Estimation of the support of the density and its boundary using Random Polyhedron

Catherine Aaron

Submitted to the Annals of Applied Probability

ESTIMATION OF THE SUPPORT OF THE DENSITY AND ITS BOUNDARY USING RANDOM POLYHEDRON

By Catherine Aaron[??],

Universit Blaise Pascal - Laboratoire de Mathmatiques UMR 6620 - CNRS

We consider random samples in $\mathbb{R}^d$ drawn from an unknown density. When the support is assumed to be convex and with sharp boundary, the convex hull is an estimator of the support that converges to S with a rate of $n^{-2/(d+1)}$. When the boundary of the support is sharp but the support is no longer assumed to be convex, the usual support estimators converges with a rate of $n^{-1/d}$ or $(\ln(n)/n)^{-1/d}$. This paper is devoted to presenting some new estimators of the support of the density, which are based on some local convexity criteria and converge to S with a rate of $(n/\ln n)^{-2/(d+1)}$ (and their boundary converges toward ∂S with the same rate) when the support is assumed to have a sharp $\mathcal{C}^2$ boundary. The convergence rate is also given when the sharpness hypothesis is relaxed (and it is close to the optimal rate when the dimension is two).

1. introduction. Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a random sample in $\mathbb{R}^d$ drawn from an unknown density f. The density support S is defined by $S = \overline{\{x \in \mathbb{R}^d, f(x) > 0\}}$, where $\overline{A}$ denotes the closure of the set A. Estimation of S has various applications in statistics such as classification and clustering (see the discussion in [2]) and has been widely studied. First let us cite the Devroye-Wise estimator [16] which is the union of small balls centered in the observations. Such an estimator has been extensively studied and converges with a rate of $(\ln n/n)^{1/d}$ [2] (and provides a boundary estimator that converges with the same rate [15]) when boundary is sharp. There are other support estimators as in [14] where the support is estimated as a level set of a kernel-based density estimation that converges, with regard to the Hausdorff distance, with rate $n^{-1/d}$. Another idea is to work with an union of bins as in [21] and once again the convergence rate is $(\ln n/n)^{1/d}$. Some estimators have also been studied without the sharpness hypothesis ([5] [4] focuses on the convergence rate of the Devroye-Wise estimator, [19] presents an optimal estimator but is limited to the case $d = 2$), and, of course that

*Footnote to the title with the 'thankstext' command.

AMS 2000 subject classifications: Primary 62-07, 62-07 ; secondary ; 62G05; 62G07; 62H12

Keywords and phrases: Delaunay complex, polyhedron, support estimation, topological data analysis, geometric inference., LaTeX 2_ε

provides slower convergence rates.

Each time (when the boundary is sharp) the convergence rate is far from the convergence rate obtained when the support is convex (and the boundary sharp) and is estimated by the convex hull of the observations (denoted in this paper $\mathcal{H}(\mathcal{X}_n)$) Some asymptotic properties of this estimator can be found for instance in [18], [3], [23] or [24] and the convergence rate is $n^{-2/(d+1)}$ (proved for L^1 and L^2 convergence).

The convexity hypothesis is a strong assumption, but it can appear natural to try to built some density support estimators using tools strongly linked with the convex hull but having a local "point of view". A first idea is to recall that the convex hull can also be defined as the union of the Delaunay simplices (see definition 1). In [17] it can be seen that a restriction of the Delaunay polyhedron allows one to recover the topology of a manifold. This work is not totally applicable for our purpose as, firstly the restriction requires knowledge of the support, and, secondly, there are no specific results for the support estimation. Nevertheless it suggests the idea that the support can be estimated using a restriction of the Delaunay polyhedron, i.e. by removing some simplices. In [1] it has be seen that a restriction according to a nearest neighbors criteria gives better results for the density support estimation than the proved results (ones again with rate $(\ln n/n)^{1/d}$). Here we propose another criteria to restrict the Delaunay polyhedron in a way that gives a support estimator.

DEFINITION 1. *(The $r-$Delaunay polyhedron D_r.) A $k-$dimensional simplex σ is the convex hull of $k+1$ points. For ease of writing we shall denote it by $\sigma = (x_1, \ldots, x_{d+1})$ (instead of $\mathcal{H}(\{x_1, \ldots, x_{d+1}\})$). Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a subset of $\mathbb{R}^d$. A simplex $\sigma = (x_1, \ldots, x_{d+1})$ belongs to $\mathcal{D}_r(\mathcal{X}_n)$ (the associated complex) if:*

- *It belongs to the Delaunay complex, i.e. all the vertices are in $\mathcal{X}_n$ and, if we denote by O_σ and r_σ the center and the radius of the hypersphere circumscribed to $\sigma = (x_1, \ldots, x_{d+1})$, $\mathcal{B}(O_\sigma, r_\sigma) \cap \mathcal{X}_n = \emptyset$;*
- *The simplex "is small" i.e. $r_\sigma \leq r$.*

We can now define the support estimator as : $D_r(\mathcal{X}_n) = \bigcup_{\sigma \in \mathcal{D}_r(\mathcal{X}_n)} \sigma$.

Another natural estimator that can be proposed to deal with non-convex support estimation is the $r-$union of convex hulls. The idea here is that a smooth enough set is locally convex and can be approximated by a union of small convex sets.

DEFINITION 2. *(The $r-$union of convex hulls H_r.) Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$*

be a subset of $\mathbb{R}^d$. *The* $r-$*union of convex hulls of this points, denoted* $H_r(\mathcal{X}_n)$ *is defined as follows:*

$$H_r(\mathcal{X}_n) = \bigcup_i \mathcal{H}(\overline{\mathcal{B}}(X_i, r) \cap \mathcal{X}_n)$$

The major problem with such an estimator is its computational cost because it may need the estimation of many convex hulls. The next proposed estimator is close to the to the $r-$union of convex hulls and does not require any convex hull computation. The underlying idea is that a convex hull is also a union of simplices.

DEFINITION 3. *(The* $r-$*Rips Polyhedron* R_r*.) Let* $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ *be a subset of* $\mathbb{R}^d$. *Let us first define the* $r-$*Rips complex as follows: A simplex* σ *belongs to* $\mathcal{R}_r$, *the* $r-$*Rips complex, if the length of each of its edges is smaller than* r *and if all its vertexes are in* $\mathcal{X}_n$.

Now, the $r-$*Rips Polyhedron denoted* $R_r(\mathcal{X}_n)$ *is defined as follows.*

$$R_r(\mathcal{X}_n) = \bigcup_{\sigma \in \mathcal{R}_r(\mathcal{X}_n)} \sigma.$$

Let us note that the $\varepsilon-$Rips complex is frequently used for geometric inference ([7], [25], [9], [8], [6], [13]). The computational cost is better than the restricted Delaunay polyhedron or the union of convex hulls and it can be computed on sets given via pairwise distances.

This paper is devoted to proving some asymptotic properties for the proposed support estimator according to the Hausdorff distance d_H. Namely, if we suppose (as in [19]) that the density decreased to 0 as a power $\alpha \geq 0$ of the distance to the boundary (sharpness hypothesis is $\alpha = 0$) then, for some fixed radius r we have $d_H(D_r(\mathcal{X}_n), S) \left(\frac{\ln n}{n}\right)^{2/(d+1+2\alpha)}$ is e.a.s. bounded, and when $r_n = (\ln n/n)^{\frac{2}{d+1+2\alpha}}$ we also have $d_H(H_{r_n}(\mathcal{X}_n), S) \left(\frac{\ln n}{n}\right)^{2/(d+1+2\alpha)}$ and $d_H(R_{r_n}(\mathcal{X}_n), S) \left(\frac{\ln n}{n}\right)^{2/(d+1+2\alpha)}$ being e.a.s. bounded. The estimators boundaries converge to the support boundary with the same rate.

Section 2 is dedicated to the notations and hypotheses used in the article. Section 3 presents the associated theorem and gives a brief discussion on the fact that the proposed estimators are expected to be homeomorphic to the support. Section 4 is devoted to the proofs of the theorems.

2. Notations, definitions and hypotheses..

2.1. *Notations and definitions.* Throughout the paper the following notations are used:

The density support S is assumed to be a compact $d-$dimensional manifold. Let us recall a definition.

