Electronic Journal of Statistics
Vol. 7 (2013) 1301–1327
ISSN: 1935-7524
DOI: 10.1214/13-EJS804

Adaptive estimation of convex polytopes and convex sets from noisy data

Victor-Emmanuel Brunel

CREST-ENSAE & University of Haifa
e-mail: victor.emmanuel.brunel@ensae-paristech.fr

Abstract: We estimate convex polytopes and general convex sets in $\mathbb{R}^d$, $d \geq 2$ in the regression framework. We measure the risk of our estimators using a L^1-type loss function and prove upper bounds on these risks. We show, in the case of convex polytopes, that these estimators achieve the minimax rate. For convex polytopes, this minimax rate is $\frac{\ln n}{n}$, which differs from the parametric rate for non-regular families by a logarithmic factor, and we show that this extra factor is essential. Using polytopal approximations we extend our results to general convex sets, and we achieve the minimax rate up to a logarithmic factor. In addition we provide an estimator that is adaptive with respect to the number of vertices of the unknown polytope, and we prove that this estimator is optimal in all classes of convex polytopes with a given number of vertices.

Keywords and phrases: Adaptive estimation, approximation, convex set, minimax, polytope, regression.

Received November 2012.

1. Introduction

1.1. Definitions and notations

Let $d \geq 2$ be a positive integer. Assume that we observe a sample of n i.i.d. pairs $(X_i, Y_i), i = 1, \ldots, n$ such that $X_1, \ldots, X_n$ have the uniform distribution on $[0,1]^d$ and

$$Y_i = I(X_i \in G) + \xi_i, i = 1, \ldots, n. \tag{1}$$

The collection $X_1, \ldots, X_n$ is called the design. The errors $\xi_i, i = 1, \ldots, n$, are i.i.d. zero-mean random variables independent of the design, G is a subset of $[0,1]^d$, and $I(\cdot \in G)$ stands for the indicator function of the set G. Here we aim to estimate the set G in model (1).

A subset $\hat{G}_n$ of $[0,1]^d$ is called a set estimator, or simply, in our framework, an estimator, if it is a Borel set and if there exists a real measurable function f defined on $([0,1]^d \times \mathbb{R})^n$ such that $I(\cdot \in \hat{G}_n) = f(\cdot, X_1, Y_1, \ldots, X_n, Y_n)$.

If G is a measurable (with respect to the Lebesgue measure on $\mathbb{R}^d$) subset of $[0,1]^d$, we denote by $|G|_d$ or, when there is no possible confusion, simply by $|G|$, its Lebesgue measure and by $\mathbb{P}_G$ the probability measure with respect to the distribution of the collection of n pairs $(X_i, Y_i), i = 1, \ldots, n$. Where it is necessary to indicate the dependence on n we use the notation $\mathbb{P}_G^{\otimes n}$. If G_1 and

G_2 are two measurable subsets of $\mathbb{R}^d$ their Nikodym pseudo distance $d_1(G_1, G_2)$ is defined as

$$d_1(G_1, G_2) = |G_1 \triangle G_2|. \tag{2}$$

Note that if $\hat{G}_n$ is a set estimator and G is a measurable subset of $[0,1]^d$, then the quantity $|G \triangle \hat{G}_n| = \int_{[0,1]^d} |I(x \in \hat{G}_n) - I(x \in G)| dx$ is well defined and by Fubini's theorem it is measurable with respect to the probability measure $\mathbb{P}_G$. Therefore one can measure the accuracy of the set estimator $\hat{G}_n$ on a given class of sets in the minimax framework ; the risk of $\hat{G}_n$ on a class $\mathcal{C}$ is defined as

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

For all the estimators that we will define in the sequel we will be interested in upper bounds on their risk, which give information about the rate at which these risks tend to zero, when the number n of available observations tends to infinity. For a given class of subsets $\mathcal{C}$, the minimax risk on this class when n observations are available is defined as

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

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

In this paper, we study minimax rates of convergence on two classes of subsets of $[0,1]^d$: the class of all compact and convex sets, and the class of all convex polytopes with at most r vertices, where r is a given positive integer. Let $\mathcal{C}$ be a given class of subsets of $[0,1]^d$. One of our aims is to provide a lower bound on the minimax risk on the class $\mathcal{C}$. This lower bound can give much information on how close the risk of a given estimator is to the minimax risk on the class that we consider. If the rate (a sequence depending on n) of the upper bound on the risk of an estimator matches with 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.

We denote by ρ the Euclidean distance in $\mathbb{R}^d$, by $B_d(y, r)$ the d-dimensional closed Euclidean ball centered at $y \in \mathbb{R}^d$ with radius r, and by β_d the volume of the Euclidean unit ball in dimension $\mathbb{R}^d$. For any positive real number x, we denote by $\lfloor x \rfloor$ the greatest integer that is less or equal to x. Any convex set that we will consider in the following is assumed to be compact and with nonempty interior in the considered topological space.

1.2. Former results and contributions

Estimation of convex sets and, more generally, of sets, has been extensively studied in the previous decades (see the nice surveys in Cuevas [6] and Cuevas and Fraiman [7] and the references therein, and related topics in [15]). First works, in the 1960's, due to Rényi and Sulanke [26, 27], and Efron [10] were motivated by issues of stochastic geometry, discussed, for instance, in the book by Kendall and Moran [16] and [1]. Most of the works on estimation of convex sets dealt with models different than ours. Rényi and Sulanke, [26, 27], were the first to study the convex hull of a sample of n i.i.d. random points in the plane. They obtained exact asymptotic formulas for the expected area and the expected number of vertices when the points are uniformly distributed over a convex set, and when they have a Gaussian distribution. They showed that if the points are uniformly distributed over a convex set K in the plane $\mathbb{R}^2$, then the expected missing area $\mathbb{E}[|K\backslash\hat{K}|]$ of the convex hull $\hat{K}$ of the collection of these points is of the order

- $n^{-2/3}$ if the boundary of K is smooth,
- $r(\ln n)/n$ if K is a polygon with r vertices.

This result was generalized to any dimension, and we refer to [2] for an overview.

Estimation of convex sets in a multiplicative regression model has been investigated by Mammen and Tsybakov [22] and Korostelev and Tsybakov [19]. The design $(X_1, \ldots, X_n)$ may be either random or deterministic, in $[0,1]^d$. In [22] Mammen and Tsybakov proposed an estimator of a convex set G, based on likelihood maximization over an ε-net whose cardinality is bounded in terms of the metric entropy [9]. They showed, with no assumption on the design, that the rate of their estimator cannot be improved.

The additive model (1) has been studied in [18] and [19], 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 γ 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, as we will discuss later (Section 5). 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 will not make such an assumption in our work. Korostelev and Tsybakov [18, 19] also looked at some non-gaussian noises, making more general assumptions. Cuevas and Rodriguez-Cazal [8], and Pateiro Lopez [24], studied the properties of set estimators of the support of a density under several geometric assumptions on the boundary of the unknown set.

One problem has not been investigated yet: What is the minimax rate of convergence if one assumes that the unknown set G in model (1) is a convex polytope with a bounded number of vertices ? This question can be reformulated in the framework of boundary fragments: What is the minimax rate of convergence if G is a fragment which belongs to a parametric family ? In the method used in [18] and [19], 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. Thus, a parametric approximation of the fragment G and not directly G itself is estimated. This idea is exploited in the present work, when we estimate convex sets by using polytopal approximations. It is easy to show that the rate of convergence of the estimator, when G belongs to a parametric family of boundary fragments of dimension M, is of the order M/n. But this is true under the assumption of uniform separation from 0 and 1. We will see below that if this assumption is dropped in a special case of a parametric family (convex polytopes with a bounded number of vertices), an extra logarithmic factor appears in the rate of the minimax risk.

