Adaptive estimation of convex and polytopal support

Victor-Emmanuel Brunel

Adaptive estimation of convex and polytopal density support

Victor-Emmanuel Brunel
CREST-ENSAE, Malakoff, France / University of Haifa, Israel
victor.emmanuel.brunel@ensae-paristech.fr

Abstract

We estimate the support of a uniform density, when it is assumed to be a convex polytope or, more generally, a convex body in $\mathbb{R}^d$. In the polytopal case, we construct an estimator achieving a rate which does not depend on the dimension d, unlike the other estimators that have been proposed so far. For $d \geq 3$, our estimator has a better risk than the previous ones, and it is nearly minimax, up to a logarithmic factor. We also propose an estimator which is adaptive with respect to the structure of the boundary of the unknown support.

Keywords: adaptation; convex set; density support; minimax; polytope.

1 Introduction

Assume we observe a sample of n i.i.d. random variables $X_i, i = 1, \ldots, n$, with uniform distribution on some subset G of $\mathbb{R}^d, d \geq 2$. We are interested in the problem of estimation of G. In particular, this problem is of interest in detection of abnormal behavior, cf. Devroye and Wise [11]. 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. When G is known to be compact and convex, the convex hull of the sample is quite a natural estimator. The properties of this random subset of $\mathbb{R}^d$ have been extensively studied since the early 1960's, from a geometric and probabilistic prospective. 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 [1] for a historical survey and extensions of Sylvester problem. Of course, this question is not well posed, since the answer should depend on the probability measure of those four points, and the many answers that were proposed, in the late 18th century, accompanied the birth of a new field: stochastic geometry. Rényi and Sulanke [22, 23] studied some basic properties of the convex hull of $X_i, i = 1, \dots, n$ when G is a compact and convex subset of the plane ($d = 2$). More specifically, if this convex hull is denoted by $CH(X_1, \dots, X_n)$, its number of vertices by V_n and its missing area $|G \backslash CH(X_1, \dots, X_n)|$ 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 G. The expected number of vertices is of the order $n^{1/3}$ when the boundary of G is smooth enough, and $r \ln n$ when G 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 G 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 [22, 23] provided asymptotic evaluations of these expectations with the explicit constants, up to two or three terms. In 1965, Efron [14] 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, if $|G|$ stands for the area of G, one has

$$\mathbb{E}[A_n] = \frac{|G|\mathbb{E}[V_{n+1}]}{n+1}, \tag{1}$$

independently of the structure of the boundary of G. In particular, (1) allows to extend the results of [22, 23] about the missing area to any convex polygon with r vertices. If G is such a polygon, $\mathbb{E}[A_n]$ is of the order $r(\ln n)/n$, up to a factor of the form $c|G|$, where c is positive and does not depend on r or G. More recently, many efforts were made to extend these results to dimensions 3 and more. We refer to [28], [15], [13] and the references therein. Notably, Efron's identity (1) holds in any dimension if $G \subseteq \mathbb{R}^d$ is a compact and convex set and $|G|$ is its Lebesgue measure.

Bàràny and Larman [4] (see [3] for a review) proposed a generalization of these results with no assumption on the structure of the boundary of G. They considered the ε-wet part of G, denoted by $G(\varepsilon)$ and defined as the union of all the caps of G of volume $\varepsilon|G|$, where a cap is the intersection of G with a half space. Here, $0 \leq \varepsilon \leq 1$. This notion, together with that of floating body (defined as $G \backslash G(\varepsilon)$) had been introduced by Dupin [12] and, later, by Blaschke [5]. In [4], the authors prove that the expected missing volume of the convex hull of independent random points uniformly distributed in a convex body is of the order of the volume of the $1/n$-wet part. Then the problem of computing this expected missing volume becomes analytical, and learning about its asymptotics reduces to analyzing the properties of the wet part, which have been studied extensively in convex analysis and geometry. In particular we refer to [24], [25], [27], [20] and the references therein. In particular, it was shown that if the boundary of the convex body G is smooth enough, then the expected missing volume is of the order $n^{-2/(d+1)}$, and if G is a polytope, the order is $(\ln n)^{d-1}n^{-1}$.

All these works were developed in a geometric and probabilistic prospective. No efforts were made at this stage to understand whether the convex hull estimator is optimal if seen as an estimator of the set G. Only in the 1990's, this question was invoked in the statistical literature. Mammen and Tsybakov [19] showed that under some restrictions on the volume of G, the convex hull is optimal in a minimax sense (see the next section for details). Korostelev and Tsybakov [17] give a detailed account of the topic of set estimation. See also [8], [9], [10], [16], for an overview of recent developments about estimation of the support of a probability measure. A different model was studied in [7], where we considered estimation of the support of the regression function. We built an estimator which achieves a speed of convergence of the order $(\ln n)/n$ when the support is a polytope in $\mathbb{R}^d, d \geq 2$. Moreover, we proved that no estimator can achieve a better speed of convergence, so the logarithmic factor cannot be dropped. Although our estimator depends on the knowledge of the number of vertices r of the true support of the regression function, we proposed an adaptive estimator, with respect to r, which achieves the same speed as for the case of known r.

However, to our knowledge, when one estimates the support of a uniform distribution, there are no results about optimality of the convex hull estimator when that support is a general convex set. In particular, when no assumptions on the location and on the structure of the boundary are made, it is not known if a convex set can be uniformly consistently estimated. In addition, the case of polytopes has not been investigated. Intuitively, the convex hull estimator can be improved, and the logarithmic factor can be, at least partially, dropped. Indeed, a polytope with a given number of vertices is completely determined by the coordinates of its vertices, and therefore belongs to some parametric family. This paper is organized as follows. In Section 2 we give all notation and definitions. In Section 3 we propose a new estimator of G when it is assumed to be a polytope and its number of vertices is known. We show that the risk of this estimator is better than that of the convex hull estimator, and achieves a rate independent of the dimension d. In Section 4, we show that in the general case, if no other assumption than compactness, convexity and positive volume is made on G, then the convex hull estimator is optimal in a minimax sense. In

Section 5 we construct an estimator which is adaptive to the shape of the boundary of G, i.e. which detects, in some sense, whether G is a polytope or not and, if yes, correctly estimates its number of vertices. Section 6 is devoted to the proofs.

2 Notation and Definitions

Let $d \geq 2$ a positive integer. Denote by ρ the Euclidean distance in $\mathbb{R}^d$ and by B_2^d the Euclidean unit closed ball in $\mathbb{R}^d$.