DEFINITION 4. *A set A is a is a $d-$dimensional manifold if for all $x \in A$, x admits a neighborhood homeomorphic to B_d or B_d^+ with:*

- $B_d = \{(x_1, \ldots, x_d) \in \mathbb{R}^d, \sum x_i^2 < 1\}$
- $B_d^+ = \{(x_1, \ldots, x_d) \in \mathbb{R}^d, \sum x_i^2 < 1, x_1 \geq 0\}$

The fact that $S \subset \mathbb{R}^d$ is a $d-$dimensional manifold implies that it has a boundary (denoted ∂S) that is assumed to be $\mathcal{C}^2$ throughout the paper. Together with the compactness hypothesis this allows us to define γ_S the maximum value (for $x \in \partial S$) of the maximum (for the d directions) principal curvature of ∂S, and $r_S = \gamma_S^{-1}$ the minimum radius of curvature.

$\mathcal{B}(x, r)$ (resp. $\overline{\mathcal{B}}(x, r)$) is the open (resp. closed) $d-$dimensional ball of radius r centered at x. When no radius or center are specified, the open (resp. closed) ball are the unit open (resp. closed) ball of $\mathbb{R}^d$.

Throughout the paper θ_d denotes the volume of the unit $d-$dimensional ball. Some other volume constants are needed in the paper:

$$\theta_{d,\alpha} = \theta_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) \text{ with } B \text{ the Beta function,}$$

$$\theta_{d,\alpha}^{ellipse} = \theta_{d-1} \int_{-1}^{1} (1+t)^{\frac{d-1+2\alpha}{2}} (1-t)^{\frac{d-1}{2}} dt = \theta_{d-1} 2^{\alpha+1} B\left(\frac{d+1+2\alpha}{2}, \frac{d+1}{2}\right),$$

$$\theta_{d,\alpha}^{lens} = 2^{\frac{d-1}{2}} \frac{\theta_{d-1}}{d+\alpha}.$$

For a set A and a constant $\varepsilon \geq 0$, the set $A + \varepsilon\overline{\mathcal{B}}$ denotes the sum of A and $\varepsilon\overline{\mathcal{B}}$ and is defined by $A + \varepsilon\overline{\mathcal{B}} = \bigcup_{a \in A} \overline{\mathcal{B}}(a, r_n)$

The distance used to characterize the difference between a support and its estimator is the Hausdorff distance defined as follows.

DEFINITION 5 (Hausdorff distance). *Let A and B be two subsets of $\mathbb{R}^d$. We denote by $A + \varepsilon\overline{\mathcal{B}}$ the set $\bigcup_{a \in A} \overline{\mathcal{B}}(a, \varepsilon)$*

$$d_H(A, B) = \inf\{r > 0, A \subset B + \varepsilon\overline{\mathcal{B}} \text{ and } B \subset A + \varepsilon\overline{\mathcal{B}}\}.$$

The distance between a point x and a set A, denoted $d(x, A)$ is the usual one, i.e. $d(x, A) = \inf_{a \in A} ||x - A||$.

To a $d-$simplex σ, one can associate r_σ and O_σ, the associated radius and center which are those of the hypersphere circumscribed to the vertices of σ.

2.2. *Hypotheses.* The theorem and properties exposed in the following requires strong hypotheses.

The first one is that the support is $d-$dimensional (with d the dimension of the observation space). It is a strong hypothesis for potential applications as it excludes dealing with sparsity phenomena where the dimension of the support is supposed to be smaller than the dimension of the observation space. However we strongly believe that it is, in fact, a "comfort hypothesis" and that an adaptation to a $d'-$dimensional smooth manifold is possible.

The second hypothesis is that the boundary of the support is a smooth $\mathcal{C}^2$ manifold. This has the great advantage of allowing us to introduce γ_S, the maximum value for the principal curvature of ∂S, and the radius $r_S = \gamma_S^{-1}$. Even if this is a strong assumption, it can be considered as not so restrictive in regard to the obtained convergence rate.

Lastly, the support has to be compact and the density f has to satisfy the following condition: there exists an $f_0 > 0$ and an $\alpha > 0$ such that $f(x) \geq f_0 d(x, \partial S)^\alpha$. When $\alpha = 0$ the boundary is sharp. It is an usual hypothesis that allows us to deal with non-sharp boundaries (as in [19]).

These three hypotheses are made throughout the paper. They are recalled in all the theorems of Section 3, which are the main ones in order to give self-contents statements. However, we will not recall them in Section 4 where the proofs are given.

3. Properties of the estimators.

3.1. *About the restricted Delaunay polyhedron.* Let us begin by considering $D_{r_n}(\mathcal{X}_n)$ under the sharpness hypothesis. First let us first imagine that $r_n(\ln(n)/n)^{-1/d} \to 0$. In this case, according to Penrose ([22]) we have $D_{r_n}(\mathcal{X}_n) \to \emptyset$ (because for every simplex of the Delaunay complex, the associated radius r_σ is larger than the minimum length of its edges and so larger than the minimum distance between two points). Let us now imagine that $r_n \geq r_S + \varepsilon$ and focus on the following example : $S = \mathcal{B}(0,1) \setminus \mathcal{B}(0, r_S)$. One can easily imagine that keeping the Delaunay simplices of radius that are close to r_S gives $P(0 \in D_{r_S+\varepsilon})(\mathcal{X}_n) \nrightarrow 0$.

This illustrates that, in order to use D_{r_n} as a S estimator, the suitable v r_n values has to be smaller than r_S, but may not decrease toward 0 too quickly. Theorem 1 gives the non-intuitive result that a constant sequence is suitable.

THEOREM 1. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a sample of n independent and identically distributed $\mathbb{R}^d-$valued random variables drawn with an unknown density f. Suppose that the support S of the density is a $d-$dimensional*

manifold with a $\mathcal{C}^2$ boundary ∂S. Suppose that there exists $f_0 > 0$ and $\alpha \geq 0$ such that $f(x) \geq f_0 d(x, \partial S)^\alpha$. Then for a fixed radius $r < r_S$, we have:

$$d_H(D_r(\mathcal{X}_n), S) \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}} \text{ is e.a.s. bounded}$$

and

$$d_H(\partial D_r(\mathcal{X}_n), \partial S) \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}} \text{ is e.a.s. bounded.}$$

The explicit expression for the constant is long but is obtained by choosing the maximum of the constants form Lemmas 2, 3 and 4. Obviously small values for r give large values for $\max_{x\in S} d(x, D_r)$ and large values for r increase $\max_{x\in R_r} d(x, S)$.

When $\alpha = 0$, a convergence rate of $(\ln(n)/n)^{\frac{2}{d+1}}$ is obtained as announced in the introduction. It is very close to the convergence rate obtained for a convex set when using $\mathcal{H}(\mathcal{X}_n)$ to estimate the support.

When $\alpha \geq 0$ and $d = 2$ the convergence rate, proved to be optimal in [19], is $n^{-\frac{2}{3+2\alpha}}$. Here again, the proposed estimators converge with the same power but applied to $(n/\ln n)$ instead of n.

3.2. *About the other proposed estimators.* Since the construction of all the three proposed estimators was based on the same idea of using a convex hull characterization with the introduction of local restriction , it is natural to look for a link between these estimators. This is given by the following property.

PROPERTY 1. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a sample of n independent and identically distributed $\mathbb{R}^d-$valued random variables. Suppose that the support S of the density is a $d-$dimensional manifold with a $\mathcal{C}^2$ boundary ∂S. Suppose that there exists $f_0 > 0$ and $\alpha \geq 0$ such that $f(x) \geq f_0 d(x, \partial S)^\alpha$. Then, for any radius sequence $r_n = \lambda \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}}$ there exists a fixed radius r_0, with $0 < r_0 < r_S$ such that:*

$$D_{r_0} \subset R_{r_n} \subset H_{r_n} \text{ e.a.s.}$$

The first inclusion is a direct corollary of Lemma 4 and holds e.a.s. The second inclusion is obvious and deterministic (we have $R_r \subset H_r$ for all values of r).

Theorems 2, 3, 4 and 5 are corollaries of Proper 1, Theorem 1 and Lemma 6.

Similar results can be easily established for every support estimator that is a polyhedron that contains D_r and such that all the edges are (e.a.s.) smaller than a $O((\ln(n)/n)^{1/(d+1)}$. The following theorems give the consistency results of R_{r_n} and H_{r_n}.