In order to estimate convex sets, we will first approximate a convex set by a convex polytope, and then estimate that polytope. There is an extensive literature on polytopal approximation of convex sets (cf. [23, 11], and the references cited therein), which is of essential use in this paper. This method provides an explicit estimator but it will be shown to be suboptimal. This is why we will propose another method, which is rather classical, using the metric entropy, and yelds a rate-minimax estimator.

For an integer $r \geq d+1$, we denote by $\mathcal{P}_r$ the class of all convex polytopes in $[0,1]^d$ with at most r vertices. This class may be embedded into the finite dimensional space $\mathbb{R}^{dr}$ since any polytope is completely defined by the coordinates of its vertices. Hence, one may expect that the problem of estimating $G \in \mathcal{P}_r$, for a given r, is parametric and therefore the minimax risk $\mathcal{R}_n(\mathcal{P}_r)$ would be of order $1/n$, cf. [13]. However this is not the case. In Section 2.1, we propose an estimator that almost achieves this rate, up to a logarithmic factor. Moreover, we prove an exponential deviation inequality for the Nikodym distance between the estimator and the true polytope. Such an exponential inequality is of interest because it is much stronger than an upper bound on the risk of the estimator, and it is the key for adaptive estimation, as we will see later. In Section 2.2, we show that this estimator has the minimax rate of convergence, so that the logarithmic factor in the rate is unavoidable. In Section 3, we extend the exponential deviation inequality of Section 2 and cover minimax estimation of any convex set. In Section 4, we propose an estimator that is adaptive to the number of vertices of the estimated polytope, using as a convention that a non polytopal convex set can be considered as a convex polytope with infinitely many vertices. In Section 5 we discuss our results, and Section 6 is devoted to the proofs. We will try as much as possible to use geometric and explicit methods, and elementary arguments in the proofs.

2. Estimation of convex polytopes

2.1. Upper bound

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

We estimate P_0 by a polytope in $\mathcal{P}_r^{(n)}$ that minimizes some criterion. The criterion that we use is the sum of squared errors

$$\mathcal{A}(P, \{(X_i, Y_i)\}_{i=1,\dots,n}) = \sum_{i=1}^{n} (1 - 2Y_i) I(X_i \in P). \tag{3}$$

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 ξ_i are supposed to be Gaussian, then minimization of $\mathcal{A}(P)$ is equivalent to maximization of the likelihood. Consider the set estimator of P_0 defined as

$$\hat{P}_n^{(r)} \in \underset{P \in \mathcal{P}_r^{(n)}}{\operatorname{argmin}} \mathcal{A}(P). \tag{4}$$

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

Let us introduce the following assumption on the law of the ξ_i.

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

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

where σ is a positive number.

Note that if the errors $\xi_i, i = 1, \dots, n$, are i.i.d. zero-mean Gaussian random variables, then Assumption 1 is satisfied.

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

Theorem 1. *Let $r \ge d+1$ be an integer, and $n \ge 2$. Consider the model (1), with $G = P$, where $P \in \mathcal{P}_r$. Let Assumption 1 be satisfied. For the estimator $\hat{P}_n^{(r)}$, there exist two positive constants C_1 and C_2, which depend on d and σ only, such that*

$$\sup_{P \in \mathcal{P}_r} \mathbb{P}_P \left[n \left(|\hat{P}_n^{(r)} \triangle P| - \frac{2dr \ln n}{C_2 n} \right) \ge x \right] \le C_1 e^{-C_2 x}, \forall x > 0.$$

The explicit expressions for the constants C_1 and C_2 are given in the proof. 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 following result.

Corollary 1. *Let $n \ge 2$. Let the assumptions of Theorem 1 be satisfied. Then, for any positive number q, there exists a constant A_q which depends on σ, d and q such that*

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

The explicit form of the constant A_q can be derived from the proof. Note that the construction of our estimator does not require the knowledge of σ.

2.2. Lower bound

Corollary 1 gives an upper bound of the order $\frac{\ln n}{n}$ for the risk of our estimator $\hat{P}_n^{(r)}$. The next result shows that $\frac{\ln n}{n}$ is the minimax rate of convergence on the class $\mathcal{P}_r$.

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

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

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

Corollary 1 together with Theorem 2 gives the following bound on the class $\mathcal{P}_r$, in the case of Gaussian noise with variance σ^2.

$$0 < \alpha^2 \sigma^2 \leq \frac{n}{\ln n} \mathcal{R}_n(\mathcal{P}_r) \leq A_1 r < \infty,$$

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 $\mathcal{P}_{d+1}$ and we use that $\mathcal{P}_r \supseteq \mathcal{P}_{d+1}$, for $r \geq d+1$. The minimax rate of convergence on any of the classes $\mathcal{P}_r, r \geq d+1$, 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 (1), the minimax risk for the estimation of segments in $[0,1]$ is of order $(\ln n)/n$.

3. Estimation of general convex sets

3.1. A first estimator

Denote by $\mathcal{C}_d$ the class of all convex sets included in $[0,1]^d$. Now we aim to estimate convex sets in the same model, without any assumption of the form of the unknown set. If G is a convex set in model (1), 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 2 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 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 defined in (2). At the same time, when r increases the upper bound of Corollary 1 increases as well. Therefore r should be chosen according to the bias-variance tradeoff.

For any integer $r \geq d+1$ consider again the estimator $\hat{P}_n^{(r)}$ defined in (4). However, now we chose a value for r that depends on n in order to achieve the bias-variance tradeoff.

Theorem 3. *Let $n \geq 2$. Consider the model (1) 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 (4). Let Assumption 1 be satisfied. Then, there exist positive constants C_1, C_2 and C_3, which depend on d and σ only, such that*

$$\sup_{G \in \mathcal{C}_d} \mathbb{P}_G \left[n \left(|\hat{P}_n^{(r)} \triangle G| - \left(\frac{C_3 \ln n}{n} \right)^{2/(d+1)} \right) \geq x \right] \leq C_1 e^{-C_2 x}, \forall x > 0.$$

The constants C_1 and C_2 are the same as in Theorem 1, and C_3 is given explicitly in the proof of the theorem. From Theorem 3 we get the next corollary.

Corollary 2. *Let the assumptions of Theorem 3 be satisfied. Then, for any positive number q there exists a positive constant A'_q which depends on σ, d and q such that*

$$\sup_{G\in\mathcal{C}_d} \mathbb{E}_G\left[|\hat{P}_n^{(r)}\triangle G|^q\right] \le A'_q\left(\frac{\ln n}{n}\right)^{\frac{2q}{d+1}}.$$

The explicit expression for A'_q can be derived in the same way as for the constants A_q in Corollary 1. Note again that the construction of our estimator does not require the knowledge of σ.

Corollary 2 shows that the estimator given in Theorem 3 achieves the rate $\left(\frac{\ln n}{n}\right)^{\frac{2}{d+1}}$. This estimator has an advantage: it is computable and constructed using 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 $\mathcal{C}_d$, and is mainly of theoretical interest. We develop this in the next subsection.

3.2. Improvement of the upper bound

We propose an estimator whose construction is similar to [22], where the multiplicative model was considered. Bronshtein [4] proves the following upper bound on the metric entropy of $\mathcal{C}_d$. If $d \ge 2$ and δ is a positive number, then there exists a δ-net in $\mathcal{C}_d$ containing not more than $\tau_1 e^{\tau_2\delta^{-(d-1)/2}}$ sets, where τ_1 and τ_2 are positive numbers and depend on d only. Another result on the metric entropy of $\mathcal{C}_d$ was obtained by Dudley [9], but in a weaker form than Bronshtein's upper bound, and could not be used in our analysis.