For brevity, we will a call convex body any compact and convex subset of $\mathbb{R}^d$ with positive volume, and we will call a polytope any compact convex polytope in $\mathbb{R}^d$ with positive volume. For an integer $r \geq d+1$, denote by $\mathcal{P}_r$ the class of all convex polytopes in $[0,1]^d$ with at most r vertices. Denote also by $\mathcal{K}$ the class of all convex bodies in $\mathbb{R}^d$.

If G is a closed subset of $\mathbb{R}^d$ and ϵ is a positive number, we denote by G^ϵ the set of all $x \in \mathbb{R}^d$ such that $\rho(x, G) \leq \epsilon$ or, in other terms, $G^\epsilon = G + \epsilon B_2^d$. If G is any set, $I(\cdot \in G)$ stands for the indicator function of G.

The Lebesgue measure on $\mathbb{R}^d$ is denoted by $|\cdot|$ (for brevity, we do not indicate explicitly the dependence on d). If G is a measurable subset of $\mathbb{R}^d$, we denote respectively by $\mathbb{P}_G$ and $\mathbb{E}_G$ the probability measure of the uniform distribution on G and the corresponding expectation operator, and we still use the same notation for the n-product of this distribution if there is no possible confusion. When necessary, we add the superscript $^{\otimes n}$ for the n product. We will use the same notation for the corresponding outer probability and expectation when it is necessary, to avoid measurability issues. 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 $|G_1 \triangle G_2|$.

A subset $\hat{G}_n$ of $\mathbb{R}^d$, whose construction depends on the sample is called a set estimator or, more simply, an estimator.

Given an estimator $\hat{G}_n$, we measure its accuracy on a given class of sets in a minimax framework. The risk of $\hat{G}_n$ on a class $\mathcal{C}$ of Borel subsets of $\mathbb{R}^d$ is defined as

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

The rate (a sequence depending on n) of an estimator on a class $\mathcal{C}$ is the speed at which its risk converges to zero when the number n of available observations tends to infinity. For all the estimators defined in the sequel we will be interested in upper bounds on their risk, in order to get information about their rate. For a given class of subsets $\mathcal{C}$, the minimax risk on this class when n observations are available is defined as

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

where the infimum is taken over all set estimators depending on n observations. If $\mathcal{R}_n(\mathcal{C})$ converges to zero, we call the minimax rate of convergence on the class $\mathcal{C}$ the speed at which $\mathcal{R}_n(\mathcal{C})$ tends to zero. For a given class $\mathcal{C}$ of subsets of $\mathbb{R}^d$, 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.

For two quantities A and B, and a parameter ϑ, which may be multidimensional, we will write $A \lesssim_\vartheta B$ (respectively $A \gtrsim_\vartheta B$) to say that for some constant positive constant $c(\vartheta)$ which depends on ϑ only, one has $A \leq c(\vartheta)B$ (respectively $A \geq c(\vartheta)B$). If we put no subscript under the signs $\lesssim$ or $\gtrsim$, this means that the involved constant is universal, i.e. depends on no parameter.

3 Estimation of polytopes

3.1 Upper bound

Let $r \geq d+1$ be a known integer. Assume that the underlying set G, denoted by P in this section, is in $\mathcal{P}_r$. The likelihood of the model, seen as a function of the compact set $G' \subseteq \mathbb{R}^d$, is defined as follows, provided that G' has a positive Lebesgue measure:

$$L(X_1, \ldots, X_n, G') = \prod_{i=1}^{n} \frac{I(X_i \in G')}{|G'|}.$$

Therefore, maximization of the likelihood over a given class $\mathcal{C}$ of candidates, $\max_{G' \in \mathcal{C}} L(X_1, \ldots, X_n, G')$, is achieved when G' is of minimum Lebesgue measure among all sets of $\mathcal{C}$ containing all the sample points. When $\mathcal{C}$ is the class of all convex subsets of $\mathbb{R}^d$, the maximum likelihood estimator is unique, and it is the convex hull of the sample. As we discussed above, this estimator has been extensively studied. In particular, using Efron's identity (1), it turns out that its expected number of vertices is of the order of $r(\ln n)^{d-1}$. However, the unknown polytope P has no more than r vertices. Hence, it seems reasonable to restrict the estimator of P to have much less vertices; the class of all convex subsets of $\mathbb{R}^d$ is too large and we propose to maximize the likelihood over the smaller class $\mathcal{P}_r$.

Assume that there exists a polytope in $\mathcal{P}_n$ with the smallest volume among all polytopes of $\mathcal{P}_r$ containing all the sample points. Let $\hat{P}_n^{(r)}$ be such a polytope, i.e.

$$\hat{P}_n^{(r)} \in \underset{P\in\mathcal{P}_r, X_i\in P, i=1,\ldots,n}{\operatorname{argmin}} |P|. \tag{2}$$

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 1. *Let $r \geq d+1$ be an integer, and $n \geq 2$. Then,*

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

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

Corollary 1. *Let the assumptions of Theorem 1 be satisfied. Then, for any positive number q,*

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

Corollary 1 shows that the risk $\mathcal{R}_n(\hat{P}_n^{(r)}; \mathcal{P}_r)$ of the estimator $\hat{P}_n^{(r)}$ on the class $\mathcal{P}_r$ 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 $\mathcal{P}_r$:

$$\mathcal{R}_n(\mathcal{P}_r) \lesssim_d \frac{r\ln n}{n}. \tag{3}$$

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

3.2 The logarithmic factor

We conjecture that the logarithmic factor can be removed in the upper bound of $\mathcal{R}_n(\mathcal{P}_r), r \geq d+1$. Specifically, for the class of all convex polytopes with at most r vertices, not necessarily included in the square $[0,1]^d$, which we denote by $\mathcal{P}_r^{all}$, we conjecture that, for the normalized version of the risk,

$$\mathcal{Q}_n(\mathcal{P}_r^{all}) \lesssim_d \frac{r}{n}.$$

What motivates our intuition is Efron's identity (1). Let us recall its proof, which is very easy, and instructive for our purposes. Let the underlying set G be a convex body in $\mathbb{R}^d$, denoted by K, and let $\hat{K}_n$ be the convex hull of the sample. Almost surely, $\hat{K}_n \subseteq K$, so

