^{j-M} \tilde{\xi}_i \right) \\
&\leq \max_{k \leq j \leq n-M} \left( -j - 2 \sum_{i=1}^j \tilde{\xi}_i \right).
\end{aligned}$$

Hence, for all  $u > 0$ ,

$$\begin{aligned}
\mathbb{P}_G[E_k] &\leq \mathbb{P}_G \left[ \max_{k \leq j} \left( -j - 2 \sum_{i=1}^j \hat{\xi}_i \right) \geq 0 \right] \\
&\leq \mathbb{P} \left[ \max_{k \leq j} \left( -j - 2 \sum_{i=1}^j \xi_i \right) \geq 0 \right] \\
&\leq \sum_{j=k}^{\infty} \mathbb{P} \left[ -2 \sum_{i=1}^j \xi_i \geq j \right] \\
&\leq \sum_{j=k}^{\infty} \frac{\mathbb{E} \left[ e^{-2u \sum_{i=1}^j \xi_i} \right]}{e^{uj}}, \text{ by Markov's inequality} \\
&\leq \sum_{j=k}^{\infty} e^{(-u+2\sigma^2 u^2)j}, \text{ by Assumption A}
\end{aligned}$$

and, by choosing  $u = 1/(4\sigma^2)$ ,

$$\mathbb{P}_G[E_k] \leq C e^{-k/(8\sigma^2)},$$

where  $C = \left(1 - e^{-1/(8\sigma^2)}\right)^{-1}$  is a positive constant. By symmetry, we obtain that :

$$\mathbb{P}_G[|\hat{M}_n - M| \geq k] \leq 2C e^{-k/(8\sigma^2)}. \quad (3.43)$$

If the design is (DD), the conclusion is straightforward, since  $|X_i - X_j| = \frac{|i-j|}{n}$ ,  $i, j = 1, \dots, n$ , and Theorem 3.8 is proved. ■

If the design is (RD), it is not clear how to go from (3.43) to an upper bound for the probability  $\mathbb{P}_G[|\hat{\theta}_n - \theta| \geq \epsilon]$ , for  $\epsilon > 0$ . This question should be addressed in a future work.

### Proof of Theorem 3.9

The proof is straightforward. Let  $G_1 = [0, 0]$  and  $G_2 = [0, 1/(2n)]$ . Then  $\mathbb{P}_{G_1} = \mathbb{P}_{G_2}$ , since no point of the design falls in  $G_1 \triangle G_2$ , and for any estimator  $\hat{G}_n$ ,

$$\begin{aligned} \sup_{G \in \mathcal{S}_0} \mathbb{E}_G [|\hat{G}_n \triangle G|] &\geq \mathbb{E}_{G_1} [|\hat{G}_n \triangle G_1|] + \mathbb{E}_{G_2} [|\hat{G}_n \triangle G_2|] \\ &\geq \mathbb{E}_{G_1} [|\hat{G}_n \triangle G_1| + |\hat{G}_n \triangle G_2|] \\ &\geq \mathbb{E}_{G_1} [|G_1 \triangle G_2|] \text{ by the triangle inequality} \\ &\geq \frac{1}{2n}. \end{aligned}$$

■

### Proof of Theorem 3.14

Let  $I_0$  be the set of even positive integers less or equal to  $n$ , and  $I_1$  the set of odd such integers. Note that  $\{X_i : i \in I_0\}$  is a deterministic and regular design, with step  $2/n$ . Let  $\hat{G}_n \in \text{ArgMax}_{G' \in \mathcal{S}} \sum_{i \in I_0} (2Y_i - 1)\mathbb{1}(X_i \in G')$  be the LSE estimator given in Theorem 3.11, built using only the subsample  $\{X_i : i \in I_0\}$ . Let  $x > 0$ , whose value will be specified in the course of the proof. Consider the event  $E_x = \{|\hat{G}_n \triangle G| < \frac{x \ln n}{n}\}$ . By Theorem 3.11, this event holds with probability at least  $1 - c_1 e^{-(c_2 x - 4) \ln n}$ . Choose  $x$  such that  $\frac{x \ln n}{n} \leq \mu/2$ . This choice implies that on the event  $E_x$ ,  $|\hat{G}_n \triangle G| < \mu \leq |G|$ , so necessary,  $\hat{G}_n$  and  $G$  must intersect. Thus, still on the event  $E_x$ ,

$$|\hat{G}_n \triangle G| = |\hat{b}_n - b| + |\hat{a}_n - a|,$$

where we denoted by  $G = [a, b]$  and  $\hat{G}_n = [\hat{a}_n, \hat{b}_n]$ . Let  $m = \frac{a+b}{2}$  and  $\hat{m}_n = \frac{\hat{a}_n + \hat{b}_n}{2}$  be, respectively, the middle points of  $G$  and  $\hat{G}_n$ . From now on, let us assume that  $E_x$  holds. Then,

$$\begin{aligned} |\hat{m}_n - m| &\leq \frac{1}{2}(|\hat{b}_n - b| + |\hat{a}_n - a|) \\ &\leq \frac{1}{2}|\hat{G}_n \triangle G| \\ &\leq \frac{x \ln n}{2n} \\ &\leq \frac{\mu}{4}. \end{aligned} \tag{3.44}$$

Therefore,  $\hat{m}_n \in G$  and, combining (3.44) with the fact that  $|G| \geq \mu$ ,

$$\min(\hat{m}_n, 1 - \hat{m}_n) \geq \frac{\mu}{4}. \tag{3.45}$$

Let us define

$$I_1^+ = \{i \in I_1 : X_i \geq \hat{m}_n\},$$

and

$$I_1^- = \{i \in I_1 : X_i \leq \hat{m}_n\}.$$

By (3.45),  $\#I_1^\epsilon \geq \frac{\mu n}{8} - 1 \geq \frac{\mu n}{16}$  for  $n$  large enough, and for  $\epsilon \in \{+, -\}$ . Note that  $\{X_i : i \in I_1^+\}$  (resp.  $\{X_i : i \in I_1^-\}$ ) is a deterministic and regular design of the segment  $[\hat{m}_n, 1]$  (resp.  $[0, \hat{m}_n]$ ), of cardinality greater or equal to  $\frac{\mu n}{16}$ , as we saw just before. Then, since we have both

$$Y_i = \mathbb{1}(X_i \leq b) + \xi_i, \forall i \in I_1^+$$

and

$$Y_i = \mathbb{1}(X_i \geq a) + \xi_i, \forall i \in I_1^-,$$

the change points  $a$  and  $b$  can be estimated as in Theorem 3.8, using the subsamples  $\{(X_i, Y_i) : i \in I_1^+\}$  and  $\{(X_i, Y_i) : i \in I_1^-\}$  respectively, and we get two estimators  $\tilde{a}_n$  and  $\tilde{b}_n$  which satisfy:

$$\mathbb{P}_G \left[ |\tilde{a}_n - a| \geq \frac{16y}{\mu n}, E_x \right] \leq c_0 e^{-y/(8\sigma^2)}$$

and

$$\mathbb{P}_G \left[ |\tilde{b}_n - b| \geq \frac{16y}{\mu n}, E_x \right] \leq c_0 e^{-y/(8\sigma^2)},$$

for all  $y > 0$ . Set  $\tilde{G}_n = [\tilde{a}_n, \tilde{b}_n]$ , on the event  $E_x$ , and  $\tilde{G}_n = \emptyset$  on its complementary  $\bar{E}_x$ . By setting  $x = \frac{\mu n}{2 \ln n}$ , which is the maximal value authorized in this proof,

$$\begin{aligned} \mathbb{P}_G \left[ |\tilde{G}_n \triangle G| \geq \frac{y}{n} \right] &\leq \mathbb{P}_G \left[ |\tilde{G}_n \triangle G| \geq \frac{y}{n}, E_x \right] + \mathbb{P}_G[\bar{E}_x] \\ &\leq 2c_0 e^{-\mu y/(256\sigma^2)} + c_1 n^4 e^{-c_2 \mu n/2}, \end{aligned}$$

which ends the proof of Theorem 3.14. ■

## 3.7 Appendix to Chapter 3

### 3.7.1 Lemmata

**Lemma 10** ([GMR95]). *Let  $r \geq d + 1$  be a positive integer. For any convex body  $G \in \mathcal{K}_d$  there exists a convex polytope  $P_r \in \mathcal{P}_r$  such that  $P_r \subseteq G$ , and*

$$|G \triangle P_r| \leq cd \frac{|G|}{r^{2/(d-1)}},$$

where  $c$  is a universal positive constant.