Let $\delta = n^{-2/(d+1)}$. Let $N = \lfloor \tau_1 e^{\tau_2\delta^{-(d-1)/2}}\rfloor$ and $G_1, \dots, G_N$ be a δ-net of $\mathcal{C}_d$. Let $G \in \mathcal{C}_d$ be the true set in model (1). We define the estimator $\tilde{G}_n = G_{\hat{j}}$, where $\hat{j}$ is the index of a set in the δ-net of $\mathcal{C}_d$ that we introduced above, which minimizes the sum of squared errors, as in Section 2.1:

$$\hat{j} \in \underset{j=1,\dots,N}{\operatorname{argmin}}\ \mathcal{A}(G_j), \tag{5}$$

where $\mathcal{A}$ is defined in (3). Note again that $\hat{j}$ may not be unique. We have the following result.

Theorem 4. *Let $n \ge 1$. Consider the model (1) with G any convex subset of $[0,1]^d$. Set $\tilde{G} = G_{\hat{j}}$, where $\hat{j}$ is defined in (5). Let Assumption 1 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 σ only, such that*

$$\sup_{G\in\mathcal{C}_d} \mathbb{P}_G\left[|\tilde{G}\triangle G| \ge C_0 n^{-2/(d+1)} + \frac{x}{n}\right] \le \tau_1 e^{-C_2 x}, \forall x > 0.$$

Here, $C_0 = \frac{\tilde{C}_1+\tau_2}{C_1}$ and the constants $\tilde{C}_1$, C_1 and C_2 are given in the proof of Theorem 1. Note again that the construction of the estimator $\tilde{G}$ does not require the knowledge of the noise level σ.

As for the estimator of the previous section, we derive from Theorem 4 an upper bound on the risk of the estimator $\tilde{G}$, and we have the following result.

Corollary 3. *Let the assumptions of Theorem 4 be satisfied. Then, for any positive number q there exists a positive constant A''_q such that*

$$\sup_{G\in\mathcal{C}_d} \mathbb{E}_G\left[|\tilde{G}\triangle G|^q\right] \leq A''_q n^{-\frac{2q}{d+1}}.$$

3.3. Lower bound

In this section we give a lower bound on the minimax risk on the class $\mathcal{C}_d$ of all convex sets in $[0,1]^d$.

Theorem 5. *Let $n \geq 125$. Consider the model (1) and assume that the errors ξ_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 σ, such that for any estimator $\hat{C}$,*

$$\sup_{C\in\mathcal{C}_d} \mathbb{E}_C\left[|C\triangle\hat{C}|\right] \geq C_4 n^{-2/(d+1)}.$$

The explicit form of the constant C_4 can be found in the proof of the theorem.

From Theorem 5 and Corollary 3, one gets, for $n \geq 125$ and in the case of Gaussian noise,

$$0 < C_4 \leq n^{\frac{2}{d+1}}\mathcal{R}_n(\mathcal{C}_d) \leq A''_1 < \infty,$$

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

4. Adaptive estimation

In Section 2, 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 $\mathcal{P}_r$. The idea of the following comes from Lepski's method for adaptation (see [21], or [5], 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 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 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 $\mathcal{C}_d$ instead of $\mathcal{P}_r$. Let us denote, for $r = d+1, \dots, R_n - 1$, $\hat{Q}_n^{(r)} = \hat{P}_n^{(r)}$, and $\hat{Q}_n^{(R_n)} = \tilde{G}$, 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 1 and 4 respectively.

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

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 σ through the definition of $\hat{r}$; it depends on the constant C_2 of Theorem 1, which depends itself on σ. We then have the following theorem.

Theorem 6. *Let $n \geq 2$. Let Assumption 1 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 σ only, such that the adaptive estimator $\hat{P}_n^{adapt}$ satisfies the following inequality.*

$$\sup_{d+1 \leq r \leq \infty} \sup_{P \in \mathcal{P}_r} \mathbb{E}_P \left[\phi_{n,r}^{-1} |\hat{P}_n^{adapt} \triangle P| \right] \leq C_5,$$

where $\mathcal{P}_\infty = \mathcal{C}_d$.

Thus, we show that one and the same estimator $\hat{P}_n^{adapt}$ attains the optimal rate simultaneously on all the classes $\mathcal{P}_r, d+1 \leq r$, and on the class $\mathcal{C}_d$ 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.

5. Discussion

Theorem 4 showed that the logarithmic factor in Corollary 2 can be dropped and that the minimax rate of convergence on the class $\mathcal{C}_d$ is $n^{-2/(d+1)}$. However, Theorems 1 and 2 show that the logarithmic factor is significant in the case of convex polytopes. We try to understand what brings this logarithmic factor in one case and not in the other.

Let us first answer the following question: What makes the estimation of sets on a given class $\mathcal{C} \subseteq \mathcal{C}_d$ difficult in the studied model ? First, it is the complexity of the class, expressed in terms of the metric entropy. We worked with this notion of complexity in our Theorem 5, using δ-nets. The second issue is how detectable the individual sets of the given class are, in our model. If the unknown subset G is too small, then, with high probability, it contains no points of the design. Conditionally to this, all the data have the same distribution and no information in the sample can be used in order to detect G. A subset 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 gives a lower bound on the rate of the minimax risk. In [14], Janson studied asymptotic properties on 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 on the minimax risk in Theorem 2 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 is the shape of their boundary, when we chose their volume of order no larger than $\frac{\ln n}{n}$. This shows that the rate $1/n$, which would come from the complexity of the parametric class $\mathcal{P}_r$, is not the right minimax rate of convergence: the order $(\ln n)/n$ is dominating. On the other hand, the proof of our lower bound of the order $n^{-2/(d+1)}$ for general convex sets 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 $(\ln n)/n$, and therefore seems to determine the best lower bound achievable on the minimax risk on the class $\mathcal{C}_d$.

Let us add two remarks in this discussion. First, if $d=2$ it is easy to prove a better lower bound on the minimax risk on the class $\mathcal{P}_r$, for any integer $r\geq 3$, using the scheme of the proof of Theorem 5 in the case $d=2$.

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

for some positive constants λ_1 and λ_2. It seems to us that this lower bound should remain true for any value of d, with constants λ_1 and λ_2 which would depend on d and σ. If $\ln n$ is larger than r, then the minimax risk is controlled from below by the rate $\frac{\ln n}{n}$. If not, i.e., if the number of vertices of the unknown convex polytope can be arbitrarily large, the order of the risk has a lower bound of the order $\frac{r}{n}$.

Our second remark is the following. Let μ_0 be a fixed positive number. If one considers the subclass $\mathcal{P}_r'(\mu_0)=\{P\in\mathcal{P}_r : |P|\geq\mu_0\}$, then subsets of $[0,1]^d$ with too small volume are excluded. Therefore, the construction used in the proof of Theorem 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.

6. Proofs

Proof of Theorem 1 Let $P_0\in\mathcal{P}_r$ be the true polytope. We have the following lemma, proven in Section 7.