$$\begin{aligned}
\mathbb{E}_K^{\otimes n}[|\hat{K}_n \triangle K|] &= \mathbb{E}_K^{\otimes n}[|K \backslash \hat{K}_n|] \\
&= \mathbb{E}_K^{\otimes n}\left[\int_K I(x \notin \hat{K}_n)dx\right] \\
&= |K|\mathbb{E}_K^{\otimes n}\left[\frac{1}{|K|}\int_K I(x \notin \hat{K}_n)dx\right] \\
&= |K|\mathbb{E}_K^{\otimes n}\left[\mathbb{P}_K[X \notin \hat{K}_n | X_1, \ldots, X_n]\right], \qquad (4)
\end{aligned}$$

where X is a random variable with the same distribution as X_1, and independent of the sample $X_1, \ldots, X_n$, and $\mathbb{P}_K[\cdot | X_1, \ldots, X_n]$ denotes the conditional distribution given $X_1, \ldots, X_n$. In what follows, we set $X_{n+1} = X$, so that we can consider the bigger sample $X_1, \ldots, X_{n+1}$. For $i = 1, \ldots, n+1$, we denote by $\hat{K}^{-i}$ the convex hull of the sample $X_1, \ldots, X_{n+1}$ from which the i-th variable X_i is withdrawn. Then $\hat{K}_n = \hat{K}^{-(n+1)}$, and by continuing (4), and by using the symmetry of the sample,

$$\begin{aligned}
\mathbb{E}_K^{\otimes n}[|\hat{K}_n \triangle K|] &= |K|\mathbb{P}_K^{\otimes n+1}[X_{n+1} \notin \hat{K}^{-(n+1)}] \\
&= \frac{|K|}{n+1}\sum_{i=1}^{n+1}\mathbb{P}_K^{\otimes n+1}[X_i \notin \hat{K}^{-(i)}] \\
&= \frac{|K|}{n+1}\sum_{i=1}^{n+1}\mathbb{P}_K^{\otimes n+1}[X_i \in V(\hat{K}_{n+1})], \qquad (5)
\end{aligned}$$

where $V(\hat{K}_{n+1})$ is the set of vertices of $\hat{K}_{n+1} = CH(X_1, \ldots, X_{n+1})$. 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 (5),

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

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.

It turns out that we can follow almost all the proof of this identity when the underlying set G is a polytope, and when we consider the estimator developed in Section 3.1. Let $r \geq d+1$ be an integer and $P \in \mathcal{P}_r^{all}$. Let $\hat{P}_n^{(r)}$ be the estimator defined in (2), where $\mathcal{P}_r$ is replaced by $\mathcal{P}_r^{all}$. In this section, we denote this estimator simply by $\hat{P}_n$. Note that this estimator does not satisfy the nice property $\hat{P}_n \subseteq P$, unlike the convex hull. However, by construction, $|\hat{P}_n| \leq |P|$, so $|P \triangle \hat{P}_n| \leq 2|P \backslash \hat{P}_n|$, and we have:

$$\begin{aligned}\mathbb{E}_P^{\otimes n}[|\hat{P}_n \triangle P|] &\leq 2\mathbb{E}_P^{\otimes n}[|P\backslash \hat{P}_n|] \\ &= 2|P|\mathbb{E}_P^{\otimes n}\left[\frac{1}{|P|}\int_P I(x \notin \hat{P}_n)dx\right] \\ &= 2|P|\mathbb{E}_P^{\otimes n}\left[\mathbb{P}_P[X \notin \hat{P}_n | X_1, \ldots, X_n]\right], \end{aligned} \tag{6}$$

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

(6),

$$\begin{aligned}
\mathbb{E}_P^{\otimes n}[|\hat{P}_n \triangle P|] &\leq 2|P|\mathbb{P}_P^{\otimes n+1}[X_{n+1} \notin \hat{P}^{-(n+1)}] \\
&= \frac{2|P|}{n+1} \sum_{i=1}^{n+1} \mathbb{P}_P^{\otimes n+1}[X_i \notin \hat{P}^{-(i)}] \\
&= \frac{2|P|}{n+1} \mathbb{E}_P^{\otimes n+1} \left[\sum_{i=1}^{n+1} I(X_i \notin \hat{P}^{-(i)}) \right] \\
&= \frac{2|P|\mathbb{E}_P^{\otimes n+1}[V'_{n+1}]}{n+1}, \qquad (7)
\end{aligned}$$

where V'_{n+1} stands for the number of points X_i falling outside of the polytope with at most r vertices, of minimum volume, containing all the other n points. Note that in this description we assume the uniqueness of such a polytope, which we conjecture to hold almost surely, as long as n is large enough. It is not clear that if a point X_i is not in $\hat{P}^{-i}$, then X_i lies on the boundary of $\hat{P}_{n+1}$. However, if this was true, then almost surely V'_{n+1} would be less or equal to $d+1$ times the number of facets of $\hat{P}_{n+1}$, since any 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 [21], [6], and we could have our conjecture proved. However, there might be some cases when some points X_i are not in $\hat{P}^{-i}$, though they do not lay on the boundary of $\hat{P}_{n+1}$. So it may be of interest to work directly on the variable V'_{n+1}. This remains an open problem.

3.3 Lower bound for the minimax risk in the case $d = 2$

In the 2-dimensional case, we provide a lower bound of the order $1/n$, with a factor that is linear in the number of vertices r. Namely, the following theorem holds.

Theorem 2. *Let $r \geq 10$ be an integer, and $n \geq r$. Assume $d = 2$. Then,*

$$\mathcal{R}_n(\mathcal{P}_r) \gtrsim \frac{r}{n}.$$

Combined with (3), this bound shows that, as a function of r, $\mathcal{R}_n(\mathcal{P}_r)$ behaves linearly in r in dimension two. In greater dimensions, it is quite easy to show that $\mathcal{R}_n(\mathcal{P}_r) \gtrsim_d \frac{1}{n}$, but this lower bound does not show the dependency in r. However, the upper bound (3) shows that $\mathcal{R}_n(\mathcal{P}_r)$ is at most linear in r.

4 Estimation of convex bodies

In this section we no longer assume that the unknown support G belongs to a class $\mathcal{P}_r, r \geq d+1$, but only that it is a convex body and we write $G = K$. Denote by $\hat{K}_n$ the convex hull of the sample. The risk of this estimator cannot be bounded from above uniformly on the class $\mathcal{K}$, since by (1) for any given n, $\mathbb{E}_K[|K \triangle \hat{K}_n|] \to \infty$ as $|K| \to \infty$. Moreover there is no uniformly consistent estimator on the class $\mathcal{K}$ of all convex bodies if the risk is defined by $(*)$. The following result holds.