**Lemma 11.** *Let  $d \geq 1$ . Assume Model (RM) holds. Let  $G_1, G_2 \in \mathcal{K}_d^{(1)}$ . The Hellinger distance between  $\mathbb{P}_{G_1}$  and  $\mathbb{P}_{G_2}$  is equal to:*

$$H^2(\mathbb{P}_{G_1}, \mathbb{P}_{G_2}) = 2(1 - e^{-\frac{1}{8\sigma^2}})|G_1 \triangle G_2|.$$

**Lemma 12.** *Let  $S$  be the sphere with center  $a_0 = (1/2, \dots, 1/2) \in \mathbb{R}^d$  and radius  $1/2$ , and  $\rho$  the Euclidean distance in  $\mathbb{R}^d$ . We still denote by  $\rho$  its restriction on  $S$ . Let  $\eta \in (0, 1)$ . Then any  $\eta$ -packing family of  $(S, \rho)$  is finite, and any maximal  $\eta$ -packing family has a cardinality  $M_\eta$  that satisfies the inequalities*

$$\frac{d\sqrt{2\pi}}{2^{d-1}\sqrt{d+2}\eta^{d-1}} \leq M_\eta \leq \frac{4^{d-2}\sqrt{2\pi d}}{3^{(d-3)/2}\eta^{d-1}}. \quad (3.46)$$

### 3.7.2 Proof of the lemmata

**Proof of Lemma 9** Note that the event  $\{X_l - X_k > h\}$  is equivalent to  $\{\#(\mathcal{X} \cap (X_k, X_k + h)) < l - k\}$ . Let us denote by  $X'_1, \dots, X'_n$  the preliminary design, from which  $X_1, \dots, X_n$  is the reordered version. The random variables  $X'_1, \dots, X'_n$  are then i.i.d., with uniform distribution on  $[0, 1]$ . Hence,

$$\begin{aligned} \mathbb{P}[X_l - X_k > h] &= \sum_{j=1}^n \mathbb{P}[\#(\mathcal{X} \cap (X_k, X_k + h)) < l - k, X_k = X'_j] \\ &\leq \sum_{j=1}^n \mathbb{P}[\#(\mathcal{X} \cap (X'_j, X'_j + h)) < l - k] \\ &\leq \sum_{j=1}^n \mathbb{E}[\mathbb{P}[\#(\mathcal{X} \cap (X'_j, X'_j + h)) < l - k | X'_j]] \\ &\leq \sum_{j=1}^n \mathbb{E}[\mathbb{P}[n - \#(\mathcal{X} \cap (X'_j, X'_j + h)) \geq n - l + k | X'_j]] \\ &\leq \sum_{j=1}^n \mathbb{E}[f(X'_j)], \end{aligned} \quad (3.47)$$

where  $f(x) = \mathbb{P}[n - \#(\mathcal{X} \cap (x, x+h)) \geq n-l+k]$ , for all  $x \in [0, 1]$ . The random variable  $n - \#(\mathcal{X} \cap (x, x+h))$  is binomial with parameters  $n$  and  $1-h$ , and by Markov's inequality, for all  $u > 0$ ,

$$f(x) \leq e^{nu} (1 - h(1 - e^{-u}))^n e^{-u(n-l+k)}, \quad (3.48)$$

and (3.47) and (3.48) yield the lemma.  $\square$

**Proof of Lemma 12** The fact that any  $\eta$ -packing family of  $(S, \rho)$  is finite is clear and comes from the fact that  $S$  is compact. Consider now a maximal  $\eta$ -packing family of  $(S, \rho)$ , denoted by  $\{y_1, \dots, y_{M_\eta}\}$ . The surface area of  $B_d(y_j, \eta/2) \cap S$  is independent of  $j \in \{1, \dots, M_\eta\}$ , and we denote it by  $V(\eta/2)$ . A simple application of the Pythagorean theorem shows that  $B_d(y_j, \eta/2) \cap S$  is a cap of height  $\eta^2/4$  of  $S$ . Therefore, using Lemma 2.3 of [RSW01]

$$V(\eta/2) \geq \beta_{d-1} \left(1 - \frac{\eta^2}{4}\right)^{(d-3)/2} \eta^{d-1}.$$

Besides, since  $\{y_1, \dots, y_{M_\eta}\}$  is an  $\eta$ -packing family of  $(S, \rho)$ , the sets  $B_d(y_j, \eta/2) \cap S, j = 1, \dots, M_\eta$  are pairwise disjoint and the surface area of their union is less than the surface area of  $S$ , which is equal to  $\frac{d\beta_d}{2^{d-1}}$ , so we get

$$M_\eta V(\eta/2) \leq \frac{d\beta_d}{2^{d-1}}.$$

Therefore,

$$M_\eta \leq \frac{d\beta_d}{2^{d-1}V(\eta/2)} \leq \frac{d\beta_d}{2^{d-1}\beta_{d-1} \left(1 - \frac{\eta^2}{4}\right)^{(d-3)/2} \eta^{d-1}}.$$

and the right inequality of Lemma 12 follows from the fact that  $\eta^2/4 \leq 1/4$  and Lemma 2.2 of [RSW01] which states that

$$\frac{\sqrt{2\pi}}{\sqrt{d+2}} \leq \frac{\beta_d}{\beta_{d-1}} \leq \frac{\sqrt{2\pi}}{\sqrt{d}}. \quad (3.49)$$

The left inequality of Lemma 12 comes from the fact that any maximal  $\eta$ -packing family is an  $\eta$ -net. Indeed, consider a maximal  $\eta$ -packing family  $\mathcal{Y}$ , and assume it is not an  $\eta$ -net. Then there exists  $x \in S$  such that for all  $y \in \mathcal{Y}$ ,  $\rho(x, y) > \epsilon$ . Therefore  $\{x\} \cup \mathcal{Y}$  is an  $\eta$ -net that contains  $\mathcal{Y}$  strictly. This contradicts maximality of  $\mathcal{Y}$ . In particular, the family  $\{y_1, \dots, y_{M_\eta}\}$  is an  $\eta$ -net of  $S$ , and the caps  $B_d(y_j, \eta) \cap S, j = 1, \dots, M_\eta$  cover the sphere  $S$ .

It follows that

$$M_\eta V(\eta) \geq \frac{d\beta_d}{2^{d-1}}.$$

Using again Lemma 2.3 of [RSW01], we bound  $V(\eta)$  from above

$$V(\eta) \leq \beta_{d-1}\eta^{d-1},$$

and then the desired result follows again from (3.49). □

## Chapter 4

# Estimation of functionals

Instead of the set  $G$  itself, it is sometimes interesting to estimate its functionals, such as its volume, surface area, diameter, center of gravity, etc. For instance, in medical imaging, one may want to estimate the volume of a tumor rather than its boundary. A plug-in estimator of a functional  $S(G)$  of  $G$  is of the form  $S(\hat{G}_n)$ , where  $\hat{G}_n$  is a preliminary estimator of  $G$ . Even if  $\hat{G}$  is an optimal estimator of  $G$ , the plug-in estimator can be suboptimal. For example, in [KT93a], it is proved that the plug-in estimator of the volume of a boundary fragment can be improved, and the risk of the resulting estimator tends infinitely faster to zero. In [Gay97], a method consisting of modifying the plug-in estimator is proposed. First, the sample is divided into three subsamples of the same size. One is used to construct the plug-in estimator, and the two others are used to estimate its error. Although a convex body can be estimated at the speed  $n^{-2/(d+1)}$  in Model (DS), its volume can be estimated much faster. We give more details in Section 4.1.