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

$$|P^*\triangle P|\leq\frac{2d^{d+1}(3/2)^d\beta_d}{n}. \tag{6}$$

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

$$\mathbb{P}_{P_0}\left[|\hat{P}_n^{(r)} \triangle P_0| \geq \epsilon\right] = \mathbb{P}_{P_0}\left[\exists P \in \mathcal{P}_r^{(n)} : \mathcal{A}(P) \leq \mathcal{A}(P^*), |P \triangle P_0| \geq \epsilon\right], \quad (7)$$

where P^* is a convex polytope chosen in $\mathcal{P}_r^{(n)}$ which satisfies the inequality $|P^* \backslash P_0| \leq \frac{2d^{d+1}(3/2)^d \beta_d}{n}$, cf. (6). For any P we have, by a simple algebra,

$$\mathcal{A}(P) - \mathcal{A}(P^*) = \sum_{i=1}^{n} Z_i, \quad (8)$$

where

$$\begin{aligned} Z_i = & I(X_i \in P) - I(X_i \in P^*) - 2I(X_i \in P_0)\left[I(X_i \in P) - I(X_i \in P^*)\right] \\ & - 2\xi_i\left[I(X_i \in P) - I(X_i \in P^*)\right], \ i = 1, \ldots, n. \end{aligned}$$

The random variables Z_i depend on P but we omit this dependence in the notation. Therefore (7) implies that

$$\begin{aligned} \mathbb{P}_{P_0}\left[|\hat{P}_n^{(r)} \triangle P_0| \geq \epsilon\right] & \leq \sum_{P \in \mathcal{P}_r^{(n)} : |P \triangle P_0| \geq \epsilon} \mathbb{P}_{P_0}\left[\sum_{i=1}^{n} Z_i \leq 0\right] \\ & \leq \sum_{P \in \mathcal{P}_r^{(n)} : |P \triangle P_0| \geq \epsilon} \mathbb{E}_{P_0}\left[\exp\left(-u \sum_{i=1}^{n} Z_i\right)\right], \quad (9) \end{aligned}$$

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

$$\mathbb{P}_{P_0}\left[|\hat{P}_n^{(r)} \triangle P_0| \geq \epsilon\right] \leq \sum_{P \in \mathcal{P}_r^{(n)} : |P \triangle P_0| \geq \epsilon} \prod_{i=1}^{n} \mathbb{E}_{P_0}\left[\exp\left(-uZ_i\right)\right]. \quad (10)$$

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

$$\begin{aligned} & \mathbb{E}_{P_0}\left[\exp(-uZ_1)\right] = \mathbb{E}_{P_0}\left[\mathbb{E}_{P_0}\left[\exp(-uZ_1)|X_1\right]\right] \\ & \quad = \mathbb{E}_{P_0}\left[\exp\left(-uW + 2uI(X_1 \in P_0)W\right)\mathbb{E}_{P_0}\left[\exp\left(2u\xi_1 W\right)|X_1\right]\right] \\ & \quad = \mathbb{E}_{P_0}\left[\exp\left(-uW + 2uI(X_1 \in P_0)W\right)\exp\left(2\sigma^2 u^2 I(X_1 \in P \triangle P^*)\right)\right] \\ & \quad = \mathbb{E}_{P_0}\left[\exp\left(2\sigma^2 u^2 I(X_1 \in P \triangle P^*) - uW + 2uI(X_1 \in P_0)W\right)\right]. \quad (11) \end{aligned}$$

We will now reduce the last expression in (11). It is convenient to use Table 1: 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 (11), that is $\exp\left(2\sigma^2 u^2 I(X_1 \in P \triangle P^*) - uW + 2uI(X_1 \in P_0)W\right)$.

Hence one can write

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

TABLE 1
Values of $\exp\left(2\sigma^2u^2I(X_1\in P\triangle P^*)-uW+2uI(X_1\in P_0)W\right)$

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

Besides by the triangle inequality,

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

which implies

$$\begin{aligned}\mathbb{E}_{P_0}\left[\exp(-uZ_1)\right]\leq\ & 1-|P\triangle P_0|+|P^*\triangle P_0|+e^{2\sigma^2u^2+u}\left(|P_0\backslash P^*|+|P^*\backslash P_0|\right)\\ &+e^{2\sigma^2u^2-u}\left(|P_0\backslash P|+|P\backslash P_0|\right)\\ \leq\ & 1-|P\triangle P_0|+|P^*\triangle P_0|+e^{2\sigma^2u^2+u}|P^*\triangle P_0|+e^{2\sigma^2u^2-u}|P\triangle P_0| \\ \leq\ & 1-|P\triangle P_0|\left(1-e^{2\sigma^2u^2-u}\right)+\frac{2d^{d+1}(3/2)^d\beta_d}{n}\left(1+e^{2\sigma^2u^2+u}\right).\end{aligned}\tag{12}$$

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

$$\mathbb{E}_{P_0}\left[\exp(-uZ_1)\right]\leq 1-\epsilon\left(1-e^{-\frac{1}{4\sigma^2}}\right)+\frac{2d^{d+1}(3/2)^d\beta_d}{n}\left(1+e^{\frac{3}{8\sigma^2}}\right).\tag{13}$$

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. From (10) and (13), and by the independence of Z_i's we have

$$\mathbb{P}_{P_0}\left[|\hat{P}_n^{(r)}\triangle P_0|\geq\epsilon\right]\leq\sum_{P\in\mathcal{P}_r^{(n)}:|P\triangle P_0|\geq\epsilon}\left(1-C_2\epsilon+\frac{2d^{d+1}(3/2)^d\beta_d\tilde{C}_1}{n}\right)^n\tag{14}$$

$$\begin{aligned}&\leq(n+1)^{dr}\left(1-C_2\epsilon+\frac{2d^{d+1}(3/2)^d\beta_d\tilde{C}_1}{n}\right)^n\\ &\leq\exp\left(dr\ln(n+1)-C_2\epsilon n+2d^{d+1}(3/2)^d\beta_d\tilde{C}_1\right)\\ &\leq\exp\left(2dr\ln n-C_2\epsilon n+2d^{d+1}(3/2)^d\beta_d\tilde{C}_1\right),\end{aligned}\tag{15}$$

where $C_1 = \exp\big(2d^{d+1}(3/2)^d\beta_d\tilde{C}_1\big)$, 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 Corollary 1 Corollary 1 follows directly from Theorem 1 and Fubini's theorem. Indeed, if we denote $Z := |\hat{P}_n^{(r)} \triangle P_0|$ and by $\mathbb{P}_Z$ its distribution measure, then Z is a continuous and nonnegative random variable and we have, by Fubini's theorem, that

$$\begin{aligned}
&\mathbb{E}_{P_0}[Z^q] = q\int_0^\infty u^{q-1}\mathbb{P}_Z[Z \geq u]du \\
&\leq q\int_0^{\frac{2dr\ln n}{C_2 n}} u^{q-1}du + q\int_0^\infty \left(u + \frac{2dr\ln n}{C_2 n}\right)^{q-1}\mathbb{P}_Z\left[Z \geq u + \frac{2dr\ln n}{C_2 n}\right]du \\
&= \left(\frac{2dr\ln n}{C_2 n}\right)^q + q\int_0^\infty \left(u + \frac{2dr\ln n}{C_2 n}\right)^{q-1}\mathbb{P}_Z\left[n\left(Z - \frac{2dr\ln n}{C_2 n}\right) \geq nu\right]du \\
&\leq \left(\frac{2dr\ln n}{C_2 n}\right)^q + q\int_0^\infty \left(u + \frac{2dr\ln n}{C_2 n}\right)^{q-1} C_1 e^{-C_2 nu}du, \quad \text{by Theorem 1}, \\
&\leq \left(\frac{2dr\ln n}{C_2 n}\right)^q + C_1 q\max(1, 2^{q-1})\int_0^\infty \left(u^{q-1} + \left(\frac{2dr\ln n}{C_2 n}\right)^{q-1}\right)e^{-C_2 nu}du \\
&\leq A_q\left(\frac{r\ln n}{n}\right)^q,
\end{aligned}$$

for some positive constant A_q which depends on σ, d and q only. Note that the fifth step of this proof comes from the easy fact that for any positive numbers a and b, $(a+b)^{q-1} \leq 2^{q-1}(a^{q-1}+b^{q-1})$ if $q-1 > 0$, and $(a+b)^{q-1} \leq a^{q-1}+b^{q-1}$ if $q-1 \leq 0$, and the sixth comes from the equality $\int_0^\infty v^{q-1}e^{-v}dv = (q-1)!$. □