Theorem 3. *For all $n \geq 1$, the minimax risk $(**)$ on the class $\mathcal{K}$ is infinite:*

$$\mathcal{R}_n(\mathcal{K}) = +\infty.$$

Therefore we will use another risk measure which is the normalized risk for an estimator $\tilde{K}_n$ of K, based of a sample of n observations:

$$\mathcal{Q}_n(\tilde{K}_n; \mathcal{K}) = \sup_{K \in \mathcal{K}} \mathbb{E}_K \left[\frac{|K \triangle \tilde{K}_n|}{|K|} \right].$$

Also define the normalized minimax risk on the class $\mathcal{K}$:

$$\mathcal{Q}_n(\mathcal{K}) = \inf_{\tilde{K}_n} \sup_{K \in \mathcal{K}} \mathbb{E}_K \left[\frac{|K \triangle \tilde{K}_n|}{|K|} \right],$$

where the infimum is taken over all estimators $\tilde{K}_n$ based on a sample of n i.i.d. observations. For the estimator $\hat{K}_n$ we do not provide a deviation inequality as in Theorem 1, but only an upper bound on the normalized risk.

Theorem 4. *Let $n \geq 2$ be an integer. Then,*

$$\mathcal{Q}_n(\hat{K}_n; \mathcal{K}) \lesssim_d n^{-\frac{2}{d+1}}.$$

Note that this result gives a bound on $\mathbb{E}_K \left[\frac{|K \triangle \hat{K}_n|}{|K|} \right]$ that is uniform over *all* convex bodies in $\mathbb{R}^d$, with no restriction on the location of the set K (such as $K \subseteq [0,1]^d$) or on the volume of K, unlike in [19]. From Theorem 4 and the lower bound of [19] (the lower bound of [19] is for the minimax risk, but the proof still holds for the normalized risk), we obtain the next corollary.

Corollary 2. *Let $n \geq 2$ be an integer. The normalized minimax risk on the class $\mathcal{K}$ satisfies*

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

and the convex hull has the minimax rate of convergence on $\mathcal{K}$, with respect to the normalized version of the risk.

Note that if instead of the class $\mathcal{K}$ we consider the class $\mathcal{K}_1$ of all convex bodies that are included in $[0,1]^d$, then $\forall K \in \mathcal{K}_1, |K| \leq 1$ and therefore, the risk of the convex hull estimator $\hat{K}_n$ on this class is bounded from above by $n^{-2/(d+1)}$, so

$$\mathcal{R}_n(\mathcal{K}_1) \lesssim_d n^{-\frac{2}{d+1}}.$$

Besides, the lower bound that is given in [19] still holds for the class $\mathcal{K}_1$, and thus we have the following corollary.

Corollary 3. *Let $n \geq 2$ be an integer. The minimax risk on the class $\mathcal{K}_1$ satisfies*

$$n^{-\frac{2}{d+1}} \lesssim_d \mathcal{R}_n(\mathcal{K}_1) \lesssim_d n^{-\frac{2}{d+1}},$$

and the convex hull has the minimax rate of convergence on $\mathcal{K}_1$, with respect to the risk defined in $(**)$.

5 Adaptative estimation

In Sections 3 and 4, 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 by construction also 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 located in $[0,1]^d$. This estimator will achieve the same rate as the estimators of Section 3.1 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, up to a logarithmic factor, as the convex hull which was studied in Section 4 in the case where the support is not polytopal, or is polytopal but with too many vertices. Note that if the support is a polytope with r vertices, where r is larger than $(\ln n)^{-1} n^{\frac{d-1}{d+1}}$, then the risk of the convex hull estimator $\hat{K}_n$ has a smaller rate than that of $\hat{P}_n^{(r)}$. The idea which we develop here is the same as in

[7], Theorem 6. The classes $\mathcal{P}_r, r \geq d+1$, are nested, that is, $\mathcal{P}_r \subseteq \mathcal{P}_{r'}$ as soon as $r \leq r'$. So it is better, in some sense, to overestimate the true number vertices of the unknown polytope P. Intuitively, it makes sense to fit some polytope with more vertices to P, while the opposite may be impossible (e.g. it is possible to fit a quadrilateral on any triangle, but not to fit a triangle on a square). We use this idea in order to select an estimator among the preliminary estimators $\hat{P}_n^{(r)}, r \geq d+1$, and $\hat{K}_n$. Note that in [7], Theorem 6, the key tools for adaptation are the deviation inequalities for the preliminary estimators, but we do not have such an inequality for the "last" one, i.e. the convex hull $\hat{K}_n$. This induces a loss of precision in our estimation procedure. Namely, an extra logarithmic factor will appear.

Set $R_n = \lfloor n^{(d-1)/(d+1)} \rfloor$, where $\lfloor \cdot \rfloor$ stands for the integer part. Let C be any positive constant greater than $16d + \frac{16}{d+1}$, and define

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

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 $r = n$. As previously, denote by V_n the number of vertices of $\hat{K}_n$, the convex hull of the sample. By definition of the convex hull, $\hat{P}_n^{(r)} = \hat{K}_n$, for all $r \geq V_n$. Therefore $\hat{r} \leq V_n$ almost surely.

The adaptive estimator is defined as follows.

$$\hat{P}_n^{adapt} = \begin{cases} \hat{P}_n^{(\hat{r})} & \text{if } \hat{r} \leq R_n \\ \hat{K}_n, & \text{otherwise.} \end{cases}$$

Then, if we denote by $\mathcal{P}_\infty = \mathcal{K}_1$, we have the following theorem.

Theorem 5. *Let $n \geq 2$. Let $\phi_{n,r} = \min\left(\frac{r \ln n}{n}, (\ln n) n^{-\frac{2}{d+1}}\right)$, for all integers $r \geq d+1$ and $r = \infty$. Then,*

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

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 $\mathcal{P}_r, d+1 \leq r$ and on the class $\mathcal{K}_1$ of all convex bodies in $[0,1]^d$.