In this chapter, we propose an overview of the known results about the estimation of a particular type of functionals. In particular, we cover estimation of the volume and the surface area, but not of the diameter, or the center of gravity. Let us recall Steiner formula for convex bodies (2.40). For  $K \in \mathcal{K}_d$ ,  $|K^\epsilon|$  is a polynomial of degree  $d$  in  $\epsilon$ ,  $\epsilon \geq 0$ :

$$|K^\epsilon| = \sum_{j=0}^d \beta_{d-j} v_j(K) \epsilon^j,$$

If  $j \geq 1$ ,  $v_j(K) = (\beta_{d-j})^{-1} L_j(K)$ , cf. (2.40). The coefficients  $v_j(K)$  are positive. They are called the intrinsic volumes of  $K$ . For  $j = 0$ ,  $v_0(K) = |K|$  is the volume of  $K$ , and  $v_1(K)$  is its surface area. The coefficient  $v_2(K)$  is called the mean width of  $K$ , and  $v_d(K) = 1$ . For  $j = 1, \dots, d-1$ , the intrinsic volume  $v_{d-j}(K)$  is the mean volume of projections of  $K$  on



$j$ -dimensional linear subspaces of  $\mathbb{R}^d$ :

$$v_{d-j}(K) = \frac{\binom{d}{j}\beta_d}{\beta_j\beta_{d-j}} \int_{\mathcal{L}_j^d} |\pi_L(K)|_j \mu_j(dL), \quad (4.1)$$

where  $\mathcal{L}_j^d$  is the Grassmannian of all  $j$ -dimensional linear subspaces of  $\mathbb{R}^d$ , equipped with the Haar probability measure  $\mu_j$  and  $\pi_L$  is the orthogonal projection onto  $L$ , for  $L \in \mathcal{L}_j^d$ . In order to simplify the notation, we denote by  $|\cdot|_j$  the  $j$ -dimensional Lebesgue measure on any  $L \in \mathcal{L}_j^d$ . In this Chapter, we are interested in the estimation of the intrinsic volumes, in both models (DS) and (RM).

## 4.1 The density support model

### 4.1.1 Estimation of the volume of a convex body

As we mentioned above, Gayraud [Gay97] has proved that the volume of a convex body, in Model (DS), can be estimated at a better speed than the convex body itself, under certain conditions. First, let us note that it is clear that the plug-in estimator associated to the convex hull estimator  $\hat{K}_n$  achieves the same speed of convergence as the estimator  $\hat{K}_n$ . Indeed, for  $K \in \mathbb{R}^d$ , since  $\hat{K}_n \subseteq K$   $\mathbb{P}_K$ -almost surely,  $\left| |K| - |\hat{K}_n| \right| = |K \setminus \hat{K}_n| = |K \triangle \hat{K}_n|$ . Therefore,  $\sup_{K \in \mathcal{K}_d^{(1)}} \mathbb{E}_K \left[ \left| |K| - |\hat{K}_n| \right| \right]$  is of the order  $n^{-\frac{2}{d+1}}$  and the plug-in estimator achieves the same rate on the class  $\mathcal{K}_d^{(1)}$  as the convex hull estimator. On the classes  $\mathcal{P}_r^{(1)}$ ,  $r \geq d+1$ , it is not clear that the risk of the plug-in estimator has the same rate of convergence as that of  $\hat{P}_n^{(r)}$ , but at least it is not worse, since, if  $P \in \mathcal{P}_r^{(1)}$ , we have  $\left| |P| - |\hat{P}_n^{(r)}| \right| \leq |P \triangle \hat{P}_n^{(r)}|$  by the triangle inequality. The inclusion  $\hat{P}_n^{(r)} \subseteq P$  does not necessarily hold, and this is why the last inequality may be strict, so that  $\mathbb{E}_P[| |P| - |\hat{P}_n^{(r)}| |]$  may converge to zero faster than  $\mathbb{E}_P[|P \triangle \hat{P}_n^{(r)}|]$ . The quite simple case of triangles in the plane ( $d=2, r=3$ ), is addressed in the next section.

We measure the error of an estimator of a functional using the expected absolute value of the difference between the estimator and its target. For convex bodies, the following theorem holds.

**Theorem 4.1.** *There exists an estimator  $\hat{V}_n$  of the volume, whose minimax risk on the class  $\mathcal{K}_d^{(1)}$  is of a smaller order than  $n^{-(d+3)/(2d+2)}$ .*

$$\sup_{K \in \mathcal{K}_d^{(1)}} \mathbb{E}_K \left[ \left| \hat{V}_n - |K| \right| \right] \leq c(d) n^{-\frac{d+3}{2d+2}}, \quad \forall n \geq n_0(d),$$

where  $n_0(d)$  is a positive integer and  $c(d)$  a positive constant, which depend on  $d$  only.

In [Gay97], this upper bound is proved if the supremum is taken over all  $K \in \mathcal{K}_d^{(1)}$  such that  $K$  contain a given Euclidean ball of positive volume. This assumption can be dropped, and Theorem 4.1 is a direct consequence of the proof of the same upper bound given in [Gay97] and of our Theorem 2.13. A lower bound of the minimax risk is also given in [Gay97], whose rate matches the upper bound in Theorem 4.1. This proves that the minimax rate of convergence on  $\mathcal{K}_d^{(1)}$ , for the estimation of the volume, is  $n^{-(d+3)/(2d+2)}$ , and that it is achieved by  $\hat{V}_n$ . Let us present how the estimator  $\hat{V}_n$  is constructed in [Gay97].

Let  $K \in \mathcal{K}_d^{(1)}$ . Assume without loss of generality that three independent samples  $\mathcal{X}_1 = \{X_1, \dots, X_n\}$ ,  $\mathcal{X}_2 = \{X'_1, \dots, X'_n\}$  and  $\mathcal{X}_3 = \{X''_1, \dots, X''_n\}$  of i.i.d. random variables uniformly distributed in  $K$  are available. Otherwise, we can divide the initial sample into three subsamples. From  $\mathcal{X}_1$ , we construct  $\hat{K}_n$ , the convex hull estimator. The error of the corresponding plug-in estimator  $|\hat{K}_n|$  is

$$\begin{aligned} \left| |K| - |\hat{K}_n| \right| &= \int_K \mathbf{1}(x \notin \hat{K}_n) dx \\ &= \mathbb{E}_K \left[ \left| |K| \mathbf{1}(X \notin \hat{K}_n) \right| \mathcal{X}_1 \right]. \end{aligned} \quad (4.2)$$

We estimate this error using the samples  $\mathcal{X}_2$  and  $\mathcal{X}_3$ . For this purpose, replace, in (4.2), the volume  $|K|$  by the volume  $|\hat{K}'_n|$  of the convex hull of  $\mathcal{X}_2$ , and the conditional expectation by its empirical version. Let  $\hat{f}_n$  be a kernel estimator of the density of the observations constructed using the subsample  $\mathcal{X}_2$ . The estimator  $\hat{V}_n$  is defined as

$$\hat{V}_n = |\hat{K}_n| + \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X''_i \notin \hat{K}_n) \min \left( 2, \frac{1}{\hat{f}_n(X''_i)} \right). \quad (4.3)$$

The same construction is used in order to estimate functionals of the form  $\int_K \phi$ , where  $\phi$  is a real valued function defined on  $\mathbb{R}^d$ . If  $\phi$  is the constant unit function, then this is the volume. Combinations of such functionals are also estimated, to get, for example, the center of gravity of  $K$ . In [Gay97], the density of the observations is not assumed to be uniform, but only separated from zero on its support  $K$ . More precisely, let  $f$  be the unknown density of the observations, and assume it satisfies  $f(x) \geq a_0, \forall x \in K$ , where  $a_0$  is a given positive number, and  $K \in \mathcal{K}_d^{(1)}$  is the unknown support. The estimator  $V_n$  of  $|K|$  is then given by

$$\hat{V}_n = |\hat{K}_n| + \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X''_i \notin \hat{K}_n) \min \left( \frac{2}{a_0}, \frac{1}{\hat{f}_n(X''_i)} \right).$$

Surprisingly, if the density of the observations is known to be uniform, only its support being

unknown, it is still better to keep an estimator  $\hat{f}_n$  of the density in (4.3), than to replace the term  $\min\left(2, \frac{1}{\hat{f}_n(X_i'')}\right)$  by  $|\hat{K}_n'|$ . This comes from the fact that the density of the uniform distribution on  $K$  is very smooth on  $K$  - since it is constant -, and thus can be estimated at a good speed. Let us define a similar estimator to  $\hat{V}_n$ , where we replace the term  $1/\hat{f}_n(X_i'')$  by  $|\hat{K}_n'|$ :

$$\tilde{V}_n = |\hat{K}_n| + \frac{|\hat{K}_n'|}{n} \sum_{i=1}^n \mathbb{1}(X_i'' \notin \hat{K}_n).$$

The advantage of this estimator is that its weighted risk can be bounded uniformly on the whole class  $\mathcal{K}_d$  of convex bodies. However, the rate of convergence of its weighted risk is minimax only for  $d \leq 5$ .