THEOREM 2. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a sample of n independent and identically distributed $\mathbb{R}^d-$valued random variables. Suppose that the support S of the density is a $d-$dimensional manifold with a $\mathcal{C}^2$ boundary ∂S. Suppose that there exists $f_0 > 0$ and $\alpha \geq 0$ such that $f(x) \geq f_0 d(x, \partial S)^\alpha$. For all sequences $r_n = \lambda \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}}$ we have*

$$d_H(R_{r_n}(\mathcal{X}_n), S) \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}} \text{ is e.a.s. bounded}$$

and

$$d_H(\partial R_{r_n}(\mathcal{X}_n), \partial S) \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}} \text{ is e.a.s. bounded.}$$

THEOREM 3. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a sample of n independent and identically distributed $\mathbb{R}^d-$valued random variables. Suppose that the support S of the density is a $d-$dimensional manifold with a $\mathcal{C}^2$ boundary ∂S. Suppose that there exists $f_0 > 0$ and a $\alpha \geq 0$ such that $f(x) \geq f_0 d(x, \partial S)^\alpha$. If $r_n \left(\frac{n}{\ln n}\right)^{\frac{1}{d+1+2\alpha}} \to +\infty$ then*

$$d_H\left(R_{r_n}(\mathcal{X}_n), S\right) r_n^{-2} \text{ is bounded e.a.s.}$$

and

$$d_H\left(\partial R_{r_n}(\mathcal{X}_n), \partial S\right) r_n^{-2} \text{ is bounded e.a.s.}$$

THEOREM 4. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a sample of n independent and identically distributed $\mathbb{R}^d-$valued random variables. Suppose that the support S of the density is a $d-$dimensional manifold with a $\mathcal{C}^2$ boundary ∂S. Suppose that there exists $f_0 > 0$ and $\alpha \geq 0$ such that $f(x) \geq f_0 d(x, \partial S)^\alpha$. For all sequences $r_n = \lambda \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}}$ we have*

$$d_H(H_{r_n}(\mathcal{X}_n), S) \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}} \text{ is e.a.s. bounded}$$

and

$$d_H(\partial H_{r_n}(\mathcal{X}_n), \partial S) \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}} \text{ is e.a.s. bounded}$$

THEOREM 5. *Let $\mathcal{X}_n = \{X_1, \dots, X_n\}$ be a sample of n independent and identically distributed $\mathbb{R}^d-$valued random variables. Suppose that the support S of the density is a $d-$dimensional manifold with a $\mathcal{C}^2$ boundary ∂S. Suppose that there exists $f_0 > 0$ and $\alpha \geq 0$ such that $f(x) \geq f_0 d(x, \partial S)^\alpha$. If $r_n \left(\frac{n}{\ln n}\right)^{\frac{1}{d+1+2\alpha}} \to +\infty$ then*

$$d_H\left(H_{r_n}(\mathcal{X}_n), S\right) r_n^{-2} \text{ is bounded e.a.s.}$$

and

$$d_H\left(\partial H_{r_n}(\mathcal{X}_n), \partial S\right) r_n^{-2} \text{ is bounded e.a.s.}$$

3.3. *A remark about topology preservation.* There exist new statistical methods that use estimation of topological invariants for real life data (see [8], [6] or [13]). This is a motivation to find support estimators that preserve the topology of the support (see [11], [10], [12]). We observed that our proposed estimators are homeomorphic to the support but we are not yet able to prove it and that is part of further work. However we can already notice two encouraging facts. First let us recall that, in [**?**] a Delaunay complex restriction is presented, let us denote it $D_{(S)}$, that is (under a reasonable hypothesis) homeomorphic to S. Our estimator D_r satisfies $D_{(S)}(\mathcal{X}_n) \subset D_r(\mathcal{X}_n)$ (e.a.s. it will be clear from the proof of Theorems 1) and so $D_{(S)}$ is included in every proposed estimator (because of Property 1).

Secondly we can notice that the $n^{1/(d+1)}$ rate used for the radius sequence make the use of the r_n-Rips complex $\mathcal{R}_{r_n}$, consistent with the estimation of the Betti numbers ([20]).

4. Proofs. In this section we prove Theorem 1. Firstly, in Section 4.1 we give the main idea of the proof and then we detail some useful lemmas and properties in Section 4.2. Proof for Property 1, and Theorems 2, 3, 4 and 5 are left to the reader because all the useful properties are established within the proof of Theorem 1, namely:

- Property 1 is a corollary of Lemma 4.
- Theorems 2, 3, 4 and 5 are corollaries of Property 1 and Lemma 6.

4.1. *Proof of Theorem 1.* The proof of Theorem 1 is divided into two parts. In the first one we consider points that belong to S but not to D_r. In the second part we consider points that belong to D_r and not to S. For both parts the aim is to prove that points are (e.a.s.) close to ∂S.

4.1.1. *First part.* Let us pick a x that belongs to S but not to D_r. There are two possible cases: x does not belongs $\mathcal{H}(\mathcal{X}_n)$ or it belongs to $\mathcal{H}(\mathcal{X}_n)$ but is in a "big" simplex.

LEMMA 1.

$$\sup_{x\in S, x\notin\mathcal{H}(\mathcal{X}_n)} d(x,\partial S) \leq 2\left(\frac{(2d+1+2\alpha)\ln(n)}{(d+1+2\alpha)\theta_{d,\alpha}^{ellipse} f_0 r_S^{\frac{d-1}{2}} n}\right)^{\frac{2}{d+1+2\alpha}} \quad e.a.s.$$

LEMMA 2. *For any fixed radius $r < r_S$,*

$$\sup_{x\in S, x\in\mathcal{H}(\mathcal{X}_n), x\notin D_r} d(x,\partial S) \leq \frac{r_S+r}{r_S}\left(\frac{\ln n}{f_0\theta_{d,\alpha}^{lens} r^{\frac{d-1}{2}} n}\right)^{\frac{2}{d+1+2\alpha}} \quad e.a.s.$$

Proof of Lemma 1. Let us denote:

$$\sup_{x\in S, x\notin\mathcal{H}(\mathcal{X}_n)} d(x,\partial S) = d_{\sup}.$$

First, one can easily prove that there exists a point $y^* \in \partial\mathcal{H}(\mathcal{X}_n)$ and a point $x^* \in \partial S$ such that $||x^* - y^*|| = d_{\sup}$ (existence because of S compactness and localization on the boundaries because of the maximum). Let $\overrightarrow{u}_{x^*}$ denote the unit vector, normal to ∂S at the point x^* which points outward from S. Then the open half space $H_{y^*}^{\overrightarrow{u}_{x^*}} = \{x \in \mathbb{R}^d, \overrightarrow{y^*x}\,\overrightarrow{u}_{x^*} > 0\}$ does not intersect $\mathcal{X}_n$ ($y^* \in \partial\mathcal{H}(\mathcal{X}_n)$ implies that there exists an open half space that contains y^* and that does not intersect $\mathcal{X}_n$ and, the fact that the half space is not the named one contradicts the maximality of the distance between the boundary of S and the boundary of $\mathcal{H}(\mathcal{X}_n)$). Let us denote by z the midpoint of the two points x^* and y^*. We have $d(z,\partial S) = a = d_{\sup}/2$. All the previous considerations imply that every ellipsoid $\mathcal{E}_{z,a}$ does not intersect $\mathcal{X}_n$. See Figure 1 for the construction. Application of Lemma 12 proves that $a = 0.5 d_{sup} \leq \left(\frac{(2d+1+2\alpha)\ln(n)}{(d+1+2\alpha)\theta_{d,\alpha}^{ellipse} f_0 r_S^{\frac{d-1}{2}} n}\right)^{\frac{2}{d+1+2\alpha}}$ e.a.s.

Proof of Lemma 2. Let us denote:

$$\max_{x\in S, x\in\mathcal{H}(\mathcal{X}_n), x\notin \mathring{D}_r} d(x,\partial S) = d_{\max}$$

This max is realized for a given point x, with $x \in S$ and $x \in \mathcal{H}(\mathcal{X}_n)$ but $x \notin \mathring{D}_r$. As x belongs to $\mathcal{H}(\mathcal{X}_n)$ there exists a simplex σ of the Delaunay complex such that x belongs to σ. This simplex does not belong to D_r so $r_\sigma > r$. As r is constant, $O_\sigma \notin S$ e.a.s. (otherwise there exists a ball centered in S and of radius r that does not contains any observation, which

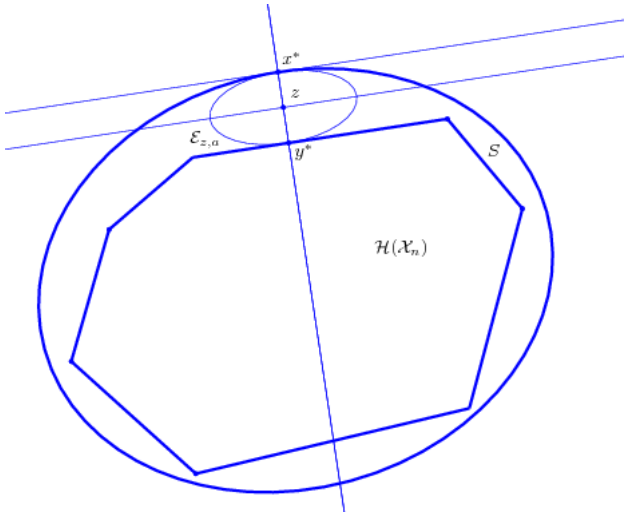

FIG 1. *Construction of an empty ellipsoid for Lemma 1*

is impossible e.a.s. according to Lemma 10). As $x \in \sigma \subset \overline{\mathcal{B}}(O_\sigma, r_\sigma)$ we have $d_{\max} \leq \max_{y \in \overline{\mathcal{B}}(O_\sigma, r_\sigma)} d(y, \partial S) = d^*$.