Remark 1. *The construction of $\hat{r}$ is inspired by Lepski's method [18]. However, we cannot use the same techniques here since they need deviation inequalities for all the preliminary estimators, while we do not have such an inequality for $\hat{K}_n$. Our proof uses directly the properties of the convex hull estimator.*

6 Proofs

6.1 Proof of Theorem 1

Let $r \geq d+1$ be an integer, and $n \geq 2$. Let $P_0 \in \mathcal{P}_r$ and consider a sample $X_1, \ldots, X_n$ 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 1. Let us define $\mathcal{P}_r^{(n)}$ as the class of all convex polytopes of $\mathcal{P}_r$ whose vertices lay on the grid $\left(\frac{1}{n}\mathbb{Z}\right)^d$, i.e. have as coordinates integer multiples of $1/n$. We use the following lemma, whose proof can be found in [7].

Lemma 1. *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 $\mathcal{P}_r$ there is a convex polytope $P^* \in \mathcal{P}_r^{(n)}$ such that :*

$$\begin{cases} |P^* \triangle P| \leq \frac{K_1}{n} \\ P^* \subseteq P^{\sqrt{d}/n}, \quad P \subseteq (P^*)^{\sqrt{d}/n}. \end{cases} \tag{8}$$

In particular, taking $P = P_0$ or $P = \hat{P}_n$ in Lemma 1, we can find two polytopes P^* and $\tilde{P}_n$ in $\mathcal{P}_r^{(n)}$ 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, $|\hat{P}| \leq |P_0|$, so $|\hat{P}_n \triangle P_0| \leq 2|P_0 \backslash \hat{P}_n|$. Besides, if G_1, G_2 and G_3 are three measurable subsets of $\mathbb{R}^d$, the following triangle inequality holds :

$$|G_1 \backslash G_3| \leq |G_1 \backslash G_2| + |G_2 \backslash G_3|. \tag{9}$$

Let us now write the following inclusions between the events.

$$\begin{aligned} \left\{|\hat{P}_n \triangle P_0| > \epsilon\right\} &\subseteq \left\{|P_0 \backslash \hat{P}_n| > \epsilon/2\right\} \\ &\subseteq \left\{|P^* \backslash \tilde{P}_n| > \epsilon/2 - \frac{2K_1}{n}\right\} \\ &\subseteq \bigcup_P \left\{\tilde{P}_n = P\right\}, \end{aligned} \tag{10}$$

where the latest union is over the class of all $P \in \mathcal{P}_r^{(n)}$ that satisfy the inequality $|P^* \backslash P| > \epsilon/2 - \frac{2K_1}{n}$. Let P be such a polytope, then if $\tilde{P}_n = P$, then necessarily the sample $\{X_1, \ldots, X_n\}$ is included in $P^{\frac{\sqrt{d}}{n}}$, by definition of $\tilde{P}_n$, and (10) becomes

$$\begin{aligned}
\mathbb{P}_{P_0}\left[\tilde{P}_n = P\right] &\leq \mathbb{P}_{P_0}\left[X_i \in P^{\frac{\sqrt{d}}{n}}, i = 1, \ldots, n\right] \\
&\leq \left(1 - \frac{|P_0 \backslash P^{\frac{\sqrt{d}}{n}}|}{|P_0|}\right)^n \\
&\leq \left(1 - |P_0 \backslash P^{\frac{\sqrt{d}}{n}}|\right)^n, \quad \text{since } |P_0| \leq 1 \\
&\leq \left(1 - |P^* \backslash P| + |P^* \backslash P_0| + |P^{\frac{\sqrt{d}}{n}} \backslash P|\right)^n, \quad \text{using (9)} \\
&\leq \left(1 - \epsilon/2 + \frac{4K_1}{n}\right)^n \\
&\leq C_1 \exp(-n\epsilon/2), \qquad (11)
\end{aligned}$$

where $C_1 = \mathrm{e}^{4K_1}$. Therefore, using (10) and (11) and denoting by $\#\mathcal{P}_r^{(n)}$ the cardinality of the finite class $\mathcal{P}_r^{(n)}$,

$$\begin{aligned}
P_{P_0}\left[|\hat{P}_n \triangle P_0| > \epsilon\right] &\leq \#\mathcal{P}_r^{(n)} C_1 \exp(-n\epsilon/2) \\
&\leq (n+1)^{dr} C_1 \exp(-n\epsilon/2) \\
&\leq C_1 \exp(-n\epsilon/2 + 2dr \ln n). \qquad (12)
\end{aligned}$$

It turns out that if we take ϵ of the form $\dfrac{4dr \ln n}{n} + \dfrac{x}{n}$, (12) becomes

$$\mathbb{P}_{P_0}\left[n\left(|\hat{P}_n \triangle P_0| - \frac{4dr \ln n}{n}\right) \geq x\right] \leq C_1 \mathrm{e}^{-x/2}, \qquad (13)$$

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

Corollary 1 comes by applying Fubini's theorem (see [7] for details).

6.2 Proof of Theorem 2

Let $r \geq 10$ 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 $A_0, A_2, \ldots, A_{r-2}$, such that for all $k = 0, \ldots, r/2 - 1$, the distance between A_{2k} and the center C is $1/2$. Let $A_1, A_3, \ldots, A_{r-1}$ be $r/2$ points built as in Figure 1: for $k = 0, \ldots, r/2 - 1$, A_{2k+1} is on the mediator of the segment $[A_{2k}, A_{2k+2}]$, outside P^*,

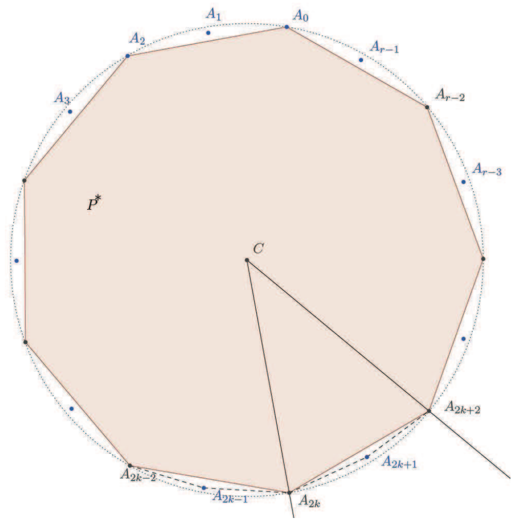

Figure 1: Construction of hypotheses for the lower bound

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 $A_0, A_2, \ldots, A_{r-2}$ 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 1. For $\omega = (\omega_0, \ldots, \omega_{r/2-1}) \in \{0,1\}^{r/2}$, we denote by P_ω the convex hull of P^* and the points $A_{2k+1}, k = 0, \ldots, r/2-1$ such that $\omega_k = 1$. Then we follow the scheme of the proof of Theorem 5 in [7].