**Theorem 4.2.** *Assume Model (DS). The weighted risk of  $\tilde{V}_n$  satisfies:*

$$\sup_{K \in \mathcal{K}_d} \mathbb{E}_K \left[ \frac{|\tilde{V}_n - |K||}{|K|} \right] \leq c(d) \max\left(n^{-\frac{4}{d+1}}, n^{-\frac{d+3}{2d+2}}\right), \forall n \geq 1,$$

where  $c(d)$  is a positive constant which depends on  $d$  only.

As soon as  $d \geq 6$ , the upper bound is of the order  $n^{-\frac{4}{d+1}}$ , but we believe this is not the minimax rate of convergence.

*Proof.* Let us write

$$\hat{V}_n - |K| = \frac{1}{n} \sum_{i=1}^n (\rho_i + \tau_i),$$

where  $\rho_i = |K| \mathbb{1}(X_i'' \notin \hat{K}_n) - |K \setminus \hat{K}_n|$  and  $\tau_i = -|K \setminus \hat{K}_n'| \mathbb{1}(X_i'' \notin \hat{K}_n)$ . Let us bound from above  $\mathbb{E}_K \left[ (\hat{V}_n - |K|)^2 \right]$ . By the Cauchy-Schwarz inequality, this will provide an upper bound for  $\mathbb{E}_K \left[ \left| \hat{V}_n - |K| \right|^2 \right]$ . First, we have

$$\mathbb{E}_K \left[ (\hat{V}_n - |K|)^2 \right] \leq 2\mathbb{E}_K \left[ \left( \frac{1}{n} \sum_{i=1}^n \rho_i \right)^2 \right] + 2\mathbb{E}_K \left[ \left( \frac{1}{n} \sum_{i=1}^n \tau_i \right)^2 \right]. \quad (4.4)$$

Let us condition the first expectation by the subsample  $\mathcal{X}_1$ . Conditionally on  $\mathcal{X}_1$ , the  $\rho_i$  have

zero mean and are i.i.d. Thus,

$$\begin{aligned}\mathbb{E}_K \left[ \left( \frac{1}{n} \sum_{i=1}^n \rho_i \right)^2 \middle| \mathcal{X}_1 \right] &= \frac{1}{n} \mathbb{E}_K [\rho_1^2 | \mathcal{X}_1] \\ &\leq \frac{|K|^2}{n} \mathbb{E}_K [\mathbf{1}(X_1'' \notin \hat{K}_n) | \mathcal{X}_1] \\ &\leq \frac{|K|^2}{n} \frac{|K \setminus \hat{K}_n|}{|K|}.\end{aligned}$$

Therefore,

$$\mathbb{E}_K \left[ \left( \frac{1}{n} \sum_{i=1}^n \rho_i \right)^2 \right] \leq c_1 |K|^2 n^{-\frac{d+3}{d+1}}, \quad (4.5)$$

by Corollary 2.2, where  $c_1$  is a positive constant which depends on  $d$  only.

Let us now condition the second expectation in (4.4) by both subsamples  $\mathcal{X}_1$  and  $\mathcal{X}_2$ . First, the expression under the expectation sign is

$$\left( \frac{1}{n} \sum_{i=1}^n \tau_i \right)^2 = \frac{1}{n^2} \left( \sum_{i=1}^n \tau_i^2 + 2 \sum_{1 \leq i < j \leq n} \tau_i \tau_j \right),$$

so that

$$\begin{aligned}\mathbb{E}_K \left[ \left( \frac{1}{n} \sum_{i=1}^n \tau_i \right)^2 \middle| \mathcal{X}_1 \cup \mathcal{X}_2 \right] &= \frac{1}{n} \mathbb{E}_K [\tau_1^2 | \mathcal{X}_1 \cup \mathcal{X}_2] + \frac{n-1}{n} \mathbb{E}_K [\tau_1 \tau_2 | \mathcal{X}_1 \cup \mathcal{X}_2] \\ &= \frac{1}{n} \frac{|K \setminus \hat{K}'_n|^2 |K \setminus \hat{K}_n|}{|K|} + \frac{n-1}{n} \frac{|K \setminus \hat{K}'_n|^2 |K \setminus \hat{K}_n|^2}{|K|^2},\end{aligned}$$

and then, using again Corollary 2.2,

$$\mathbb{E}_K \left[ \left( \frac{1}{n} \sum_{i=1}^n \tau_i \right)^2 \right] \leq c_2 |K|^2 \left( n^{-\frac{d+7}{d+1}} + n^{-\frac{8}{d+1}} \right), \quad (4.6)$$

where  $c_2$  is a positive constant which depends on  $d$  only.

Hence, by (4.4), (4.5) and (4.6),

$$\mathbb{E}_K \left[ (\hat{V}_n - |K|)^2 \right] \leq c |K|^2 \max \left( n^{-\frac{d+3}{d+1}}, n^{-\frac{8}{d+1}} \right).$$

The constant  $c$  does not depend on  $K$ , and Theorem 4.2 is proved.  $\square$

### 4.1.2 Estimation of the area of a triangle

Let us focus on the case  $d = 2$ , and assume that  $K = T$  is a triangle in the plane. We will show that the plug-in estimator  $|\hat{P}_n^{(3)}|$ , which is the area of the smallest triangle containing the whole sample, attains the rate  $1/n$ , with no extra logarithmic factor. Let  $T$  be a triangle, and let  $A, B, C$  be its vertices. Let  $\epsilon$  and  $\lambda$  in  $(0, 1)$ . At each vertex, we define three new points, as on Figure 4.1. At vertex  $A$ , for example: Let  $A'_1 \in [A, B]$  such that  $\rho(A, A'_1) = \lambda\rho(A, B)$ , and let  $A_1 \in T$ , such that the lines  $(A'_1 A_1)$  and  $(AC)$  are parallel, and such that  $\rho(A'_1, A_1) = \epsilon\rho(A, C)$ . Denote by  $R_{A,1}$  the parallelogram with  $A, A'_1$  and  $A_1$  as vertices. Similarly, define  $R_{A,2}, R_{B,1}, R_{B,2}, R_{C,1}$  and  $R_{C,2}$ . These six subsets of  $T$  have the same area  $2\epsilon\lambda|T|$ , and the intersection of  $R_{M,1}$  and  $R_{M,2}$ , for  $M \in \{A, B, C\}$ , has area  $2\epsilon^2|T|$ . Denote by  $R_1, \dots, R_6$  these six parallelograms. Denote by  $A'$  the intersection point of the lines  $(A_1 B_2)$  and  $(A_2 C_1)$ , by  $B'$  that of the lines  $(A_1 B_2)$  and  $(B_1 C_2)$ , and by  $C'$  that of the lines  $(B_1 C_2)$  and  $(C_1 A_2)$ , as on Figure 4.1.