As in the previous proof we can establish the existence of $y^* \in \mathcal{S}(O_\sigma, r_\sigma) \cap S$ and $x^* \in \partial S$ such that $||x^* - y^*|| = d^*$. Moreover the fact that we are at the maximum of the distance ensures that the plane tangent to S at the point x^* is parallel to the plan tangent to $\mathcal{S}(O_\sigma, r_\sigma) \cap S$ at the point y^*. So the vector $\overrightarrow{x^* y^*}$ normal to S at the point x^* is parallel to the vector $\overrightarrow{O_\sigma y^*}$ which is normal to $\mathcal{S}(O_\sigma, r_\sigma) \cap S$ at the point y^*. This implies that y^*, x^* and O_σ are collinear. Let us now define O_r such that $\overrightarrow{y^* O_r} = (r_n / r_\sigma) \overrightarrow{y^* O_\sigma}$. We obviously have: $\mathcal{B}(O_r, r_n) \subset \mathcal{B}(O_\sigma, r_\sigma)$ so $\mathcal{B}(O_r, r_n) \cap \mathcal{X}_n = \emptyset$ and $d(O_r, \partial S) \leq r_n - d^*$. See Figure 2 for construction.

According to Lemma 11, we have

$$d_{\max} \leq d^* \leq \frac{r_S + r}{r_S} \left(\frac{\ln n}{f_0 \theta_{d,\alpha}^{lens} r^{\frac{d-1}{2}} n} \right)^{\frac{2}{d+1+2\alpha}} \quad \text{e.a.s.}$$

4.1.2. *Second part.* Let us pick a x that belongs to D_r but not to S. Here x belongs to a (unique) simplex σ and we will study the two possible cases : $O_\sigma \in S$ and $O_\sigma \notin S$.

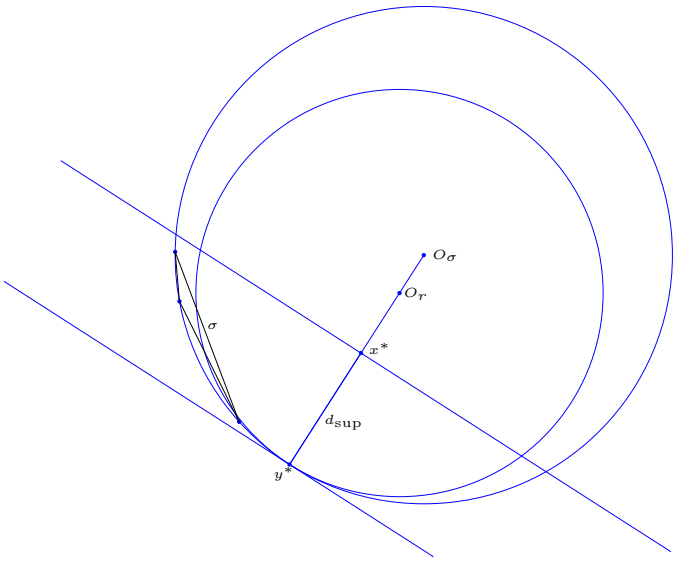

FIG 2. *Construction of an empty $\mathcal{C}$ for Lemma 2*

LEMMA 3.

$$\max_{x\in\sigma\in\mathcal{D}_r, O_\sigma\in S} d(x,\partial S) \leq \frac{1}{2r_S}\left(\frac{(2d+\alpha)}{(d+\alpha)\theta_{d,\alpha}f_0}\frac{\ln(n)}{n}\right)^{\frac{2}{d+\alpha}} \quad e.a.s.$$

LEMMA 4.

$$\max_{x\in\sigma\in\mathcal{D}_r, O_\sigma\notin S} d(x,\partial S) \leq a_2\left(\ln(n)/n\right)^{2/d} \quad e.a.s.$$

Proof of Lemma 3. For all σ such that belongs to $\mathcal{D}_r$ and such that $O_\sigma \in S$ we have $r_\sigma \leq \left(\frac{(2d+\alpha)}{(d+\alpha)\theta_{d,\alpha}f_0}\frac{\ln(n)}{n}\right)^{\frac{1}{d+\alpha}}$ (this is a corollary to Lemma 10) and so its edges are smaller than $2\left(\frac{(2d+\alpha)}{(d+\alpha)\theta_{d,\alpha}f_0}\frac{\ln(n)}{n}\right)^{\frac{1}{d+\alpha}}$. We conclude, according to Lemma 6, that for all $x \in D_r$ such that $O_\sigma \in S$ we have $d(x,\partial S) \leq \frac{1}{2r_S}\left(\frac{(2d+\alpha)}{(d+\alpha)\theta_{d,\alpha}f_0}\frac{\ln(n)}{n}\right)^{\frac{2}{d+\alpha}}$ e.a.s.

Proof of Lemma 4. Let σ be simplex that belongs to D_r and such that $O_\sigma \notin S$. If $r_\sigma \leq r$ we can pick a point $x^* \in \sigma$ that realizes the minimum of the distances $\max_{x\in\partial S}||x-O_\sigma||$. Let us denote $\overrightarrow{u}_{x^*}$ the unit vector, tangent to ∂S at the point x^* and that points outward from S. Let us denote O_S^+ (resp. O_S^-) the point that satisfy $\overrightarrow{x^*O_S^+} = -r_S\overrightarrow{u}_{x^*}$ (resp. $\overrightarrow{x^*O_S^+} = +r_S\overrightarrow{u}_{x^*}$).

Let us now define the point O_r such that $\mathcal{S}(O_r, r) \cap \mathcal{S}(O_S^-, r_S) = \mathcal{S}(O_\sigma, r_\sigma) \cap \mathcal{S}(O_S^-, r_S)$. Figure 3 shows that $\mathcal{B}(O_r, r) \cap S \subset \mathcal{B}(O_\sigma, \sigma) \cap S$. This first step allows us to fix the radius as r and avoid working with a radius r_σ varying between 0 and r.

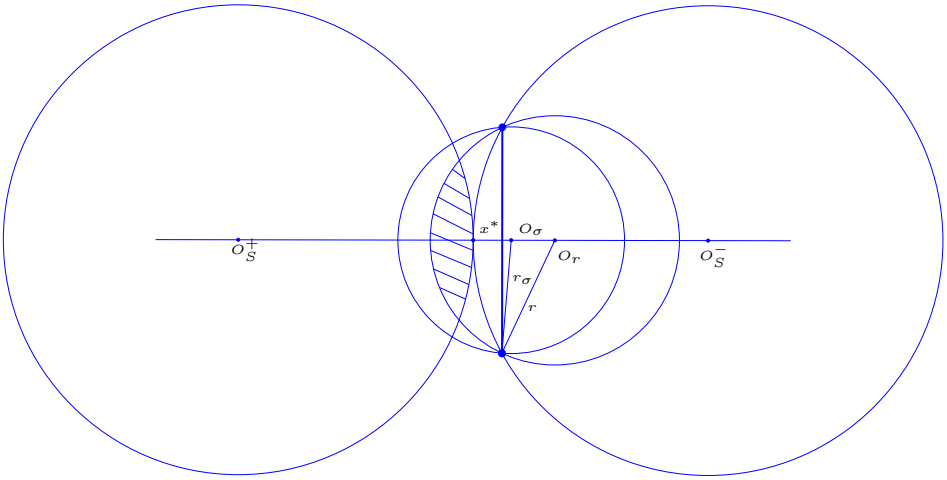

FIG 3. *First construction for Lemma 4*

Now, as $\mathcal{B}(O_r, r) \cap \mathcal{X}_n = \emptyset$, Lemma 11 implies that $y' + l' \leq a(\ln(n)/n)^{2/(d+1)}$ e.a.s. (see Figure 4 for a depiction of the definitions for l, l', y' and h, and choose the value for a according to Lemma 11). We have the following system of equation that links l, l', y' and h, when we fix $l' + y' = \varepsilon_n$:

$$\begin{cases} (r_S - l)^2 + h^2 = r_S^2 \\ (r - l - l' - y')^2 + h^2 = r^2 \\ (r_S - l')^2 + h'^2 = r_S^2 \\ (r - y')^2 + h'^2 = r^2 \end{cases} \Rightarrow \begin{cases} h^2 = \frac{2rr_S}{r_S - r}\varepsilon_n + o(\varepsilon_n) \\ l = \frac{r}{r_S - r}\varepsilon_n + o(\varepsilon_n) \\ h'^2 = \frac{2rr_S}{r_S - r}\varepsilon_n + o(\varepsilon_n) \\ y' = \frac{r_S}{r + r_S}\varepsilon_n + o(\varepsilon_n) \end{cases}$$