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

$$\omega^{(k,0)} = (\omega_1, \ldots, \omega_{k-1}, 0, \omega_{k+1}, \ldots, \omega_{r/2-1}) \text{ and by}$$
$$\omega^{(k,1)} = (\omega_1, \ldots, \omega_{k-1}, 1, \omega_{k+1}, \ldots, \omega_{r/2-1}).$$

Note that for $k = 0, \ldots, r/2-1$, and $(\omega_1, \ldots, \omega_{k-1}, \omega_{k+1}, \ldots, \omega_{r/2-1}) \in \{0,1\}^{r/2-1}$,

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

Let H be the Hellinger distance between probability measures. For the definition and some properties, see [26], Section 2.4. We have, by a simple computation,

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

since $|P_{\omega^{(k,1)}}| \geq |P^*| \geq 1/2$. Now, let $\hat{P}_n$ be any estimator of P^*, based on a sample of n i.i.d. random variables. By the same computation as in the proof of Theorem 5 in [7], based on the $2^{r/2}$ hypotheses that we constructed, we get

$$
\begin{aligned}
\sup_{P\in\mathcal{P}_r}\mathbb{E}_P\left[|P\triangle\hat{P}_n|\right] &\geq \frac{r\delta\cos(2\pi/r)}{8}\left(1-\frac{\delta\cos(2\pi/r)}{4}\right)^n \\
&\geq \frac{rh\cos\left(\frac{2\pi}{r}\right)^2\tan\left(\frac{4\pi}{r}\right)}{8}\left(1-\frac{h\cos\left(\frac{2\pi}{r}\right)^2\tan\left(\frac{4\pi}{r}\right)}{8}\right)^n. \qquad (14)
\end{aligned}
$$

Note that if we denote by $x=\dfrac{2\pi}{r}>0$ and $\phi(x)=\dfrac{1}{x}\cos(x)\tan(2x)$, then $\phi(x)\gtrsim 1$ since r is supposed to be greater or equal to 10. Therefore, by the choice $h=r/n\leq 1$ (we assumed that $n\geq r$), (14) becomes

$$
\sup_{P\in\mathcal{P}_r}\mathbb{E}_P\left[|P\triangle\hat{P}_n|\right]\gtrsim\frac{r}{n},
$$

and Theorem 2 is proved.

6.3 Proof of Theorem 3

Let $t>0$ be fixed. Let $G_1=t^{1/d}B_2^d$ and $G_2=(2t)^{1/d}B_2^d$. Let us denote respectively by $\mathbb{P}_1$ and $\mathbb{P}_2$ the uniform distributions on G_1 and G_2, and by $\mathbb{E}_1$ and $\mathbb{E}_2$ the corresponding expectations. We denote by $\mathbb{P}_1^{\otimes n}$ and $\mathbb{P}_2^{\otimes n}$ the n-product of $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively, i.e. the probability distribution of a sample of n i.i.d. random variables of distribution $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively. The corresponding expectations are still denoted by $\mathbb{E}_1$ and $\mathbb{E}_2$. Then, for any estimator $\hat{G}_n$ based on a sample of n random variables, we bound from bellow the minimax risk by the Bayesian one.

$$\begin{aligned}\sup_{G\in\mathcal{K}} \mathbb{E}_G[|\hat{G}_n \triangle G|] &\geq \frac{1}{2}\left(\mathbb{E}_1[|\hat{G}_n \triangle G_1|] + \mathbb{E}_2[|\hat{G}_n \triangle G_2|]\right) \\ &\geq \frac{1}{2}\int_{(\mathbb{R}^d)^n} \left(|\hat{G}_n \triangle G_1| + |\hat{G}_n \triangle G_2|\right) \min(d\mathbb{P}_1^{\otimes n}, d\mathbb{P}_2^{\otimes n}) \\ &\geq \frac{1}{2}\int_{(\mathbb{R}^d)^n} |G_1 \triangle G_2| \min(d\mathbb{P}_1^{\otimes n}, d\mathbb{P}_2^{\otimes n}) \\ &\geq \frac{t|B_2^d|}{4}\left(1 - \frac{H(\mathbb{P}_1^{\otimes n}, \mathbb{P}_2^{\otimes n})^2}{2}\right)^2,\end{aligned}$$

where H is the Hellinger distance between probability measures, as in the proof of Theorem 2. Therefore,

$$\sup_{G\in\mathcal{K}} \mathbb{E}_G[|\hat{G}_n \triangle G|] \geq \frac{t|B_2^d|}{4}\left(1 - \frac{H(\mathbb{P}_1, \mathbb{P}_2)^2}{2}\right)^{2n}, \tag{15}$$

and a simple computation shows that

$$1 - \frac{H(\mathbb{P}_1, \mathbb{P}_2)^2}{2} = \frac{1}{\sqrt{2}},$$

and (15) becomes

$$\sup_{G\in\mathcal{K}} \mathbb{E}_G[|\hat{G}_n \triangle G|] \geq \frac{t|B_2^d|}{2^{n+2}}.$$

This ends the proof of Theorem 3 by taking t arbitrarily large.

6.4 Proof of Theorem 4

We first state the following result, due to Groemer [15].

Lemma 2. *Let $K \in \mathcal{K}$ and n be an integer greater than the dimension d. Let $\hat{K}_n$ be the convex hull of a sample of n i.i.d. random variables uniformly distributed in K. Then, if B denotes a Euclidean ball in $\mathbb{R}^d$, of the same volume as K,*

$$\mathbb{E}_K[|K \triangle \hat{K}_n|] \leq \mathbb{E}_B[|B \triangle \hat{K}_n|].$$

If $K \in \mathcal{K}$, we denote by $K_1 = \frac{1}{|K|^{1/d}}K$, so K_1 is homothetic to K and has volume 1. Besides, if $X_1, \dots, X_n$ are i.i.d. with uniform distribution in K, then $\frac{1}{|K|^{1/d}}X_1, \dots, \frac{1}{|K|^{1/d}}X_n$ are i.i.d. with uniform distribution in K_1, and

$$\mathbb{E}_K\left[\frac{|K \triangle \hat{K}_n|}{|K|}\right] = \mathbb{E}_{K_1}[|K_1 \triangle \hat{K}_n|]$$

Therefore, one gets, from the previous lemma, that if $B = B_2^d$ is the unit Euclidean ball in $\mathbb{R}^d$,

$$\mathbb{E}_K\left[\frac{|K \triangle \hat{K}_n|}{|K|}\right] \leq \mathbb{E}_B\left[\frac{|B \triangle \hat{K}_n|}{|B|}\right], \forall K \in \mathcal{K}.$$

In other terms,

$$\mathcal{Q}_n(\tilde{K}_n; \mathcal{K}) \leq \mathbb{E}_B\left[\frac{|B \triangle \hat{K}_n|}{|B|}\right].$$

By [2], $n^{\frac{2}{d+1}}\mathbb{E}_B\left[\frac{|B \triangle \hat{K}_n|}{|B|}\right]$ tends to the affine surface area of $B = B_2^d$ times some positive constant which depends on d only, that is,

$$n^{\frac{2}{d+1}}\mathbb{E}_B\left[\frac{|B \triangle \hat{K}_n|}{|B|}\right] \to C(d),$$

where $C(d)$ is a positive constant which depends on d only. Thus, Theorem 4 is proved.