Proof of Theorem 2 This proof is a simple application of Fano's method, see Corollary 2.6 in [28] or, for a more general setting, [12]. Let M be a positive integer, and $h = \frac{1}{M+1}$. Let $T_k, k = 0, \ldots, M$ be M disjoint convex polytopes in $\mathcal{P}_{d+1}$ and with same volume: $|T_0| = \cdots = |T_M| = h/2$. 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 $k = 1, \ldots, M$ we denote by $\mathbb{P}_k$ the probability distribution of the observations $(X_i, Y_i), i = 1, \ldots, n$ when $G = T_k$ in (1), and by $\mathbb{E}_k$ the expectation with respect to this distribution. 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 [28] with the pseudo distance defined in (2), we set for $r \geq d+1$ the following inequality

$$\inf_{\hat{P}} \sup_{P \in \mathcal{P}_{d+1}} \mathbb{E}_P\left[|\hat{P} \triangle P|\right] \geq \frac{1}{M+1}\left(\frac{\ln(M+1) - \ln 2}{\ln M} - \alpha\right).$$

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

$$\inf_{\hat{P}} \sup_{P \in \mathcal{P}_{d+1}} \mathbb{E}_P\left[|\hat{P} \triangle P|\right] \geq \frac{\alpha}{M+1} \geq \frac{\alpha}{2M} \geq \frac{\alpha \ln n}{\gamma n} \geq \frac{\alpha^2\sigma^2 \ln n}{n}.$$

This immediately implies Theorem 2. $\square$

Proof of Theorem 3 The idea of the proof is very similar to that of Theorem 1. Here we need to control an extra bias term, due to the approximation of R by a r-vertex convex polytope. We give the following lemma (cf. [11]).

Lemma 2. *Let $r \geq d+1$ be a positive integer. For any convex set $G \subseteq \mathbb{R}^d$ there exists a convex polytope P_r with at most r vertices such that*

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

where A is a positive constant that does not depend on r, d and G.

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

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

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 (7) and (9) we have

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

Repeating the argument in (8) with G instead of P_0 we set

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

where

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

The rest of the proof is very similar to the one of Theorem 1. Indeed, replacing P_0 by G in that proof between (7) and (12), and $\frac{2d^{d+1}(3/2)^d\beta_d}{n}$ by $\frac{2d^{d+1}(3/2)^d\beta_d}{n} + \frac{Ad}{r^{2/(d-1)}}$ in (13) and (15) one gets

$$\begin{aligned}&\mathbb{P}_G\left[|\hat{P}_n^{(r)}\triangle G| \geq \epsilon\right] \\ &\leq \sum_{P\in\mathcal{P}_r^{(n)}:|P\triangle G|\geq\epsilon} \left(1 - C_2\epsilon + \tilde{C}_1\left(\frac{Ad}{r^{2/(d-1)}} + \frac{2d^{d+1}(3/2)^d\beta_d}{n}\right)\right)^n \\ &\leq (n+1)^{dr}\left(1 - C_2\epsilon + \tilde{C}_1\left(\frac{Ad}{r^{2/(d-1)}} + \frac{2d^{d+1}(3/2)^d\beta_d}{n}\right)\right)^n \\ &\leq \exp\left(2dr\ln n - C_2\epsilon n + \tilde{C}_1\left(\frac{Adn}{r^{2/(d-1)}} + 2d^{d+1}(3/2)^d\beta_d\right)\right).\end{aligned}$$

Therefore if we set $\epsilon = \frac{2dr\ln n}{C_2 n} + \frac{\tilde{C}_1 Ad}{C_2 r^{2/(d-1)}} + \frac{x}{n}$ for a positive number x, we get the following deviation inequality

$$\mathbb{P}_G\left[n\left(|\hat{P}_n^{(r)}\triangle G| - \frac{2dr\ln n}{C_2 n} - \frac{\tilde{C}_1 Ad}{C_2 r^{2/(d-1)}}\right) \geq x\right] \leq C_1 e^{-C_2 x},$$

where the constants are defined as in the previous section. That ends the proof of Theorem 3 by choosing $r = \lfloor\left(\frac{n}{\ln n}\right)^{\frac{d-1}{d+1}}\rfloor$, and the constant C_3 is given by

$$C_3 = \left(1 + \tilde{C}_1 A\right)\frac{d}{C_2} = \left(1 + (1 + e^{3/(8\sigma^2)})A\right)\frac{d}{1 - e^{-1/(4\sigma^2)}}. \qquad \square$$

Proof of Theorem 4 The proof is similar to the proof of Theorem 1. The difference is that we now use a δ-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 δ-net whose distance to G is not greater than δ:

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

It follows, from the definition of the estimator that

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

This leads to the same inequality as (14) where the sum is now over $i = 1,\ldots,N$, 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 $\tilde{C}_1\delta$.

$$\begin{aligned}
\mathbb{P}_G\left[|\tilde{G}\triangle G| \geq \epsilon\right] &\leq \sum_{i\in\{1,\ldots,N\}:|G_i\triangle G|\geq\epsilon} \left(1 - C_2\epsilon + \tilde{C}_1\delta\right)^n \\
&\leq N\exp\left(-C_2\epsilon n + \tilde{C}_1\delta n\right) \\
&\leq \tau_1\exp\left(-C_2\epsilon n + \tilde{C}_1\delta n + \tau_2\delta^{-(d-1)/2}\right) \\
&\leq \tau_1\exp\left(-C_2\epsilon n + (\tilde{C}_1 + \tau_2)\delta n\right),
\end{aligned}$$

since our choice of δ guarantees that $n\delta = \delta^{-(d-1)/2}$. Hence, by choosing $\epsilon = \frac{x}{n} + \frac{\tilde{C}_1+\tau_2}{C_2}$, we get Theorem 4.

Proof of Theorem 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 [20] (which is similar to the proof of Assouad's lemma, see [28]). 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 $D_1,\ldots,D_M$, and for any $\omega = (\omega_1,\ldots,\omega_M) \in \{0,1\}^M$ we denote by G_ω the set made of G out of which we took all the caps D_j for which $\omega_j = 0$, $j = 1,\ldots,M$.

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

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

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

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

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

For two probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined on the same probability space and having densities denoted respectively by p and q with respect to a common measure ν (we also denote by $d\mathbb{P} = pd\nu$ and $d\mathbb{Q} = qd\nu$), we call $H(\mathbb{P}, \mathbb{Q})$ the Hellinger distance between $\mathbb{P}$ and $\mathbb{Q}$, defined as

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

Some useful properties of the Hellinger distance can be found in [28], Section 2.4.