The maximum edge length of σ, z_σ, is bounded above by $2\sqrt{h^2 + l^2}$ and so $z_\sigma \leq 2\sqrt{\frac{2rr_S}{r_S - r}\varepsilon_n + o(\varepsilon_n)}$ (e.a.s.). Now we can apply Lemma 6 to obtain that $d(\sigma, S) \leq \frac{1}{8r_S} 4 \frac{2rr_S}{r_S - r}\varepsilon_n + o(\varepsilon_n) = \frac{r}{r_S - r}\varepsilon_n + o(\varepsilon_n)$. To conclude, we apply Lemma 11 which proves that $\varepsilon_n \leq \frac{r_S + r}{(f_0 \theta_{d,\alpha}^{lens})^{\frac{2}{d+1+2\alpha}} r^{\frac{d-1}{d+1+2\alpha}} r_S} \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}}$ e.a.s. That concludes the proof as we now have:

$$d(\sigma, S) \leq \frac{(r_S + r) r^{\frac{2\alpha}{d+1+2\alpha}}}{(r_S - r) r_S (f_0 \theta_{d,\alpha}^{lens})^{\frac{d-1}{d+1+2\alpha}}} \left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}} \text{ e.a.s.}$$

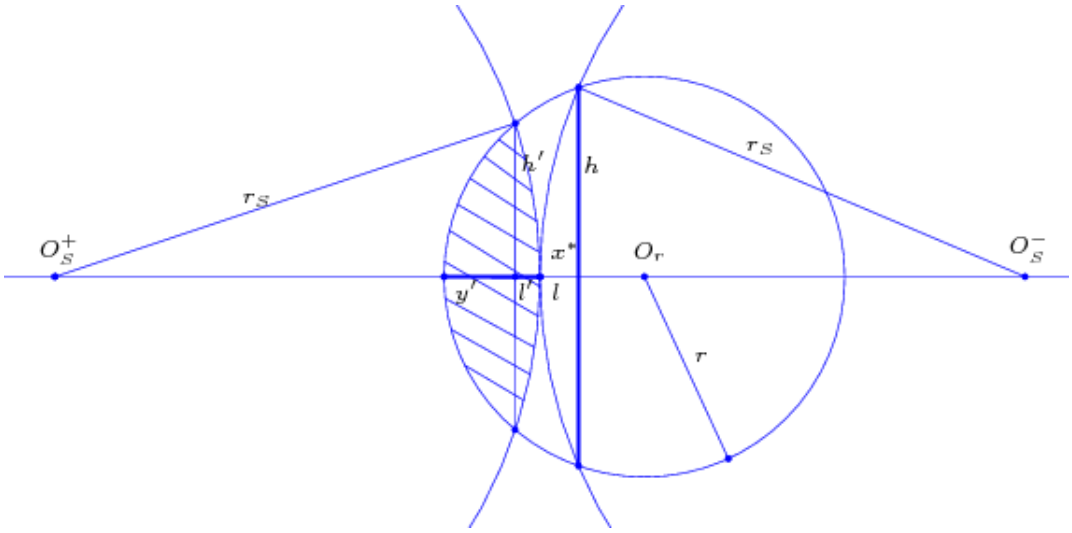

FIG 4. *Definition of some useful lengths for Lemma 4*

4.1.3. *Conclusions.*

For Theorem 1. Lemmas 1, 2, 3 and 4 prove that both $d(D_r, S)$ and $d(\partial D_r, \partial S)$ are e.a.s. bounded. Namely, let us denote:

$$a_1 = (2\left(\frac{2d+1+2\alpha}{d+1+2\alpha}\right)^{\frac{2}{d+1+2\alpha}} (f_0\theta_{d,\alpha}^{ellipse})^{-\frac{2}{d+1+2\alpha}},$$

$$a_2 = \frac{r_S + r}{r_S r^{\frac{d-1}{d+1+2\alpha}}} (f_0\theta_{d,\alpha}^{lens})^{-\frac{2}{d+1+2\alpha}},$$

$$a_3 = \frac{(r_S + r) r^{\frac{2\alpha}{d+1+2\alpha}}}{r_S - r} (f_0\theta_{d,\alpha}^{lens})^{-\frac{2}{d+1+2\alpha}}.$$

The bound is $\max(a_1, a_2, a_3)$. The constants a_1 and a_2 bound $\max_{x\in S}(d(x, D_r))$ and decrease as r increases. The constant a_3 bounds $\max_{x\in D_r}(d(xS))$ and increases as r increases.

For Property 1. Within the proof of Lemma 4 one can see that the maximum edge length of $\mathcal{D}_r$ is e.a.s. bounded by $2\sqrt{2a_3 r_S}\left(\frac{\ln n}{n}\right)^{\frac{1}{d+1+2\alpha}}$. The fact that $a_3(r)$ is a bijective function from $]0, r_S[$ to $\mathbb{R}_+^*$ gives Property 1.

4.2. *Proof of useful Lemmas.* We are now going to describe several Lemmas which will be used to finish the proofs of Lemmas 1, 2, 3 and 4. Let us

recall that the proofs of these Lemmas are based on the fact that in each case the contrary implies the existence of some subset of S that does not contain any observation. To prove that this is not possible (e.a.s.) we will first set some gives some geometric lemmas, then we will give lower bounds for the probability that an observation belongs to these sets. Finally we will give the proofs of Lemmas 10, 12 and 11 which are required for Lemmas 1, 2, 3 and 4.

4.2.1. *Geometric preliminaries.*

Covering numbers.

LEMMA 5. *Let ε_n be sequence that converges to 0. There exists N_0 and a constant $c_0(S)$ such that, for all $n > N_0$, $\nu_0(n)$ the covering number of S with small balls of radius ε_n centered in S satisfies $\nu_0(n) \leq c_0(S) r_n^{-d}$. That is, one can deterministically find points $x_1, \ldots, x_{\nu_0(n)}$ in S such that $S \subset \bigcup \mathcal{B}(x_i, \varepsilon_n)$ with $\nu_0(n) \leq c_0(S)\varepsilon_n^{-d}$*

This is a well known result that we will use to get probabilistic inequalities. Namely, the two constants used will be : $c_0(S)$ when we will deal with empty balls centered in S or empty ellipsoids centered in S (for Lemmas 1 and 3) and $c_0(S + r\overline{\mathcal{B}})$ when we deal with empty balls centered outside of S but that intersect (Lemmas 2 and 3).

Interpolation type Lemma.. To bound the distance between a point that belongs to a simplex σ (whose vertices belongs to S) and S (used in the second part of the proof of Theorem 1) we need the following lemma whose proof is similar to the error bound when estimation is done via linear interpolation. This proof is left to the reader.

LEMMA 6. *If x belongs to a simplex $\sigma = (x_1, \ldots, x_{k+1})$ such that every x_i belongs to S then $d(x, S) \leq \max(||x_i - x_j||^2)/8r_S$.*

Local Properties of S and some probability bounds.

LEMMA 7. *There exists a ε_0 such that, for all $x \in \mathbb{R}^d$ such that $d(x, \partial S) = \varepsilon \leq \varepsilon_0$, there exists a unique $x^* = \operatorname{argmin}_{y \in \partial S} ||\overrightarrow{yx}||$.*

PROOF. Let us orient ∂S so that, for all $y \in \partial S$, the vector $\overrightarrow{u}_y$ is the unique unit vector normal to ∂S at the point y that points outward from S. As ∂S is $\mathcal{C}^2$, the Gauss map that associates $\overrightarrow{u}_y$ to y is $\mathcal{C}^1$ and $\overrightarrow{u}_y - \overrightarrow{u}_z = A_y\overrightarrow{yz} + o_y(||\overrightarrow{yz}||)$. Compactness of ∂S implies that there exists $\varepsilon_1 > 0$ such

that, for all $\varepsilon \leq \varepsilon_1$, for all $x \in \partial S$ and for any vector $\overrightarrow{v}$, $||(Id - \varepsilon A_x)\overrightarrow{v}|| \geq \frac{2}{3}||\overrightarrow{v}||$. Compactness of ∂S also implies that there exists an ε_2 such that for all $x \in \partial S$ and for any vector $\overrightarrow{v}$ with $||\overrightarrow{v}|| \leq \varepsilon_2$, $||o_y(||\overrightarrow{v}||)|| \leq ||\overrightarrow{v}||/3$. Let us choose $\varepsilon_0 = 0.5 \min(\varepsilon_1, \varepsilon_2)$.