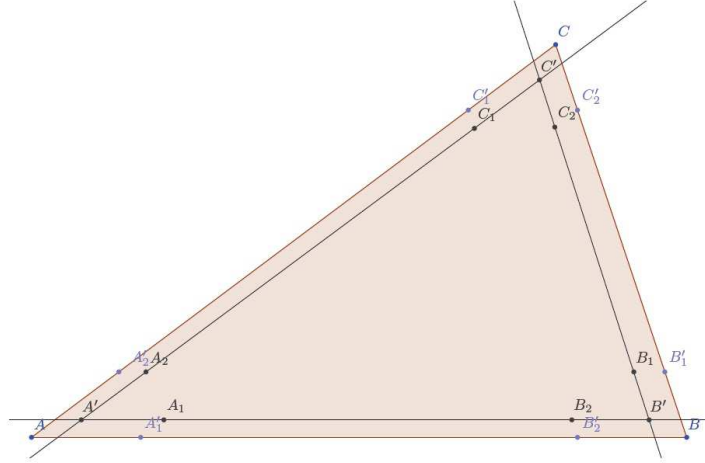


Figure 4.1: Construction for the estimation of the area of a triangle

If each of the  $R_j, j = 1, \dots, 6$  meets the sample, i.e., if for each  $j = 1, \dots, 6$ , there exists  $i \in \{1, \dots, n\}$ , such that  $X_i \in R_j$ , then  $\hat{P}_n^{(3)}$  contains  $A_1, A_2, B_1, B_2, C_1$  and  $C_2$ . Therefore, its area is larger than the area of the smallest triangle containing those six points. If  $\lambda \leq 1/3$ , then the smallest triangle containing  $A_1, A_2, B_1, B_2, C_1$  and  $C_2$  is the triangle with vertices  $A', B'$  and  $C'$ . Its area is greater or equal to  $|T| - 6\epsilon|T|$ . Conversely, if  $|\hat{P}_n^{(3)}| \leq |T| - 6\epsilon|T|$ , then one

of the  $R_j, j = 1, \dots, 6$  does not meet the sample, and

$$\begin{aligned}\mathbb{P}_T \left[ |\hat{P}_n^{(3)}| \leq |T| - 6\epsilon|T| \right] &\leq \sum_{j=1}^6 \mathbb{P}_T [X_i \notin R_j, \forall i = 1, \dots, n] \\ &\leq 6(1 - 2\epsilon\lambda)^n \\ &\leq 6e^{-2\lambda\epsilon n}.\end{aligned}$$

Note that, by definition,  $\left| |\hat{P}_n^{(3)}| - |T| \right| = |T| - |\hat{P}_n^{(3)}|$ . Taking  $\lambda = 1/3$  yields the following theorem:

**Theorem 4.3.** *Let  $d = 2$ . Let Model (DS) hold, with  $G = T$  being a triangle in  $\mathbb{R}^2$ , and let  $\hat{P}_n^{(3)}$  be the estimator defined in (2.6). Then,*

$$\mathbb{P}_T \left[ n \frac{\left| |\hat{P}_n^{(3)}| - |T| \right|}{|T|} \geq x \right] \leq 6e^{-x/9}, \forall x \geq 0.$$

This theorem yields the corollary:

**Corollary 4.1.** *Let  $d = 2$ . Let Model (DS) hold. Then,*

$$\sup_{T \in \mathcal{P}_3} \mathbb{E}_T \left[ \left( \frac{\left| |\hat{P}_n^{(3)}| - |T| \right|}{|T|} \right)^q \right] \leq \frac{6q9^q(q-1)!}{n^q}.$$

*In particular, the minimax weighted risk for the estimation of the area of a triangle in Model (DS) is of the order  $1/n$ .*

### 4.1.3 Estimation of intrinsic volumes

In order to estimate intrinsic volumes  $v_j(K)$  of a convex body  $K$ , for  $j = 1, \dots, d-1$ , (4.1) suggests to estimate the volumes of the projections of  $K$  on linear subspaces of  $\mathbb{R}^d$ . Assume that Model (DS) holds. If  $L \in \mathcal{L}_d^j$ , then  $\pi_L(X_1), \dots, \pi_L(X_n)$  are i.i.d. random points in  $L$ , with a distribution that is supported on  $\pi_L(K)$ . This distribution admits a density with respect to the  $j$ -dimensional Lebesgue measure on  $L$ , and this density is concave on  $\pi_L(K)$ . In addition, the convex hull of  $\pi_L(X_1), \dots, \pi_L(X_n)$  is equal to  $\pi_L(\hat{K}_n)$ , and its volume  $|\pi_L(\hat{K}_n)|_j$  is the plug-in estimator of  $|\pi_L(K)|_j$ . The plug-in estimator of  $v_j(K)$  is defined as

$$\hat{v}_j = v_j(\hat{K}_n) = \frac{\binom{d}{j} \beta_d}{\beta_j \beta_{d-j}} \int_{\mathcal{L}_d^j} |\pi_L(\hat{K}_n)|_j \mu_j(dL).$$

The risk of the plug-in estimator has been computed in [Bà92] when the boundary of  $K$  is three times continuously differentiable with everywhere positive Gauss curvature, in [Rei04] when it is twice differentiable with everywhere positive curvature, and in [BHH08] under weaker assumptions on  $\partial K$ . We say that a ball rolls freely inside  $K$  if and only if there exists some positive number  $r$ , such that for each point  $x \in \partial K$ , there exists a Euclidean ball  $B$  of radius  $r$ , containing  $x$  and included in  $K$ . In particular, if  $\partial K$  is twice differentiable, then a ball rolls freely in  $K$  (cf. [Lei98]). For  $j = 1, \dots, d-1$  and  $x \in \partial K$ , denote by  $\sigma_j(x)$  the  $j$ -th normalized elementary symmetric function of the principal curvatures of  $\partial K$  at point  $x$  (see [Sch93a, Section 2.5, p. 106]). The following result holds:

**Theorem 4.4** ([BHH08]). *Let  $K \in \mathcal{K}_d$ , in which a ball rolls freely. Then, for  $j = 1, \dots, d-1$ ,*

$$\lim_{n \rightarrow \infty} \left( \frac{n}{|K|} \right)^{\frac{2}{d+1}} \mathbb{E}_K \left[ v_j(K) - v_j(\hat{K}_n) \right] = c_{d,j} \int_{\partial K} \sigma_{d-1}(x)^{\frac{1}{d+1}} \sigma_j(x) d\mu(x),$$

where  $\mu$  is the Lebesgue measure on  $\partial K$  and  $c_{d,j}$  is a positive constant which depends on  $d$  and  $j$  only.

Note also that in [BFV10], the variance of  $v_j(\hat{K}_n)$  is shown to be of the order  $n^{-(d+3)/(d+1)}$  under the same assumptions as in [Rei04].

The condition of a freely rolling ball in  $K$  allows for flat parts on  $\partial K$ , i.e., the intersection of  $K$  with a supporting hyperplane can have a  $(d-1)$ -dimensional affine hull. This would not be allowed by the assumptions of theorems in [Bà92] or [Rei04]. However, the freely rolling ball condition does not allow for situations where  $\partial K$  has a corner. In particular, Rényi and Sulanke [RS63] showed that if  $K$  is a square in the plane ( $d=2$ ), then the expected perimeter of  $\hat{K}_n$  makes a difference with that of the square of the order of  $n^{-1/2}$ , which is much larger than the  $n^{-2/3}$  provided by the theorem above. It is shown (cf. [BHH08]) that if  $j < d/2$ , there exists  $K \in \mathcal{K}_d$ , whose boundary is infinitely many times differentiable everywhere except at one point of  $\partial K$  where it is only continuously differentiable, with positive Gauss curvature everywhere on  $\partial K$ , and such that

$$\lim_{n \rightarrow \infty} \left( \frac{n}{|K|} \right)^{\frac{2}{d+1}} \mathbb{E}_K \left[ v_j(K) - v_j(\hat{K}_n) \right] = \infty.$$

This shows that the rate of the risk of the plug-in estimator of  $v_j(K)$  is not bounded from above by  $n^{-2/(d+1)}$ , uniformly on the class  $\mathcal{K}_d^{(1)}$ . On the class  $\mathcal{K}_d^{(1,+)}$  of such  $K \in \mathcal{K}_d^{(1)}$  in which a ball rolls freely, it is not clear if this estimator achieves the rate  $n^{-2/(d+1)}$  uniformly, since the integral in the limit in Theorem 4.4 may be arbitrarily large. It seems that some additional assumptions should be added on that class to guarantee that the risk of the plug-in estimator

is bounded from above by  $n^{-2/(d+1)}$  up to a constant factor independent of  $K$ .