6.5 Proof of Theorem 5

Let $\hat{r}$ be chosen as in Section 5. Let $d+1 \leq r^* \leq R_n$ and $P \in \mathcal{P}_{r^*}$. We distinguish two cases,

$$\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|] = \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} \leq r^*)] + \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} > r^*)], \qquad (16)$$

and we bound separately the two terms in the right side. Note that if $\hat{r} \leq r^*$, then, $\hat{r} \leq R_n$, so $\hat{P}_n^{adapt} = \hat{P}_n^{(\hat{r})}$, and by definition of $\hat{r}$

$$|\hat{P}_n^{(r^*)} \triangle \hat{P}_n^{(\hat{r})}| \leq \frac{Cr^* \ln n}{n}. \qquad (17)$$

Therefore, using the triangle inequality,

$$\begin{aligned}
&\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} \leq r^*)] \\
&\quad \leq \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle \hat{P}_n^{(r^*)}| I(\hat{r} \leq r^*)] + \mathbb{E}_P[|\hat{P}_n^{(r^*)} \triangle P| I(\hat{r} \leq r^*)] \\
&\quad \lesssim_d \frac{r^* \ln n}{n}, \text{ by (17) and Corollary 1.} \qquad (18)
\end{aligned}$$

The second term of (16) is bounded differently. First note that $\hat{P}_n^{adapt} \subseteq [0,1]^d$, so $P \triangle \hat{P}_n^{adapt} \subseteq [0,1]^d$ and $|P \triangle \hat{P}_n^{adapt}| \leq 1$ almost surely. Besides, note that if $\hat{r} > r^*$, then for some $r \in \{r^*+1, \ldots, n\}$, $|\hat{P}_n^{(r^*)} \triangle \hat{P}_n^{(r)}| > \frac{Cr \ln n}{2n}$. Otherwise, for any $r_1, r_2 \in \{r^*, \ldots, n\}$, one would have, by the triangle inequality, $|\hat{P}_n^{(r_1)}\triangle\hat{P}_n^{(r_2)}| \leq \frac{Cr\ln n}{n}$, and this would contradict that $\hat{r} > r^*$. Thus, we have the following inequalities.

$$\begin{aligned}\mathbb{E}_P[|\hat{P}_n^{adapt}\triangle P|I(\hat{r} > r^*)] &\leq \mathbb{P}_P[\hat{r} > r^*] \\ &\leq \sum_{r=r^*+1}^{n} \mathbb{P}_P\left[|\hat{P}_n^{(r^*)}\triangle\hat{P}_n^{(r)}| > \frac{Cr\ln n}{2n}\right]\end{aligned}$$

$$\begin{aligned}&\leq \sum_{r=r^*+1}^{n} \mathbb{P}_P\left[|\hat{P}_n^{(r^*)}\triangle P| + |\hat{P}_n^{(r)}\triangle P| > \frac{Cr\ln n}{2n}\right] \\ &\leq \sum_{r=r^*+1}^{n} \left(\mathbb{P}_P\left[|\hat{P}_n^{(r^*)}\triangle P| > \frac{Cr\ln n}{4n}\right] + \mathbb{P}_P\left[|\hat{P}_n^{(r)}\triangle P| > \frac{Cr\ln n}{4n}\right]\right). \qquad (19)\end{aligned}$$

Note that since $P \in \mathcal{P}_{r^*}$, it is also true that $P \in \mathcal{P}_r, \forall r \geq r^*$. Therefore, for $d+1 \leq r^* \leq r \leq n$, we have, using Theorem 1, with $x = (C/4 - 4d)r\ln n$,

$$\mathbb{P}_P\left[|\hat{P}_n^{(r)}\triangle P| > \frac{Cr\ln n}{4n}\right] \leq e^{-(C/8-2d)r\ln n} \leq n^{-(C/8-2d)(d+1)} \lesssim_d n^{-1},$$

by the choice of C.

It comes from (19) that

$$\mathbb{E}_P[|\hat{P}_n^{adapt}\triangle P|I(\hat{r} > r^*)] \lesssim_d n^{-1}. \qquad (20)$$

Finally, using (18) and (20),

$$\mathbb{E}_P[|\hat{P}_n^{adapt}\triangle P|] \lesssim_d \frac{r^*\ln n}{n}.$$

Let us now assume that the unknown support, which we now denote by K, is any convex body in $\mathcal{P}_\infty = \mathcal{K}_1$, possibly a polytope with many (more than R_n) vertices. We write, similarly to the previous case,

$$\mathbb{E}_K[|\hat{P}_n^{adapt}\triangle K|] = \mathbb{E}_K[|\hat{P}_n^{adapt}\triangle K|I(\hat{r} \leq R_n)] + \mathbb{E}_K[|\hat{P}_n^{adapt}\triangle K|I(\hat{r} > R_n)], \qquad (21)$$

and we bound separately the two terms of the right side. If $\hat{r} \leq R_n$, then $\hat{P}_n^{adapt} = \hat{P}_n^{(\hat{r})}$. As we already explained, if V_n is the number of vertices of $\hat{K}_n$, then $\hat{r} \leq V_n$ almost surely, and $\hat{K}_n = \hat{P}_n^{(V_n)}$. So we have, $|\hat{P}_n^{adapt} \triangle \hat{K}_n| = |\hat{P}_n^{(\hat{r})} \triangle \hat{P}_n^{(V_n)}| \leq \frac{CV_n \ln n}{n}$. Therefore, using the triangle inequality,

$$\begin{aligned}\mathbb{E}_K[|\hat{P}_n^{adapt} \triangle K| I(\hat{r} \leq R_n)] &\leq \mathbb{E}_K[|\hat{P}_n^{adapt} \triangle \hat{K}_n| I(\hat{r} \leq R_n)] \\ &\quad + \mathbb{E}_K[|\hat{K}_n \triangle K| I(\hat{r} \leq R_n)] \\ &\leq \mathbb{E}_K\left[\frac{CV_n \ln n}{n} I(\hat{r} \leq R_n)\right] + \mathbb{E}_K[|\hat{K}_n \triangle K|] \\ &\leq \frac{C\mathbb{E}_K[V_n] \ln n}{n} + \mathbb{E}_K[|\hat{K}_n \triangle K|]. \end{aligned} \tag{22}$$

We have the following lemma, which is a consequence of Efron's equality (1) and Theorem 4.