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.

$$\begin{aligned}
&\sup_{G \in \mathcal{C}_2} \mathbb{E}_G \left[|G \triangle \hat{G}| \right] \\
&\geq \frac{1}{2^M} \sum_{\omega \in \{0,1\}^M} \mathbb{E}_{G_\omega} \left[|G_\omega \triangle \hat{G}| \right] \\
&\geq \frac{1}{2^M} \sum_{\omega \in \{0,1\}^M} \sum_{j=1}^M \mathbb{E}_{G_\omega} \left[|(G_\omega \cap A_j) \triangle (\hat{G} \cap A_j)| \right] \\
&= \frac{1}{2^M} \sum_{j=1}^M \sum_{\omega \in \{0,1\}^M} \mathbb{E}_{G_\omega} \left[|(G_\omega \cap A_j) \triangle (\hat{G} \cap A_j)| \right] \\
&= \frac{1}{2^M} \sum_{j=1}^M \sum_{\omega_1, \dots, \omega_{j-1}, \omega_{j+1}, \dots, \omega_M} \!\!\!\! \cdots \sum \Big(\mathbb{E}_{G_\omega^{(j,0)}} \left[|(G_\omega^{(j,0)} \cap A_j) \triangle (\hat{G} \cap A_j)| \right] \\
&\qquad\qquad + \mathbb{E}_{G_\omega^{(j,1)}} \left[|(G_\omega^{(j,1)} \cap A_j) \triangle (\hat{G} \cap A_j)| \right] \Big). \qquad (16)
\end{aligned}$$

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

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

$$\geq \frac{h}{2}\left(1-\frac{H^2(\mathbb{P}^{\otimes n}_{G_\omega^{(j,0)}},\mathbb{P}^{\otimes n}_{G_\omega^{(j,1)}})}{2}\right)^2$$

$$= \frac{h}{2}\left(1-\frac{H^2(\mathbb{P}_{G_\omega^{(j,0)}},\mathbb{P}_{G_\omega^{(j,1)}})}{2}\right)^{2n}, \tag{17}$$

using properties of the Hellinger distance (cf. Section 2.4. in [28]). To compute the Hellinger distance between $\mathbb{P}_{G_\omega^{(j,0)}}$ and $\mathbb{P}_{G_\omega^{(j,1)}}$ we use the following lemma.

Lemma 3. *For any integer $d \geq 2$, if G_1 and G_2 are two subsets of $[0,1]^d$, then*

$$H^2(\mathbb{P}_{G_1},\mathbb{P}_{G_2}) = 2(1-e^{-\frac{1}{8\sigma^2}})|G_1\triangle G_2|.$$

Then if we denote by $C_9 = 1-e^{-\frac{1}{8\sigma^2}}$, it follows from (16) and (17) that

$$\begin{aligned}\sup_{G\in\mathcal{C}_2}\mathbb{E}_G\left[|G\triangle\hat{G}|\right] &\geq \frac{1}{2^M}M2^{M-1}\frac{h}{2}(1-C_9h)^{2n}\\ &\geq \frac{Mh}{4}(1-C_9h)^{2n}\\ &\geq \frac{\pi^3}{12M^2}(1-\pi^3C_9/M^3)^{2n}.\end{aligned}$$

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

$$\pi^3C_9/M^3 \leq \pi^3C_9/6^3 = \frac{\pi^3}{6^3}\left(1-\exp(-\frac{1}{8\sigma^2})\right) \leq \frac{\pi^3}{6^3} < 1,$$

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

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

where $C_{14} = \frac{\pi^3}{12}\exp\left(\frac{432\ln(1-\pi^3/216)(1-e^{-\frac{1}{8\sigma^2}})}{\pi^3}\right)$ is a positive constant that depends only on σ. 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 [17].

Definition 1. *Let (S,ρ) be a metric space and η a positive number. A family $\mathcal{Y} \subseteq S$ is called an η-packing family if and only if $\rho(y,y') \geq \eta$, for $(y,y') \in \mathcal{Y}$ with $y \neq y'$. An η-packing family is called maximal if and only if it is not strictly included in any other η-packing family. A family $\mathcal{Z}$ is called an η-net if and only if for all $x \in S$, there is an element $z \in \mathcal{Z}$ which satisfies $\rho(x,z) \leq \eta$.*

We now give a Lemma.

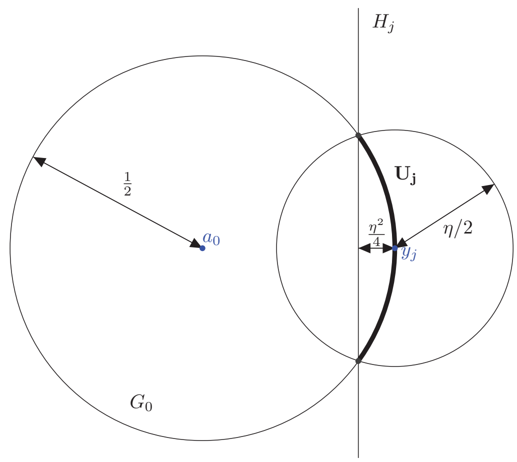

FIG 1. *Construction of the hypotheses.*

Lemma 4. *Let S be the sphere with center $a_0 = (1/2, \dots, 1/2) \in \mathbb{R}^d$ and radius $1/2$, and ρ the Euclidean distance in $\mathbb{R}^d$. We still denote by ρ its restriction on S. Let $\eta \in (0, 1)$. Then any η-packing family of (S, ρ) is finite, and any maximal η-packing family has a cardinality M_η that satisfies the inequalities*

$$\frac{d\sqrt{2\pi}}{2^{d-1}\sqrt{d+2}\eta^{d-1}} \leq M_\eta \leq \frac{4^{d-2}\sqrt{2\pi d}}{3^{(d-3)/2}\eta^{d-1}}. \tag{18}$$

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 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 η-packing family of $S = \partial G_0$. The integer M_η satisfies (18) by Lemma 4. For $j \in \{1, \dots, M_\eta\}$, 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

$$G_\omega = G_0 \backslash (\bigcap_{j=1,\dots,M_\eta : \omega_j = 0} H_j^-).$$

The set G_ω 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

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

since as mentioned before $\eta^2/4$ is the height of the cap cut off by H_j, or in order words the distance between y_j and the hyperplane H_j, 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

$$\frac{3^{(d-1)/2} \eta^{d+1} \beta_{d-1}}{2^{3d}(d+1)} \le |G_{\omega^{(j,0)}} \triangle G_{\omega^{(j,1)}}| \le \frac{\eta^{d+1} \beta_{d-1}}{2^{2d+1}(d+1)}. \tag{19}$$

Now, continuing (16) and (17), replacing M by M_η and h by the lower bound in (19) and using lemmas 3 and 4, we get that

$$\sup_{G \in \mathcal{C}_d} \mathbb{E}_G \left[|G \triangle \hat{G}| \right] \ge C_8 \eta^2 \left(1 - C_9 \eta^{d+1}\right)^{2n}, \tag{20}$$

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

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

and this shows that $C_9 < 1$. Therefore, since $\eta < 1$ as well, the concavity of the logarithm leads (20) to

$$\sup_{G \in \mathcal{C}_d} \mathbb{E}_G \left[|G \triangle \hat{G}| \right] \geq C_8 \eta^2 \exp\left(2n \ln(1 - C_9)\eta^{d+1}\right).$$

Let us choose $\eta = n^{-1/(d+1)}$, so that (20) becomes

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

where $C_{10} = C_8(1 - C_9)^2 > 0$. □

Proof of Theorem 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 $\mathcal{P}_{r^*} \subseteq \mathcal{P}_r \subseteq \mathcal{P}_{r'}$. Therefore if $P \in \mathcal{P}_{r^*}$ and $G = P$ in model (1), by Theorem 1 it is likely that with high probability we have, using the triangle inequality,

$$|\hat{P}_n^{(r)} \triangle \hat{P}_n^{(r')}| \leq \frac{Cdr' \ln n}{n}, \tag{21}$$

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 (21).

Let $\hat{r}$ be chosen as in Theorem 5. For $r = d + 1, \ldots, R_n$, let us denote by A_r the event following event.

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

where C_2 is the same constant as in Theorem 1. Then $\hat{r}$ is the smallest integer $r \leq R_n$ such that A_r holds.