Let us notice that the compactness of ∂S implies that for all $x \in \mathbb{R}^d$ there exists a $x^* = \text{argmin}_{y \in \partial S} ||\overrightarrow{yx}||$ and $\overrightarrow{xx^*}$ is a vector normal to ∂S at the point x^*.

Let us suppose that there exists a point $x \in S$ (the proof when $x \in S^c$ is exactly the same) such that $d(x, \partial S) = \varepsilon \leq \varepsilon_0$, and that there exist two points x_1^* and x_2^* in ∂S that realize the minimum i.e. $||\overrightarrow{x_1^* x}|| = ||\overrightarrow{x_2^* x}|| = \min_{y \in \partial S} ||\overrightarrow{yx}||$. We have $x = x_1^* - \varepsilon \overrightarrow{u}_{x_1^*} = x_2^* - \varepsilon \overrightarrow{u}_{x_2^*}$ which implies $||\overrightarrow{x_1^* x_2^*}|| \leq 2\varepsilon$. Let us recall that $\overrightarrow{u}_{x_2^*} = A_{x_1^*} \overrightarrow{x_1^* x_2^*} + o_{x_1^*}(||\overrightarrow{x_1^* x_2^*}||)$ and so

$$x_2^* - \varepsilon \overrightarrow{u}_{x_2^*} - (x_1^* - \varepsilon \overrightarrow{u}_{x_1^*}) = \overrightarrow{x_1^* x_2^*} + \varepsilon A_{x_1^*} \overrightarrow{x_1^* x_2^*} + \varepsilon o_{x_1^*}(||\overrightarrow{x_1^* x_2^*}||).$$

This concludes the proof as it implies that $O = ||x - x|| \geq ||\overrightarrow{x_1^* x_2^*}||/3$. □

Lemma 7 implies the following corollary.

COROLLARY 1. *Let X be a random variable of density f.*

For all x such that $d(x, \partial S) \leq \varepsilon_0$ we can define the unique $x^* \in \partial S$ such that $||\overrightarrow{xx^*}|| = d(x, \partial S)$ and denote $\overrightarrow{u_x} = \overrightarrow{xx^*}/||\overrightarrow{xx^*}||$.

Thus, for all $r_n \leq d(x, \partial S)$ one can define uniquely the following ellipsoid:

$$\mathcal{E}^*_{x,r_n} = \{y, \frac{(\overrightarrow{xy}.\overrightarrow{u}_x)^2}{r_n^2} + \frac{||\overrightarrow{xy} - (\overrightarrow{xy}.\overrightarrow{u}_x)\overrightarrow{u}_x||^2}{r_S r_n} \leq 1\}$$

which satisfies:

$$P(X \in \mathcal{E}_{x,r_n}) \geq f_0 \theta^{ellipse}_{d,\alpha} r_S^{\frac{d-1}{2}} r_n^{\frac{d+1+2\alpha}{2}} (1 + o(1))$$

The sketch of the proof is as follows: First, the uniqueness is a corollary of Lemma 7 (it is important to define uniquely the ellipsoid to prove Lemma 12). Then the probability bounds comes from the hypothesis on f. Without giving details of the calculation they are based on

$$P(X \in \mathcal{E}_{x,r_n}) \leq P(X \in \mathcal{E}_{0,r_n})$$

with O the point that satisfies $\overrightarrow{x^* O} = r_n \overrightarrow{x^* x}/||\overrightarrow{x^* x}||$, and:

$$P(X \in \mathcal{E}_{O,r_n}) \sim \int_{-r_n}^{r_n} f_0(t+r_n)^\alpha \theta_{d-1} \left(\sqrt{r_n r_S} \sqrt{1 - \frac{t^2}{r_n^2}} \right)^{d-1} dt.$$

(See Figure 5). Applying a variable changes $t/r_n = u$ gives the result.

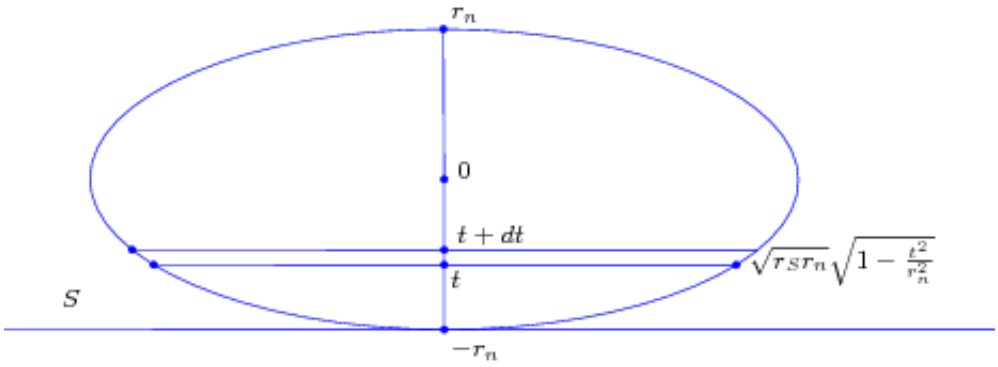

FIG 5. *Probability integration in an ellipsoid*

LEMMA 8. *If X is a random variable of density f, $x \in S$ is a deterministic point, and r_n a radius sequence such that $r_n \to 0$ then*

$$P(X \in \mathcal{B}(x, r_n)) \geq \frac{1}{2} f_0 \theta_{d,\alpha} r_n^{d+\alpha} (1 + o(1)).$$

Here again we only present the idea of the proof. The limit case that bounds the probability is given when $x \in \partial S$ and, for this case,

$$P(X \in \mathcal{B}(x, r_n)) \sim \int_0^{r_n} \theta_{d-1} f_0 t^\alpha (r_n^2 - t^2)^{\frac{d-1}{2}} dt.$$

LEMMA 9. *Let r be a radius with $r < r_S$, let ε_n be a sequence that converges toward 0 and let x_n be a point sequence such that that x_n does not belong to S and such that $d(x, \partial S) = r - \varepsilon_n$. If X is a random variable of density f then*

$$P(\mathcal{B}(x_n, r) \cap S) \geq f_0 \theta_{d,\alpha}^{lens} \frac{r^{\frac{d-1}{2}} r_S^{\frac{d+1+2\alpha}{2}}}{(r_S + r)^{\frac{d+1+2\alpha}{2}}} \varepsilon_n^{\frac{d+1+2\alpha}{2}} (1 + o(1)).$$

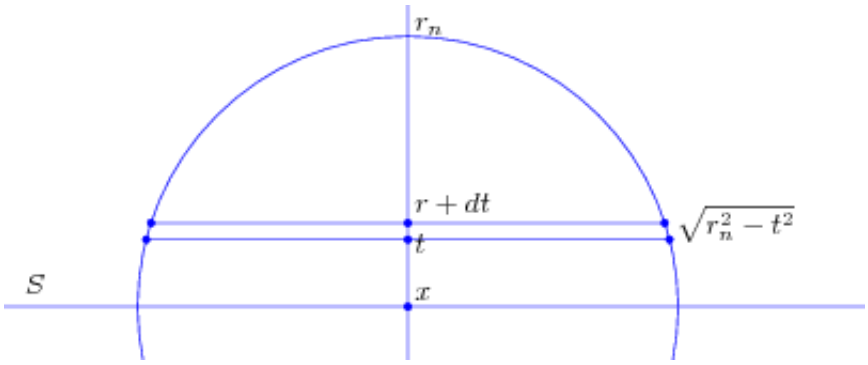

FIG 6. *Probability integration in a disc*

This case is the same as illustrated in Figure 4. Let us recall that the solutions for y', l, l' and h' (with $y'+l'=\varepsilon_n$):

$$\begin{cases} h^2 = \frac{2rr_S}{r_S-r}\varepsilon_n + o(\varepsilon_n) \\ l = \frac{r}{r_S-r}\varepsilon_n + o(\varepsilon_n) \\ h'^2 = \frac{2rr_S}{r_S-r}\varepsilon_n + o(\varepsilon_n) \\ y' = \frac{r_S}{r+r_S}\varepsilon_n + o(\varepsilon_n) \end{cases}$$

Referring to Figure 7 we see that :

$$P(\mathcal{B}(x_n,r)\cap S) \leq \int_0^{y'} f_0(t+l')^\alpha \theta_{d-1}\left(\frac{t}{y'}\right)^{d-1} dt \leq \int_0^{y'} f_0 t^\alpha \theta_{d-1}\left(\frac{t}{y'}\right)^{d-1} dt.$$

4.2.2. *Probability bounds for the existence of a given "empty" set.* The previous section establishes the uniqueness of the ellipsoid-type set and gives lower bounds for the probability for one observation to fall in a given set (ellipsoid, ball or lens). To conclude the proof we now need to bound the probability that there exists an empty set.