As for the estimation of the volume, a natural question is whether the plug-in estimator of the  $j$ -th intrinsic volume can be improved. In particular, it seems that a similar technique as for the plug-in estimator of the volume can be used. The error of the plug-in estimator of the  $j$ -th intrinsic volume  $v_j(K)$  is equal to:

$$v_j(K) - v_j(\hat{K}_n) = \frac{\binom{d}{j}\beta_d}{\beta_j\beta_{d-j}} \int_{\mathcal{L}_j^d} |\pi_L(K) \setminus \pi_L(\hat{K}_n)|_j \mu_j(dL).$$

As before, assume without loss of generality that we have three independent samples  $\mathcal{X}_1 = \{X_1, \dots, X_n\}$ ,  $\mathcal{X}_2 = \{X'_1, \dots, X'_n\}$  and  $\mathcal{X}_3 = \{X''_1, \dots, X''_n\}$  of i.i.d. random variables uniformly distributed in  $K$ . Let  $\hat{K}_n$  be the convex hull of the first sample. Let  $L \in \mathcal{L}_j^d$ . The distribution of the variables  $\pi_L(X'_i)$ ,  $i = 1, \dots, n$  is not uniform on  $\pi_L(K)$  and their density denoted by  $f_L$  is concave on its support  $\pi_L(K)$ . Precisely,  $f_L(x) = |x + (K \cap L^\perp)|_{d-j} / |K|$ ,  $x \in L$ , where  $L^\perp$  is the orthogonal linear subspace to  $L$ . Let  $\hat{f}_{n,L}$  be an estimator of  $f_L$ , based on the sample  $\mathcal{X}_2$ . Similarly to (4.3), one could define an estimator of  $v_j(K)$  as:

$$\tilde{v}_j = v_j(\hat{K}_n) + \int_{\mathcal{L}_j^d} \sum_{i=1}^n \frac{1}{\hat{f}_{n,L}(X''_i)} \mathbb{1}(X''_i \notin \hat{K}_n) dL. \quad (4.7)$$

Note that  $f_L$  decreases to zero near the boundary of  $\pi_L(K)$ , and therefore there is no  $a_0 > 0$  such that  $f_L(x) \geq a_0$ ,  $\forall x \in \pi_L(K)$ . Thus, the technique used in [Gay97] needs to be adapted here. The risk of the estimator  $\tilde{v}_j$  depends on the choice of the estimators  $\hat{f}_{n,L}$ ,  $L \in \mathcal{L}_j^d$ . This problem will be a subject of future work.

## 4.2 The regression model

To our knowledge, in a regression setup, the previous works considered mainly the estimation of the volume [KT93b] and of the surface area of a convex body. In [CFRC07], the surface area of a set  $G \in [0, 1]^d$  is interpreted as its *Minkowski content*:

$$v_0(G) = \lim_{\epsilon \rightarrow 0} \frac{|(\partial G)^\epsilon|}{2\epsilon}, \quad (4.8)$$

provided that this limit exists and is finite. If  $G$  is a convex body, the Minkowski content of  $G$  is well defined. Cuevas, Fraiman and Rodriguez-Casal [CFRC07] propose a consistent estimator of  $v_0(G)$ , using observations generated by Model (RM), which they assume to be noise free, i.e.,  $\xi_i = 0$ ,  $\forall i = 1, \dots, n$ . At each point of the design  $X_1, \dots, X_n$ , the perfect label  $\mathbb{1}(X_i \in G)$  is



observed. These authors cover a broader class of sets than convex bodies, but let us focus here on the class  $\mathcal{K}_d^{(1)}$ . If  $G \in \mathcal{K}_d^{(1)}$ , then the limit in (4.8) exists, is finite, and is equal to the surface area of  $G$ . The estimator of  $v_0(G)$  is defined as follows. Let  $\epsilon_n$  be a positive sequence. Let  $P_n = \{X_i : Y_i = 1\}$  be the set of all points in the design which are labeled 1, i.e., which belong to  $G$ , and let  $N_n = \{X_i : Y_i = 0\}$  be all the other points. Let  $\hat{T}_n$  be the set of all points  $z \in \mathbb{R}^d$  such that both  $B_2^d(z, \epsilon_n) \cap P_n$  and  $B_2^d(z, \epsilon_n) \cap N_n$  are nonempty. The points in  $\hat{T}_n$  are close to the boundary  $\partial G$  in the sense that  $\hat{T}_n \subseteq (\partial G)^{\epsilon_n}$ . The proposed estimator of  $v_0(G)$  is defined as

$$\hat{v}_n = \frac{|\hat{T}_n|}{2\epsilon_n}. \quad (4.9)$$

Note that the computation of this estimator does not require that of a preliminary estimator of  $G$ . The sequence  $\epsilon_n$  should be chosen so it converges to zero not too fast. In particular, it is shown in [CFRC07] that if  $\epsilon_n = n^{-1/(2d)}$  then, under some conditions on  $G$ ,

$$\mathbb{E}_G[|\hat{v}_n - v_0(G)|] \leq c(G, d)n^{-1/(2d)}, \quad (4.10)$$

where the positive constant  $c(G, d)$  depends on  $d$ , and on  $G$ . The conditions on  $G$  is that its boundary should not have neither inner, nor outer peaks: there should exist a positive constant  $C$ , such that for small enough  $\epsilon > 0$ ,

$$\left\{ \begin{array}{l} \frac{|B_2^d(z, \epsilon) \cap G|}{|G|} \geq C|B_2^d(z, \epsilon)|, \\ \text{and} \\ \frac{|B_2^d(z, \epsilon) \setminus G|}{1 - |G|} \geq C|B_2^d(z, \epsilon)|, \end{array} \right. \quad (4.11)$$

for all  $z \in \partial G$ .

These results on estimation of the Minkowski content - or surface area -  $v_0(G)$  are not embedded in a minimax setup. In particular, the constants that enter into consideration in (4.10) depend on  $G$  and it is not clear if such a bound can be given uniformly on the class of sets in  $\mathcal{K}_d^{(1)}$  satisfying (4.11). It is not clear whether the same procedure can be used if the labels  $Y_i$  are contaminated by some noise. In that case, an idea would be to define  $P_n = \{X_i : Y_i \geq 1/2\}$  and  $N_n = \{X_i : Y_i \leq 1/2\}$ . The set  $T_n$  would be defined as the set of all points  $z \in \mathbb{R}^d$  such that the cardinality of  $B_2^d(z, \epsilon_n) \cap P_n$  is at least  $a_n$  and that of  $B_2^d(z, \epsilon_n) \cap N_n$  is at least  $b_n$ , for some real sequences  $a_n$  and  $b_n$  to be chosen.

## Chapter 5

# Conclusion, contributions

This thesis dealt with estimation of convex bodies, and more specifically of convex polytopes. Two setups were addressed. The first one consists in estimating the support of the density of  $n$  observed random points. In most of the cases, we assumed that the density was uniform. In the second setup, a set of  $n$  design points is given. To each of these points, the label 1 is assigned if the point belongs to the unknown set and 0 otherwise. The labels are not observed directly but a contaminated version of them with some noise is available. In both setups, the unknown sets were assumed to be convex bodies. We focused on the case of convex polytopes when the number of vertices is known. In this case, we proposed estimators and proved deviation inequalities for their risks when the accuracy is measured using the Nikodym distance. The constants in these deviation inequalities do not depend on the unknown set, which makes these inequalities uniform on the considered classes of sets. These deviation inequalities provide upper bounds on the risks of the corresponding estimators. On the other hand, we proved lower bounds for the minimax risks on the considered classes. For the class of convex polytopes with given number of vertices, we showed that the minimax rate of convergence is between  $1/n$  and  $(\ln n)/n$ . Although we showed that the logarithmic factor is unavoidable in the regression setup, we still do not know if it is also the case in the density support setup. Our conjecture is that the right rate of convergence is the parametric rate  $1/n$ . In the one-dimensional case, this conjecture is true. For higher dimensions, the question remains open. In the regression setup, in dimension one, we proved that the logarithmic factor which appears in the minimax rate, can be dropped if either the unknown set is not too small, i.e., its length is greater than a given positive number, or it contains a known and given point.