Lemma 3.

$$\sup_{K \in \mathcal{K}} \mathbb{E}_K[V_n] \lesssim_d n^{\frac{d-1}{d+1}}.$$

Therefore, (22) becomes, using Theorem 4,

$$\mathbb{E}_K[|\hat{P}_n^{adapt} \triangle K| I(\hat{r} \leq R_n)] \lesssim_d (\ln n) n^{-\frac{2}{d+1}}. \tag{23}$$

The second term is easily bounded. If $\hat{r} > R_n$, then $\hat{P}_n^{adapt} = \hat{K}_n$ and

$$\begin{aligned}\mathbb{E}_K[|\hat{P}_n^{adapt} \triangle K| I(\hat{r} > R_n)] &\leq \mathbb{E}_K[|\hat{K}_n \triangle K|] \\ &\lesssim_d n^{-\frac{2}{d+1}}, \end{aligned} \tag{24}$$

by Theorem 4.

From (23) and (24) we get

$$\mathbb{E}_K[|\hat{P}_n^{adapt} \triangle K|] \lesssim_d (\ln n) n^{-\frac{2}{d+1}}.$$

Theorem 5 is then proven.

References

[1] Bárány I. (2001) A note on Sylvester's four-point problem. Studia Math. Hungarica, 38 73 - 77.

[2] Bárány I. (1992) Random polytopes in smooth convex bodies, Mathematika 39, 81-92.

[3] Bárány I. (2007) Random Polytopes, convex bodies, and approximation, in Baddeley, A. J., Bárány, I., Shneider, R., Weil, W. (2004) Stochastic Geometry, C.I.M.E. Summer School, Martina Franca, Italy (W. Weil, ed.), Lecture Notes in Mathematics 1892, pp. 77-118, Springer, Berlin.

[4] Bárány I., Larman D. G. (1988) Convex bodies, economic cap coverings, random polytopes, Mathematika, 35, pp. 274-291.

[5] Blaschke W. (1923) Vorlesungen über Differentialgeometrue II, Springer-Verlag.

[6] Bronsted A. (1983) An Introduction to Convex Polytopes, Springer Verlag.

[7] Brunel V.-E. (2013) Adaptive estimation of convex polytopes and convex sets from noisy data, Electronic Journal of Statistics, Vol. 7, pp. 1301-1327

[8] Cuevas A. (2009) Set estimation : Another bridge between statistics and geometry, Boletìn de Estadìstica e Investigaciòn Operativa, Vol. 25, No. 2, pp. 71-85.

[9] Cuevas A., Fraiman R. (2010) Set estimation, in New Perspectives on Stochastic Geometry, W.S. Kendall and I. Molchanov, eds., pp. 374-397 Oxford University Press.

[10] Cuevas A., Rodríguez-Casal A. (2004) On boundary estimation. Adv. in Appl. Probab. 36, pp. 340-354.

[11] Devroye L., Wise G. L. (1980) Detection of abnormal behaviour via nonparametric estimation of the support, SIAM J. Applied Math. 38, 480-488.

[12] Dupin C. (1822) Applications de géométrie et de méchanique à la marine, aux ponts et chaussées, Paris.

[13] Dwyer R. A. (1988) On the Convex Hull of Random Points in a Polytope, J. Appl. Prob. 25, 688-699

[14] Efron B. (1965) The Convex Hull of a Random Set of Points, Biometrika, Vol. 52, No 3/4, pp. 331-343.

[15] Groemer H. (1974) On the mean value of the volume of a random polytope in a convex set, Arch. Math., 25, pp. 86-90.

[16] Guntuboyina A. (2012) Optimal rates of convergence for convex set estimation from support functions, Annals of Statistics, vol. 40, pp. 385-411.

[17] Korostelev, A.P.,Tsybakov, A.B. (1993) Minimax Theory of Image Reconstruction. Lecture Notes in Statistics, v.82. Springer, NY e.a.

[18] Lepski, O.V. (1991) Asymptotically minimax adaptive estimation i. upper bounds. optimally adaptive estimates. Theory Probab. Appl., 36:682-697.

[19] Mammen, E., Tsybakov, A. (1995) Asymptotical Minimax Recovery of Sets with Smooth Boundaries, The Annals of Statistics, Vol. 23, No. 2, pp. 502-524.

[20] Meyer M., Schütt C., Werner E. (2013) Affine invariant points, *to appear*.

[21] McMullen P. (1970) The maximum numbers of faces of a convex polytope, Mathematika, Volume 17, Issue 02, pp 179-184.

[22] Rényi A., Sulanke R. (1963) Über die konvexe Hülle von n zufällig gewählten Punkten. Z.Wahrscheinlichkeitsth. Verw. Geb. 2 pp. 75-84.

[23] Rényi A., Sulanke R. (1964) Über die konvexe Hülle von n zufällig gewählten Punkten. II, Z.Wahrscheinlichkeitsth. Verw. Geb. 3 pp. 138-147.

[24] Schütt C., Werner E. (1990) The convex floating body, Mathematica Scandinavica 66, 275–290.

[25] Schütt C. (1994) Random polytopes and affine surface area, Mathematische Nachrichten 170, 227–249.

[26] Tsybakov, A.B. (2009) Introduction to nonparametric estimation, Springer.

[27] Werner E. (2006) Floating bodies and Illumination bodies, Proceedings of the Conference "Integral Geormetry and Convexity", Wuhan 2004, World Scientific, Singapore.

[28] Wieacker J. A. (1987) Einige Probleme der polyedrischen Approximation, Diplomarbeit (Frieburg i. Br.)