LEMMA 10. *Let r_n be the sequence $r_n^{d+\alpha} = \frac{2\lambda}{\theta_{d,\alpha} f_0}\frac{\ln n}{n}$ with $\lambda \geq 1$ then*

$$P(\textit{exists } x \in S \textit{ such that } \mathcal{B}(x,r_n)\cap \mathcal{X}_n = \emptyset) = O(n^{\frac{d}{d+\alpha}-\lambda+o(1)}).$$

As a corollary ; for all $\lambda > 1+\frac{d}{d+\alpha}$, and for all $x \in S$, $\mathcal{B}\left(x, \left(\frac{2\lambda}{\theta_{d,\alpha} f_0}\frac{\ln n}{n}\right)^{\frac{1}{d+\alpha}}\right)$ contains at least one observation e.a.s.

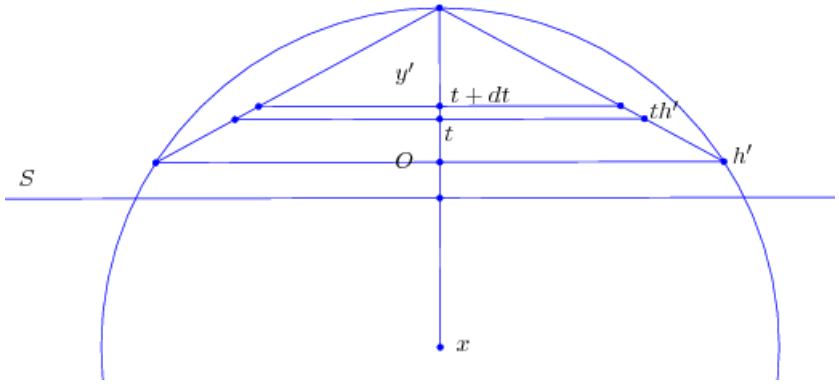

FIG 7. *Probability integration in a lens*

PROOF. Let us first cover S with $\nu_0 \leq c_0(S)\varepsilon_n^{-d} r_n^{-d}$ small balls of radius $r_n\varepsilon_n$ as in Lemma 5 and let us denote x_i^* the centers of these balls. Let us suppose that there exists an $x \in S$ such that $\mathcal{B}(x, r_n) \cap \mathcal{X}_n = \emptyset$. Then there exists an i such that $\mathcal{B}(x_i^*, r_n(1-\varepsilon_n)) \cap \mathcal{X}_n = \emptyset$. Lemmas 5 and 8 imply that

$$P(\exists i,\, \mathcal{B}(x_i^*, r_n) \cap \mathcal{X}_n = \emptyset) \leq c_0(S)\ln(n) r_n^{-d}\varepsilon_n^{-d}\left(1 - \frac{f_0\theta_{d,\alpha}}{2} r_n^{d+\alpha}(1+o(1))\right)^n.$$

and so, for the given r_n:

$$P(\exists i,\, \mathcal{B}(x_i^*, r_n) \cap \mathcal{X}_n = \emptyset) = O\left(\varepsilon_n^{-d}\ln(n)^{-\frac{d}{d+\alpha}} n^{\frac{d}{d+\alpha}-\lambda+o(1)}\right).$$

The choice of $\varepsilon_n = \ln(n)^{-\frac{1}{d+\alpha}}$ concludes the proof. □

LEMMA 11. *For all $x \notin S$ such that $d(x_n, \partial S) \leq r - (a\ln(n)/n)^{2/(d+1+2\alpha)}$ with $a = \lambda \frac{(r_S+r)^{\frac{d+1+2\alpha}{2}}}{f_0\theta_{d,\alpha}^{lens} r^{\frac{d-1}{2}} r_S^{\frac{d+1+2\alpha}{2}}}$ and $\lambda > 1$ we have $\mathcal{B}(x,r) \cap \mathcal{X}_n \neq \emptyset$ e.a.s.*

PROOF. Let us define $E_{n,a,r} = \{x \notin S, d(x, \partial S) \leq r - (a\ln(n)/n)^{2/(d+1+2\alpha)}\}$ and $E_0 = \{x \notin S, d(x, \partial S) \leq r\}$. Let us now cover $E_{n,a,r}$ with $\nu_n \leq c(E_{n,a,r})\varepsilon_n^{-d} \leq c(E_0)\varepsilon_n^{-d}$ small deterministic balls of radius ε_n, and centered at the x_i.

Let us suppose that there exists an $x \in E_{n,a,r}$ such that $\mathcal{B}(x,r) \cap \mathcal{X}_n$ is empty. Then there exists a deterministic $x_i \in E_{n,a,r}$ such that $\mathcal{B}(x_i, r(1-\varepsilon_n))$ does not contains any observation.

Let us denote

$$p_n = P(\exists x \in E_{n,a,r}, \mathcal{B}(x,r) \cap \mathcal{X}_n = \emptyset).$$

Applying Lemma 9 and a reasoning similar to previous one we have:

$$p_n \leq p_n^* = c(E_{0,r_S})\varepsilon_n^{-d}\left(1 - \lambda\frac{\ln n}{n}(1+o(1))\right)^n = O(\varepsilon_n^{-d} n^{-\lambda+o(1)}).$$

When $\lambda > 1$, the choice of $\varepsilon = (\ln n)^{-1}$, for instance, leads to $\sum p_n < \infty$ and so concludes the proof. □

COROLLARY 2. *When, $r < r_S$, if $x \notin S$ and $\mathcal{B}(x,r) \cap \mathcal{X}_n = \emptyset$ then*

$$d(x, \partial S) \geq r - \frac{(r_S + r)}{f_0 \theta_{d,\alpha}^{lens} r^{\frac{d-1}{d+1+2\alpha}} r_S}\left(\frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}} \quad e.a.s.$$

LEMMA 12. *For all $x \in S$ such that $d(x, \partial S) \geq \left(\frac{\lambda}{f_0 \theta_{d,\alpha}^{ellipse} r_S^{\frac{d-1}{2}}} \frac{\ln n}{n}\right)^{\frac{2}{d+1+2\alpha}}$ and $\lambda > 1 + \frac{1}{d+1+2\alpha}$ we have*

$$\mathcal{E}_{x,d(x,\partial S)} \cap \mathcal{X}_n \neq \emptyset \ e.a.s.$$

PROOF. The proof is basically the same as for the previous lemmas (Lemmas 10 and 11).

Let us first cover S with $c_0(S)\varepsilon_n^{-d}$ small deterministic balls of radius ε_n centered at the x_i. Let us suppose that there exists an $x \in S$ such that $d(x, \partial S) = \rho_n$ and $\mathcal{E}_{x,\rho_n} \cap \mathcal{X}_n = \emptyset$. There exists an x_i such that $x \in \mathcal{B}(x_i, \varepsilon_n)$.

First, the $\mathcal{C}^2$ smoothness of the boundary implies that $\overrightarrow{u}_{x_i} = \overrightarrow{u}_x + \varepsilon_n O_x(1)$ and the compactness of S allows us to have a uniform upper bound i.e. $\overrightarrow{u}_{x_i} = \overrightarrow{u}_x + \varepsilon_n O(1)$.

Let us first note that $\mathcal{E}_{x,d(x,\partial S)} \subset \mathcal{B}(x, \sqrt{r_S \rho_n})$. For all $y \in \mathcal{E}_{x,d(x,\partial S)}$ and so $||x - y|| = O(\rho_n)$.