The deviation inequalities also allow us to construct estimators which adapt to the number of vertices of the unknown set when it is assumed to be a polytope. In particular, our adaptive

estimators achieve the same rates of convergence as the previous estimators that required the knowledge of the true number of vertices. In addition, these estimators achieve the minimax rate of convergence if the true set is not necessarily a polytope, but a general convex body.

In the density support setup, we were particularly interested in the estimator which is the convex hull of the available observations. This random polytope has attracted much attention in the literature during the last seventy years. We proved a new deviation inequality for this random polytope, uniformly on the class of all  $d$ -dimensional convex bodies. This inequality yields upper bounds on all the moments of the missing volume of the random polytope. We also proved upper bounds on the moments of the number of vertices of this random polytope. These bounds are uniform on the class of all  $d$ -dimensional convex bodies and we showed that they are tight.

In the regression setup, in dimension higher than 1, we have conjectured that the minimax rate of convergence on the class of polytopes with known number of vertices, and whose volume is separated from zero, is  $1/n$ . In other words, if we add a restriction on the volume of the unknown polytope, the logarithmic factor can be dropped. This seems to be a reasonable extension of the one dimensional case.

Estimation of functionals, such as intrinsic volumes, is still a new topic, and very few results are known. We provided some refinements of the upper bounds in estimation of the volume, a discussion of open problems and some ideas to construct new estimators.

# Bibliography

- [Bà92] I. Bárány. Random polytopes in smooth convex bodies. *Mathematika*, 39:81–92, 1992.
- [Bà01] I. Bárány. A note on Sylvester’s four-point problem. *Studia Math. Hungarica*, 38:73–77, 2001.
- [BB93] I. Bárány and C. Buchta. Random polytopes in a convex polytope, independence of shape, and concentration of vertices. *Math. Ann.*, 297:467–497, 1993.
- [BBSW07] A. J. Baddeley, I. Bárány, R. Schneider, and W. Weil. *Stochastic Geometry, C.I.M.E. Summer School, Martina Franca, Italy (W. Weil, ed.), Lecture Notes in Mathematics*. Springer, 2007.
- [BC01] A. Baillo and A. Cuevas. On the estimation of a star-shaped set. *Advances in Applied Probability*, 33:717–726, 2001.
- [BCJ00] A. Baillo, A. Cuevas, and A. Justel. Set estimation and nonparametric detection. *Canad. J. Statist.*, 28:765–782, 2000.
- [BFV10] I. Bárány, F. Fodor, and V. Vigh. Intrinsic volumes of inscribed random polytopes in smooth convex bodies. *Adv. Appl. Prob.*, 42:605–619, 2010.
- [BHH08] K. J. Böröczky, L. M. Hoffmann, and D. Hug. Expectation of intrinsic volumes of random polytopes. *Periodica Mathematica Hungarica*, 57:143–164, 2008.
- [BL88] I. Bárány and D. G. Larman. Convex bodies, economic cap coverings, random polytopes. *Mathematika*, 35:274–291, 1988.
- [BL92] L. D. Brown and M. G. Low. Superefficiency and lack of adaptability in functional estimation. *Technical report, Cornell Univ.*, 1992.
- [Bla23] W. Blaschke. *Vorlesungen über Differentialgeometrie II*. Springer-Verlag, 1923.

- [BM71] H. Bruggesser and P. Mani. Shellable decompositions of cells and spheres. *Math. Scand.*, 29:197–205, 1971.
- [BMM10] A. Beresteanu, I. Molchanov, and F. Molinari. Partial identification using random set theory. *cemmap working paper, No. CWP40/10*, 2010.
- [Bro76] E. M. Bronshtein.  $\epsilon$ -entropy of convex sets and functions. *Siberian Mathematical Journal*, 17:303–398, 1976.
- [Bro83] A. Bronsted. *An Introduction to Convex Polytopes*. Springer-Verlag, 1983.
- [Bru13] V.-E. Brunel. Adaptive estimation of convex polytopes and convex sets from noisy data. *Electronic Journal of Statistics*, 7:1301–1327, 2013.
- [Bru14a] V.-E. Brunel. Adaptive estimation of convex and polytopal density support. *Submitted, arXiv:1309.6602*, 2014.
- [Bru14b] V.-E. Brunel. Convex set detection. *Submitted, arXiv:1404.6224*, 2014.
- [Bru14c] V.-E. Brunel. A universal deviation inequality for random polytopes. *submitted, arXiv:1311.2902*, 2014.
- [CF97] A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *The Annals of Statistics*, 25:2300–2312, 1997.
- [CF10] A. Cuevas and R. Fraiman. *Set estimation, in New Perspectives on Stochastic Geometry, Kendall, W.S. and Molchanov, I., pp. 374-397*. Oxford University Press, 2010.
- [CFRC07] A. Cuevas, R. Fraiman, and A. Rodríguez-Casal. A nonparametric approach to the estimation of lengths and surface areas. *The Annals of Statistics*, 35:1031–1051, 2007.
- [Che76] J. Chevalier. Estimation du support et du contour du support d’une loi de probabilité. *Ann. Inst. H. Poincaré Sec. B*, 12:339–364, 1976.
- [Chi12] M. Chichignoud. Performances statistiques d’estimateurs non-linéaires. *PhD Thesis*, 2012.
- [CHM06] D. Cheng, X. Hu, and C. F. Martin. On the smallest enclosing balls. *Communications in Information and Systems*, 6:137–160, 2006.

- [CRC04] A. Cuevas and A. Rodríguez-Casal. On boundary estimation. *Adv. in Appl. Probab.*, 36:340–354, 2004.
- [CS10] M. Cule and R. Samworth. Theoretical properties of the log-concave maximum likelihood estimation of a multidimensional density. *Electron. J. Stat.*, 4:254–270, 2010.
- [CSS10] M. Cule, R. Samworth, and M. Stewart. Maximum likelihood estimation of a multidimensional log-concave density. *J. Roy. Statist. Soc., Ser. B (with discussion)*, 72:545–607, 2010.
- [Cue09] A. Cuevas. Set estimation : Another bridge between statistics and geometry. *Sociedad de Estadística e Investigación Operativa*, 25:71–85, 2009.
- [CW13] H. P. Chan and G. Walther. Detection with the scan and the average likelihood ratio. *Statistica Sinica*, 23:409–428, 2013.
- [DMM07] L. D. Drager, Lee J. M., and C. F. Martin. On the geometry of the smallest circle enclosing a finite set of points. *Journal of the Franklin Institute*, 344:929–940, 2007.
- [Dud74] R. M. Dudley. Metric entropy os some classes of sets with differentiable boundaries. *Journal of Approximation Theory*, 10:227–236, 1974.
- [Dup22] C. Dupin. *Applications de géométrie et de mécanique à la marine*. Ponts et Chaussées, Paris, 1822.
- [DW80] L. Devroye and G. L. Wise. Detection of abnormal behaviour via nonparametric estimation of the support. *SIAM J. Applied Math.*, 38:480–488, 1980.
- [DW13] C. Doss and J. A. Wellner. Global rates of convergence of the mles of log-concave and s-concave densities. *Submitted, arXiv:1306.1438*, 2013.
- [Dwy88] R. A. Dwyer. On the convex hull of random points in a polytope. *J. Appl. Prob.*, 25:688–699, 1988.
- [Efr65] B. Efron. The convex hull of a random set of points. *Biometrika*, 52:331–343, 1965.
- [FMSnt] S. Frick, A. Munk, and H. Sieling. Multiscale change-point inference. *arXiv:1301.7212*, preprint.