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

$$\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|] = \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} \leq r^*)] + \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} > r^*)], \tag{22}$$

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,

$$\begin{aligned}
&\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} \leq r^*)] \\
&\quad \leq \mathbb{E}_P[|\hat{P}_n^{adapt} \triangle \hat{Q}_n^{(r^*)}| I(\hat{r} \leq r^*)] + \mathbb{E}_P[|\hat{Q}_n^{(r^*)} \triangle P| I(\hat{r} \leq r^*)] \\
&\quad \leq \frac{2C_a r^* \ln n}{n} + \frac{A_1 d r^* \ln n}{n} \text{ by Corollary 1, since } \hat{Q}_n^{(r^*)} = \hat{P}_n^{(r^*)} \\
&\quad \leq \frac{C_{11} r^* \ln n}{n},
\end{aligned} \tag{23}$$

where C_{11} depends only on d and σ. The second term of (22) is bounded differently. First note that for all $r = d+1, \ldots, R_n$, $\hat{Q}_n^{(r)} \subseteq [0,1]^d$, 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.

$$
\begin{aligned}
&\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} > r^*)] \\
&\quad \leq 2\mathbb{P}_P[\hat{r} > r^*] \\
&\quad \leq 2\mathbb{P}_P\left[\overline{A_{r^*}}\right] \\
&\quad \leq 2 \sum_{r=r^*}^{R_n} \mathbb{P}_P\left[|\hat{Q}_n^{(r^*)} \triangle \hat{Q}_n^{(r)}| > \frac{2C_a r \ln n}{n}\right] \\
&\quad \leq 2 \sum_{r=r^*}^{R_n} \mathbb{P}_P\left[|\hat{Q}_n^{(r^*)} \triangle P| + |\hat{Q}_n^{(r)} \triangle P| > \frac{2C_a r \ln n}{n}\right] \\
&\quad \leq 2 \sum_{r=r^*}^{R_n} \left(\mathbb{P}_P\left[|\hat{Q}_n^{(r^*)} \triangle P| > \frac{C_a r \ln n}{n}\right] + \mathbb{P}_P\left[|\hat{Q}_n^{(r)} \triangle P| > \frac{C_a r \ln n}{n}\right]\right) \\
&\quad \leq 2 \sum_{r=r^*}^{R_n - 1} \left(\mathbb{P}_P\left[|\hat{P}_n^{(r^*)} \triangle P| > \frac{C_a r^* \ln n}{n}\right] + \mathbb{P}_P\left[|\hat{P}_n^{(r)} \triangle P| > \frac{C_a r \ln n}{n}\right]\right) \\
&\qquad\quad + 2\mathbb{P}_P\left[|\hat{P}_n^{(r^*)} \triangle P| > \frac{C_a r^* \ln n}{n}\right] + 2\mathbb{P}_P\left[|\tilde{G} \triangle P| > \frac{C_a R_n \ln n}{n}\right] \qquad (24)
\end{aligned}
$$

Note that since $P \in \mathcal{P}_{r^*}$, it is also true that $P \in \mathcal{P}_r, \forall r \geq r^*$. Therefore, if $r^* \leq r \leq R^* - 1$, we have, using Theorem 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 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 (24) that

$$
\begin{aligned}
\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P| I(\hat{r} > r^*)] &\leq 2 \sum_{r=r^*}^{R_n - 1} 2C_1 n^{-(d+1)} + 2C_1 n^{-(d+1)} + 2\tau_1 n^{-(d+1)} \\
&\leq 4 \max(C_1, \tau_1) R_n n^{-(d+1)}. \qquad (25)
\end{aligned}
$$

Finally, using (23) and (25),

$$
\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|] \leq \frac{C_{12} r^* \ln n}{n},
$$

where C_{12} is a positive constant that depends on d and σ. Let us now assume that r^* is a given integer larger than R_n, possibly infinite, and that $P \in \mathcal{P}_{r^*}$. As in Theorem 6, if $r^* = \infty$ we denote by $\mathcal{P}_\infty$ the class $\mathcal{C}_d$. 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,

$$\begin{aligned}\mathbb{E}_P[|\hat{P}_n^{adapt} \triangle P|] &\leq 2C_a n^{-\frac{2}{d+1}} + \mathbb{E}_P[|\hat{Q}_n^{(R_n)} \triangle P|] \\ &\leq 2C_a n^{\frac{-2}{d+1}} + A_1'' n^{-\frac{2}{d+1}},\end{aligned}$$

by Corollary 3, since $P \in \mathcal{P}_{r^*} \subseteq \mathcal{P}_\infty$ and $\hat{Q}_n^{(R_n)}$ is the estimator of Theorem 4. Theorem 6 is then proven. □

7. Appendix: Proof of the lemmas

Proof of Lemma 1 Let us first state the following lemma, which gives the Steiner formula in the case of convex polytopes. It can also be found in [3]. If $R \subseteq \mathbb{R}^d$ and $\lambda > 0$, we denote by R^λ the set of all $x \in \mathbb{R}^d$ such that the Euclidean distance between x and R is less or equal to λ ;

$$R^\lambda = \{x \in \mathbb{R}^d, \rho(x, R) \leq \lambda\} = R + \lambda B_d(0, 1).$$

Lemma 5. *For any convex polytope $R \subseteq \mathbb{R}^d$ the volume of R^λ is polynomial in λ, with degree d, that is there exists $(L_0(R), \ldots, L_d(R)) \in \mathbb{R}^{d+1}$*

$$|R^\lambda| = \sum_{k=0}^{d} L_k(R)\lambda^k, \ \forall \lambda \geq 0.$$

Besides, $L_0(R) = |R|$, $L_1(R)$ is the surface area of R and $L_d(R) = |B_d(0, 1)|$, independent of R, and all the $L_i(R), i = 0, \ldots, d$ are nonnegative.

Note that in this lemma, if R is included in $B_d(a, u)$ for some $a \in \mathbb{R}^d$ and $u > 0$, then for all positive λ,

$$R^\lambda \subseteq B_d(a, u)^\lambda = B_d(a, u + \lambda)$$

and if we denote by $\beta_d = |B_d(0, 1)|$,

$$|R^\lambda| = \sum_{k=0}^{d} L_k(R)\lambda^k \leq (u + \lambda)^d \beta_d. \tag{26}$$

Therefore, since all the $L_i(R)$ are nonnegative, one gets

$$L_i(R) \leq (u + 1)^d \beta_d, i = 1, \ldots, d \tag{27}$$

by taking $\lambda = 1$ in (26).

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^*, then one can 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 $P^* \in \mathcal{P}_r^{(n)}$. For any set $G \subseteq \mathbb{R}^d$ and $\epsilon > 0$ we denote by G^ϵ the set

$$G^\epsilon = G + \epsilon B_d(0,1) = \{x \in \mathbb{R}^d : \rho(x, G) \leq \epsilon\}.$$

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 $P \subseteq (P^*)^\epsilon$.

Since the two polytopes P and P^* are included in $B_d\big(a, \frac{\sqrt{d}}{2}\big)$, for $a = (1/2, \ldots, 1/2)$, one gets from (27) that

$$L_i(R) \leq \left(\frac{\sqrt{d}}{2} + 1\right)^d \beta_d \leq \left(\frac{3\sqrt{d}}{2}\right)^d, \ i = 0, \ldots, d$$

for $R = P$ or P^*.