This allows us to have

$$(\overrightarrow{u}_x . \overrightarrow{xy})^2 = (\overrightarrow{u}_{x_i} . \overrightarrow{x_i y})^2 + \varepsilon_n \sqrt{\rho_n} O(1)$$

and

$$||\overrightarrow{xy} - (\overrightarrow{u}_x . \overrightarrow{xy})\overrightarrow{u}_x||^2 = ||\overrightarrow{x_i y} - (\overrightarrow{u}_{x_i} . \overrightarrow{x_i y})\overrightarrow{u}_{x_i}||^2 + \varepsilon_n O(1).$$

That implies that there exists a constant A such that:

$$\mathcal{E}_{x_i,\frac{\rho_n}{1+\frac{A\varepsilon_n(1+r_S)}{\sqrt{\rho_n}}}} \subset \mathcal{E}_{x,\rho_n}$$

We can now apply Lemma 1:

$$P(\exists x \text{ such that } d(x,\partial S)=\rho_n \text{ and } \mathcal{E}_{x,\rho_n}\cap\mathcal{X}_n=\emptyset) \leq$$
$$c_0(S)\varepsilon_n^{-d}\left(1-f_0\theta_{d,\alpha}^{ellipse}r_S^{\frac{d-1}{2}}\left(\frac{\rho_n}{1+\frac{A\varepsilon_n(1+r_S)}{\sqrt{\rho_n}}}\right)^{\frac{d+1+2\alpha}{2}}(1+o(1))\right)^n$$

When $\rho_n^{\frac{d+1+2\alpha}{2}} = \frac{\lambda}{f_0\theta_{d,\alpha}^{ellipse}r_S^{\frac{d-1}{2}}}\frac{\ln n}{n}$ and $\varepsilon_n\sqrt{\rho_n}^{-1}\to 0$ we have:

$$P(\exists i \text{ such that } \mathcal{E}_{x,\rho_n}\cap\mathcal{X}_n=\emptyset) \leq c_0(S)\varepsilon_n^{-d}n^{-\lambda+o(1)}.$$

If we now choose $\varepsilon_n=(\ln n)^{-a}\sqrt{\rho_n}$ and $\lambda > 1+\frac{1}{d+1+2\alpha}$ we have
$\sum P(\exists x \text{ such that } d(x,\partial S)\geq\rho_n \text{ and } \mathcal{E}_{x,\rho_n}\cap\mathcal{X}_n=\emptyset)<\infty$
This concludes the proof. □

5. Conclusion and Perspective. First, as mentioned in the discussion on the hypothesis, it would be interesting to remove the dimension hypothesis on the support to allow one to deal with sparsity problems.

The radius for D_r (or radii sequence for R_{r_n} or H_{r_n}) that have to be chosen and the convergence rate depends on some unknown parameters: f_0, α and r_S. That is a classical problem in statistics and one may replace these unknown quantities by their estimators. When $\alpha=0$ there exist estimators for f_0 (see [22] for instance), but, to our knowledge, there is none for f_0 and α in the general case, and we have not seen any for r_S.

Another way to search is to have the usual dual approach replacing a fixed radius sequence with a local one with a nearest-neighbors method. We believe that this may remove the unknown f_0 for the choice of the parameter and give better results in practice when the density is far from uniform.

The proposed estimators have a long computational time (which is very sensitive to the observation dimension) but we have chosen to present the H_r estimator in a way that allows us to now introduce now H_r^o and $H_r^{\square}$. The idea is very simple: the practical problem in H_r is the computation of the convex hull. Let us replace it by the smallest set that contains $\overline{\mathcal{B}}(X_i,r)\cap\mathcal{X}_n$ with a given geometrical shape.

DEFINITION 6. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a subset of $\mathbb{R}^d$. The $r-$Union of ellipsoids, denoted $H_r^o(\mathcal{X}_n)$, is defined as follows:*

$$H_r^o(\mathcal{X}_n) = \bigcup_i \mathcal{E}(\overline{\mathcal{B}}(X_i, r) \cap \mathcal{X}_n)$$

where $\mathcal{E}(A)$ is the smallest ellipsoid that contains A.

DEFINITION 7. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a subset of $\mathbb{R}^d$. The $r-$Union of bins, denoted $H_r^{\square}(\mathcal{X}_n)$, is defined as follows:*

$$H_r^{\square}(\mathcal{X}_n) = \bigcup_i C(\overline{\mathcal{B}}(X_i, r) \cap \mathcal{X}_n)$$

where $C(A)$ is the smallest bin that contains A (with the following bin definition : B is a bin if it is isometric to a $\pi_i[0, a_i]$)

We believe that these new estimators may give similar results to the one proposed here but with a better computational time (we already have a part of the proof, as we obviously have $H_r \subset H_r^o$ and $H_r \subset H_r^{\square}$). A final result on this new estimator together with theoretical result that allows one to deal with a manifold of dimension smaller than d should give very interesting and practical methods to deal with sparse data.

Another possible perspective is the introduction of the support estimation in the nearest neighbor density estimation. When the density is estimated with the nearest neighbors method there exists a bias for the points located on (or close to) the boundary when the boundary is sharp. The idea is to use the support estimation to reduce this bias. Once this point achieve it can be imagined to adapt this method to the level set estimation that has much more practical interest that the whole support estimation.

To conclude this long list of further potential work we can also notice that the restricted Delaunay estimator provides a piecewise linear approximation of the boundary with extremities lying on a subset of the observations, a smoother estimator as a polynomial one, based on the identified points may now be built and may over-perform ours (as in [19] but for higher dimensions).

REFERENCES

[1] C. Aaron. Using the k-nearest neighbor restricted delaunay polyhedron to estimate the density support and its topological properties. submitted to Electronic Journal of statistics.

[2] A. Baillo, A. Cuevas, and A. Justel. Set estimation and nonparametric detection. *The Canadian Journal of Statistics*, 28:765–782, 2000.

[3] I. Bárány. Random polytopes in smooth convex bodie. *Mathematika*, 39:81–92, 1982.

[4] G. Biau, B. Cadre, D.M. Mason, and B. Pelletier. Asymptotic normality in density support estimation. *Electronic Journal of Probability*, pages 2617–2635, 2009.

[5] G. Biau, B. Cadre, and B. Pelletier. Exact rates in density support estimation. *Journal of Multivariate Analysis*, 99:2185–2207, 2008.

[6] P. Bubenik, G. Carlsson, P. T. Kim, and Z. M. Luo. Statistical topology via morse theory, persistence, and nonparametric estimation. contemporary mathematics, 516 (2010). *Algebraic Methods in Statistics and Probability II.Contemporary Mathematics,*, 516:75–92, 2010.

[7] G. Carlsson. Persistent homology and the analysis of high dimensional data. In *Symposium on the Geometry of Very Large Data Sets , Fields Institute for Research in Mathematical Sciences*, 2005.

[8] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46:255–308, 2009.

[9] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11:149–187, 2005.

[10] F. Chazal, D. Cohen-Steiner, and A. Lieutier. Normal cone approximation and offset shape isotopy. *Computational Geometry : Theory and Applications*, 42:566–581, 2009.

[11] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in euclidean spaces. *Discrete and Computational Geometry*, 41:461–479, 2009.

[12] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures,. *Journal on Foundations of computational Mathematics*, 11:733–751, 2011.

[13] M.K. Chung, P. Bubenik, and P.T. Kim. *Information Processing in Medical Imaging 2009*, chapter Persistence Diagrams of Cortical Surface Data., pages 386–397. Lecture Notes in Computer Science, 2009.

[14] A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *The Annals of Statistics*, 25:2300–2312, 1997.

[15] A. Cuevas and A. Rodríguez-Casal. On boundary estimation. *Advanced in Applied Probability*, 36:340–354, 2004.

[16] L. Devroye and G.L. Wise. Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal of Applied Mathematics*, 38:480–488, 1980.

[17] H. Edelsbrunner and N. R. Shah. Triangulating topological spaces. *Internat. J. Comput. Geom. Appl.*, 7:365–378, 1997.

[18] B. Efron. The convex hull of a random set of points. *biometrika*, 15:331–343, 1965.

[19] W. Hardle, B.U. Park, and A.B. Tsybakov. Estimation of non-sharp support boundary. *Journal of Multivariate Analysis*, 55:205–218, 1995.

[20] M. Kahle and E. Meckes. Limit theorems for betti numbers of random simplicial complexes. arXiv:1009.4130.

[21] J. Klemelä. Complexity penalized support estimation. *Journal of Multivariate Analysis*, 88:274–297, 2004.

[22] M.D. Penrose. A strong law for the largest nearest-neighbour link between random points. *Journal of the London Mathematical Society*, 60:951–960, 1999.

[23] M. Reitzner. Random polytopes and the efronstein jackknife inequality,. *Ann. Probab.,*, 31:21362166., 2003.

[24] C. Schütt. Random polytopes and affine surface area. *Math. Nachr.*, 170:227–249, 1994.

[25] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:247–274, 2005.