- [Gay97] G. Gayraud. Estimation of functionals of density support. *Mathematical Methods of Statistics*, 6:26–47, 1997.
- [Gay01] G. Gayraud. Minimax hypothesis testing about the density support. *Bernoulli*, 7:507–525, 2001.
- [Gef64] J. Geffroy. Sur un problème d’estimation géométrique. *Publications de l’Institut de Statistique des Universités de Paris*, 13:191–210, 1964.
- [GG97] S. Glasauer and P. M. Gruber. Asymptotic estimates for best and stepwise approximation of convex bodies iii. *Forum Math.*, 9:383–404, 1997.
- [GKKW02] L. Györfy, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer, 2002.
- [GMR95] Y. Gordon, M. Meyer, and S. Reisner. Constructing a polytope to approximate a convex body. *Geometricae Dedicata*, 57:217–222, 1995.
- [Gro74] H. Groemer. On the mean value of the volume of a random polytope in a convex set. *Arch. Math.*, 25:86–90, 1974.
- [Gru67] B. Grünbaum. *Convex polytopes*. John Wiley and Sons, London-New York-Sidney, 1967.
- [Gru91] P. M. Gruber. Volume approximation of convex bodies by circumscribed polytopes. *Amer. Math. Soc.*, 4:309–317, 1991.
- [Gru93] P. M. Gruber. Asymptotic estimates for best and stepwise approximation of convex bodies ii. *Forum Math.*, 5:521–538, 1993.
- [Gun11] A. Guntuboyina. Lower bounds for the minimax risk using f-divergences, and applications. *IEEE Transactions on Information Theory*, 57:2386–2399, 2011.
- [Gun12] A. Guntuboyina. Optimal rates of convergence for convex set estimation from support functions. *Annals of Statistics*, 40:385–411, 2012.
- [HRW08] S. Hart, Y. Rinott, and B. Weiss. Evolutionarily stable strategies of random games, and the vertices of random polygons. *The Annals of Applied Probability*, 18:259–297, 2008.

- [HS07] D. Hug and R. Schneider. Stability result for a volume ratio. *Israel J. Math.*, 161:209–219, 2007.
- [IK84] I. A. Ibragimov and R. Z. Khasminskii. *Statistical Estimation: Asymptotic Theory*. Springer, 1984.
- [Jan87] S. Janson. Maximal spacings in several dimensions. *The Annals of Probability.*, 15:274–280, 1987.
- [Kor06] A. P. Korostelev. On minimax estimation of a discontinuous signal (translation of a paper in russian, published in 1986). *Theory Probab. Appl.*, 32:727–730, 2006.
- [KST95] A. P. Korostelev, L. Simar, and A. B. Tsybakov. On estimation of monotone and convex boundaries. *Publications de l’Institut de Statistique de l’Université de Paris*, 39:3–18, 1995.
- [KT59] A. N. Kolmogorov and V. M. Tikhomirov.  $\varepsilon$ -entropy and  $\varepsilon$ -capacity of sets in function spaces (in russian). *Uspekhi Mat. Nauk*, 14:3–86, 1959.
- [KT92] A. P. Korostelev and A. B. Tsybakov. Asymptotically minimax image reconstruction problems. in topics in nonparametric estimation. (*R. Z. Khasminskii, ed.*) *Amer. Math. Soc., Providence, RI*, pages 45–86, 1992.
- [KT93a] A. P. Korostelev and A. B. Tsybakov. Estimation of the density support and its functionals (in russian). *Probl. Peredachi Inf.*, 29:1–15, 1993.
- [KT93b] A. P. Korostelev and A. B. Tsybakov. *Minimax Theory of Image Reconstruction*. Springer, NY, 1993.
- [KT94] A. P. Korostelev and A. B. Tsybakov. Asymptotic efficiency in estimation of a convex set (in russian). *Problems of Information Transmission*, 30:317–327, 1994.
- [Leb03] E. Lebarbier. Detecting multiple change-points in the mean of gaussian process by model selection. *Rapport de recherche de l’INRIA, n. 4740*, pages 1–24, 2003.
- [Lei59] K. Leichtweiß. Über die affine Exzentrizität konvexer Körper. *Arch. Math.*, 10:187–199, 1959.
- [Lei98] K. Leichtweiß. *Affine geometry of convex bodies*. Johann Ambrosius Barth Verlag, 1998.



- [Lep91] O. V. Lepski. Asymptotically minimax adaptive estimation i. upper bounds. optimally adaptive estimates. *Theory Probab. Appl.*, 36:682–697, 1991.
- [LT00] O. V. Lepski and A. B. Tsybakov. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, 117:17–48, 2000.
- [Man03] C. F. Manski. *Partial identification of probability distributions*. Springer Verlag, New York, 2003.
- [McM70] P. McMullen. The maximum numbers of faces of a convex polytope. *Mathematika*, 17:179–184, 1970.
- [MCRF10] S. N. Majumdar, A. Comtet, and J. Randon-Furling. Random convex hulls and extreme value statistics. *Journal of Statistical Physics*, 138:955–1009, 2010.
- [Mol05] I. Molchanov. *Theory of Random Sets*. Springer Verlag, 2005.
- [MT95] E. Mammen and A. Tsybakov. Asymptotical Minimax Recovery of Sets with Smooth Boundaries. *Annals of Statistics*, 23:502–524, 1995.
- [MT99] E. Mammen and A. Tsybakov. Smooth discrimination analysis. *Annals of Statistics*, 27:1808–1829, 1999.
- [Par12] J. Pardon. Central limit theorems for uniform random polygons. *Journal of Theoretical Probability*, 25:823–833, 2012.
- [PL08] B. Pateiro-Lopez. Set estimation under convexity type restrictions. *PhD Thesis*, 2008.
- [Pol95] W. Polonik. Measuring mass concentrations and estimating density contour clusters - an excess mass approach. *The Annals of Statistics*, 23:855–881, 1995.
- [Rei03] M. Reitzner. Random polytopes and the Efron-Stein jackknife inequality. *The Annals of Probability*, 31:2136–2166, 2003.
- [Rei04] M. Reitzner. Stochastic approximation of smooth convex bodies. *Mathematika*, 51:11–29, 2004.
- [RS63] A. Rényi and R. Sulanke. Über die konvexe Hülle von  $n$  zufällig gewählten Punkten. *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 2:75–84, 1963.

- [RS64] A. Rényi and R. Sulanke. Über die konvexe Hülle von  $n$  zufällig gewählten Punkten. ii. *Z. Wahrscheinlichkeitsth. Verw. Geb.*, 3:138–147, 1964.
- [RSW01] S. Reisner, C. Schütt, and E. Werner. Dropping a vertex or a facet from a convex polytope. *Forum Math.*, 13:359–378, 2001.
- [Sch93a] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*. Cambridge University Press, 1993.
- [Sch93b] C. Schütt. On the affine surface area. *Proceedings of the American Mathematical Society*, 118:1213–1218, 1993.
- [Sch94] C. Schütt. Random polytopes and affine surface area. *Mathematische Nachrichten*, 170:227–249, 1994.
- [SW90] C. Schütt and E. Werner. The convex floating body. *Math. Scand*, 66:275–290, 1990.
- [Syl57] J. J. Sylvester. A question in the geometry of situation. *Quarterly Journal of Mathematics*, pages 1–79, 1857.
- [SZ10] X. Shao and X. Zhang. Testing for change points in time series. *Journal of the American Statistical Association*, 105:1228–1240, 2010.
- [Tsy94] A. B. Tsybakov. Multidimensional change-point problems and boundary estimation. *IMS Lecture Notes - Monograph Series*, 23, 1994.
- [Tsy97] A. Tsybakov. On nonparametric estimation of density level sets. *Ann. Stat.*, 25:948–969, 1997.
- [Tsy09] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer, 2009.
- [Vu05] V. H. Vu. Sharp concentration of random polytopes. *Geom. Funct. Anal.*, 15:1284–1318, 2005.
- [Wel91] E. Welzl. Smallest enclosing disks (balls and ellipsoids). *New Results and New Trends in Computer Science, H. Maurer, Editor, Lecture Notes in Computer Science 555*, pages 359–370, 1991.
- [Wer06] E. Werner. Floating bodies and Illumination bodies. *Proceedings of the Conference Integral Geomerty and Convexity, Wuhan 2004, World Scientific, Singapore*, 2006.

- [Wie78] J. A. Wieacker. Einige Probleme der polyedrischen Approximation. *Diplomarbeit (Frieburg i. Br.)*, 1978.
- [Zie95] G. M. Ziegler. *Lectures on polytopes*. Springer, 1995.