We can now bound the Nikodym distance between P and P^*

$$\begin{aligned} |P \triangle P^*| &= |P \backslash P^*| + |P^* \backslash P| \leq |(P^*)^\epsilon \backslash P^*| + |P^\epsilon \backslash P| \\ &\leq 2\left(\frac{3\sqrt{d}}{2}\right)^d \beta_d \sum_{k=1}^d \left(\frac{\sqrt{d}}{n}\right)^k \leq \frac{2d^{d+1}(3/2)^d \beta_d}{n}. \end{aligned}$$

□

Proof of Lemma 3 First note that if $G \subseteq [0,1]^d$, then the density of the probability measure $\mathbb{P}_G$ with respect to the Lebesgue measure on $[0,1]^d \times \mathbb{R}$ is

$$p_G(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y - I(x \in G))^2}.$$

Therefore, by a simple algebra, if G_1 and G_2 are two subsets of $[0,1]^d$, then

$$\begin{aligned} &\int_{[0,1]^d \times \mathbb{R}} \sqrt{p_{G_1}(x,y) p_{G_2}(x,y)} dx dy \\ &= \int_{[0,1]^d} \exp\left(-\frac{I(x \in G_1 \triangle G_2)}{8\sigma^2}\right) dx \\ &= |G_1 \triangle G_2| e^{-\frac{1}{8\sigma^2}} + 1 - |G_1 \triangle G_2|, \end{aligned}$$

and Lemma 3 follows from [28], Section 2.4. □

Proof of Lemma 4 The fact that any η-packing family of (S,ρ) is finite is clear and comes from the fact that S is compact. Consider now a maximal η-packing family of (S,ρ), 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 [25]

$$V(\eta/2)\geq \beta_{d-1}\left(1-\frac{\eta^2}{4}\right)^{(d-3)/2}\eta^{d-1}.$$

Besides, since $\{y_1,\dots,y_{M_\eta}\}$ is an η-packing family of (S,ρ), 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 4 follows from the fact that $\eta^2/4\leq 1/4$ and Lemma 2.2 of [25] which states that

$$\frac{\sqrt{2\pi}}{\sqrt{d+2}}\leq\frac{\beta_d}{\beta_{d-1}}\leq\frac{\sqrt{2\pi}}{\sqrt{d}}. \tag{28}$$

The left inequality of Lemma 4 comes from the fact that any maximal η-packing family is an η-net. Indeed, consider a maximal η-packing family $\mathcal{Y}$, and assume it is not an η-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 η-net that contains $\mathcal{Y}$ strictly. This contradicts maximality of $\mathcal{Y}$. Therefore the family $\{y_1,\dots,y_{M_\eta}\}$ is an η-net of S, and the caps $B_d(y_j,\eta)\cap S, j=1,\dots,M_\eta$ cover the sphere S, so that

$$M_\eta V(\eta)\geq\frac{d\beta_d}{2^{d-1}}.$$

Using again Lemma 2.3 of [25], we bound $V(\eta)$ from above

$$V(\eta)\leq\beta_{d-1}\eta^{d-1},$$

and then the desired result follows again from (28). ☐

Acknowledgements

This work is part of my PhD thesis under supervision of Prof. Alexander Goldenshluger and Prof. Alexandre Tsybakov, whom I would like to thank for their useful comments and the interesting discussions about set estimation and related topics. I would also like to thank the referees for indicating the useful reference [4], which I did not find by myself before submitting this article, and the Caesarea Rothschild Institute at the University of Haifa, for hosting me when I began working on this article.

References

[1] Alon N., Spencer J. H. (2008) The probabilistic method, Wiley. MR2437651

[2] 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. MR2327291

[3] Berger M. (1978) Géométrie, Tome 3: Convexes et polytopes, polyèdres réguliers, aires et volumes, Ed. Cedic/Fernand Nathan.

[4] Bronshtein E. M. (1976) ϵ-entropy of convex sets and functions, Siberian Mathematical Journal 17, 393-398. MR0415155

[5] Chichignoud M. (2010) Performances statistiques d'estimateurs non-linéaires, Thèse de Doctorat, Université de Provence U.F.R. M.I.M., available at http://stat.ethz.ch/people/michaech/PDF_thesis.

[6] Cuevas A. (2009) Set estimation: Another bridge between statistics and geometry, Boletín de Estadística e Investigación Operativa, Vol. 25, No. 2, pp. 71-85. MR2750781

[7] 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. MR2654684

[8] Cuevas A., Rodríguez-Casal A. (2004). On boundary estimation. Adv. in Appl. Probab. 36, pp. 340-354. MR2058139

[9] Dudley R. M. (1974) Metric Entropy of Some Classes of Sets with Differentiable Boundaries, Journal of Approximation Theory 10, 227-236. MR0358168

[10] Efron B. (1965) The Convex Hull of a Random Set of Points, Biometrika, Vol. 52, No 3/4, pp. 331-343. MR0207004

[11] Gordon Y., Meyer M., Reisner, S. (1995) Constructing a Polytope to Approximate a Convex Body, Geometricae Dedicata 57: 217-222, 1995. MR1347327

[12] Guntuboyina A. (2011) Lower bounds for the minimax risk using f-divergences, and applications. IEEE Transactions on Information Theory, vol. 57, pages 2386-2399 MR2809097

[13] Ibragimov I. A., Khasminiskii R. Z. (1984) Statistical Estimation: Asymptotic Theory, New York: Springer-Verlag. MR0620321

[14] Janson S. (1987) Maximal spacings in several dimensions, The Annals of Probability, Vol. 15, No. 1, pp. 274-280. MR0877603

[15] KENDALL W. S., MOLCHANOV I. (2010) New Perspectives in Stochastic Geometry. Eds., Oxford University Press. MR2668353
[16] KENDALL M. G., MORAN P. A. P. (1963) Geometrical Probability, Griffin's Statistical Monographs and Courses, no.10. MR0174068
[17] KOLMOGOROV A. N., TIKHOMIROV V. M. (1959) ε-entropy and ε-capacity of sets in function spaces, Uspekhi Mat. Nauk, 14:2(86), 3–86 (in Russian). MR0112032
[18] KOROSTELEV A. P., TSYBAKOV A. B. (1992) Asymptotically minimax image reconstruction problems. In Topics in Nonparametric Estimation (R. Z. Khasminskii, ed.) 45-86. Amer. Math. Soc., Providence, RI. MR1191691
[19] KOROSTELEV A. P., TSYBAKOV A. B. (1993) Minimax Theory of Image Reconstruction. Lecture Notes in Statistics, v.82. Springer, NY e.a. MR1226450
[20] KOROSTELEV A. P., TSYBAKOV A. B. (1994) Asymptotic efficiency in estimation of a convex set (in russian), Problems of Information Transmission, v.30, n.4, 317-327. MR1310060
[21] LEPSKI O. V. (1991) Asymptotically minimax adaptive estimation i. upper bounds. optimally adaptive estimates. Theory Probab. Appl., 36:682-697. MR1147167
[22] MAMMEN E., TSYBAKOV A. (1995) Asymptotical Minimax Recovery of Sets with Smooth Boundaries, The Annals of Statistics, Vol. 23, No. 2, pp. 502-524. MR1332579
[23] MCCLURE D. E., VITALE R. A. (1975) Polygonal Approximation of Plane Convex Bodies, Journal of Mathematical Analysis and Applications, Vol. 51, No.2. MR0385714
[24] PATEIRO-LÓPEZ B. (2008) Set estimation under convexity type restrictions, PhD thesis, available at `http://eio.usc.es/pub/pateiro/files/THESIS_BeatrizPateiroLopez.pdf`.
[25] REISNER S., SCHÜTT C., WERNER E. (2001) Dropping a vertex or a facet from a convex polytope, Forum Math., 359-378. MR1831090
[26] RÉNYI A., SULANKE R. (1963) Über die konvexe Hülle von n zufällig gewählten Punkten. Z.Wahrscheinlichkeitsth. Verw. Geb. 2 pp. 75-84. MR0156262
[27] RÉNYI A., SULANKE R. (1964) Über die konvexe Hülle von n zufällig gewählten Punkten. II, Z.Wahrscheinlichkeitsth. Verw. Geb. 3 pp. 138-147. MR0169139
[28] TSYBAKOV A. B. (2009) Introduction to nonparametric estimation, Springer. MR2724359