



**HAL**  
open science

## Local Convex Hull support and boundary estimation

Catherine Aaron

► **To cite this version:**

| Catherine Aaron. Local Convex Hull support and boundary estimation. 2013. hal-00786393v3

**HAL Id: hal-00786393**

**<https://hal.science/hal-00786393v3>**

Preprint submitted on 2 Dec 2013 (v3), last revised 19 Dec 2014 (v5)

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

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

# Local Convex Hull support and boundary estimation

Catherine Aaron, Université Blaise Pascal

December 2, 2013

## Abstract

We consider random samples in  $\mathbb{R}^d$  drawn from an unknown density. This paper is devoted to presenting a new estimator of the support of the density, which is based on a local convexity criteria. We prove that the estimator is consistent and that it also provides a consistent estimator for the boundary. Some convergence rates are given depending on different assumptions and one can also prove that when the boundary is smooth enough and when the density go to 0 as a power of the distance to the boundary the estimator is (eventually almost surely) homeomorph to the support.

**Key Words:** Convex-Hull polyhedron, support estimation, topological data analysis, geometric inference.

## 1 Introduction and further notations

Let  $X_1, \dots, X_n$  be a sample drawn from  $f$  an unknown density function defined on  $\mathbb{R}^d$ . We are concerned in the problem of estimating

$$S = \overline{\{x \in \mathbb{R}^d, f(x) > 0\}} \text{ and } \partial S = \overline{S} \setminus \overset{\circ}{S}$$

the support,  $\partial S$  its boundary (with  $\overline{A}$  denoting the closure of the set  $A$  and  $\overset{\circ}{A}$  its interior). We also expect that the estimated support is, eventually almost surely homeomorph to the support in a way to be able estimate some topological invariants (a sequence of event  $\mathcal{A}_n$  occurs eventually almost surely when  $1_{\mathcal{A}_n} \xrightarrow{a.s.} 1$ ).

All this points has been, one by one, studied for a long time let us cite, [Che76], [DW80], [CF97], [Kle04], [Mei06] (a review of different method can be found in [CF09]) for the problem of the support estimation, [CRC04] [HPT95], for the problem of the problem of the boundary estimation and

[CCSM11], [CCSL09], [CGOS11], [Car05], [ZC05], [CZCG05] for the problem of the topological recognition but they are rarely mixed together (However one can cite).

To understand a little the kind of problem that occurs let us focus on two specific methods. First, if we now that the support is convex (but that is a strong assumption), one can naturally estimate it with the convex hull of the observations:  $\hat{H} = \mathcal{H}(\mathcal{X}_n)$  (here  $\mathcal{H}(A)$  denotes the convex hull of the set  $A$  and  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  denotes the set associated to the sample). This estimator clearly fulfill our three goals. The support estimator is  $\hat{H}$  is obviously homeomorph to the support its convergence rate has been extensively studied (see for instance [Efr65], [B82], [Rei03], [Sch88] or [DW96]), , its natural boundary estimator is  $\partial\hat{H}$  also converges.

Second, if no hypothesis on the support can be made (and that is a more challenging case), the most used (because the simplest) support estimator is the Devroye-Wise estimator ([Che76], [?]) defined as follows:

$$\hat{S}_r = \bigcup_{X \in \mathcal{X}_n} \mathcal{B}(X, r)$$

Here  $\mathcal{B}(X, r)$  denotes the closed ball centered in  $x$  and of radius  $r$ .

The properties of  $\hat{S}_r$  as a support estimator has been extensively studied (see [BCP08] or [BCMP09]). Properties of  $\partial\hat{S}_r$  as an estimator of  $\partial S$  has also been studied in ([CRC04]). Such an estimator can also be related to the topology recognition via  $\alpha$ -shape method ([ES97], [EKS83]). The main problem is that we empirically observed that the “best” parameter  $r$  when using the Devroye-Wise estimator to estimate the support can gives not converging estimator for its boundary (and the topological recognition problem). To understand this phenomena is quite easy. To evaluate the estimator it is usually proposed to evaluate the measure of the symmetric difference  $|\hat{S}_r \Delta S|$  or  $d_H(S, \hat{S}_r)$  the Hausdorff distance defined as follows:

$|A \Delta B| = |(A \cap B^c) \cup (A^c \cap B)|$  where  $A^c$  denotes the complementary of the set  $A$  and  $|A|$  the Lebesgue measure of the set  $A$ .

$d_H(A, B) = \max(\max_{a \in A}(\min_{b \in B} \|a-b\|), \max_{b \in B}(\min_{a \in A} \|a-b\|))$  where  $\|a-b\|$  denotes the euclidean distance between the two points.

For both criteria it is easy to see that we can have very small “distances” between sets but the existence of very small “holes” in the estimator (i.e. with the existence of  $x \in S$  such that  $\mathcal{B}(x, \varepsilon) \subset \hat{S}_r$ ). If there exists such “holes” located far from the boundary (and that empirically happen, it can be observed in section 5)  $\partial\hat{S}_r$  is not a good estimator of  $\partial S$ . To avoid such a problem, it is common ([CRC04]) to impose  $S \subset \hat{S}_r$  when estimating the boundary. Obviously imposing such a condition will decrease the perfor-

mance of the support estimator. More precisely here the support estimator will “overflow” too much the support.

We propose, in this paper to study the following support estimator:

$$\hat{H}_r = \bigcup_{X \in \mathcal{X}_n} \mathcal{H}(\mathcal{B}(X, r) \cap \mathcal{X}_n)$$

The first idea was to decrease the “overflowing problem” (see Figure 1), as, even if  $\mathcal{B}(X, r)$  overflow  $S$  the fact that we now use  $\mathcal{H}(\mathcal{B}(X, r) \cap \mathcal{X}_n)$  may decrease this problem when the boundary of the support is smooth enough. We so expect to improve the boundary estimation and the support estimation in the case of support close to a  $d'$ -dimensional manifold (with  $d' \leq d$  the dimension of the observation).

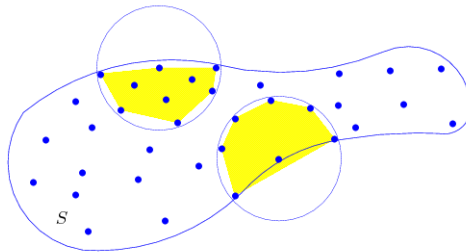


Figure 1: Even if the two balls “overflow” a lot the support, the local convex hulls does not overflow that much the support.

Let us notice that the first idea of using a local-convex hull estimator has been introduced in ??, with nearest-neighbors instead of fixed radius, and apply it to estimate not only the support but also the level line of the density and use it in home-range estimation. As the method gives very good results it has been applied in various paper in ecologic problem, home-range estimation and so on. However, it has never been mathematically studied.

The paper is dedicated to a first study of such an estimator, focusing on the estimation of the support, its boundary and the recognition of the topological properties.

In Section 2 we settle a general theorem that only requires that the support is compact. Then, in section 3 we study some convergence rates according to two sets of shape conditions. The first one is the standarness that has been originally introduced in [CF97] and that is an usual assumption in support estimation. The second one concern the smoothness of the boundary (with a condition close to  $\mathcal{C}^2$ ) and the density that we assume to decrease to

0 as a power  $\alpha$  of the distance to the boundary (this set of assumption is a case of this used in ??).

To conclude the theoretical study of  $\hat{H}_r$ , in section 4 we focus on the recognition of the topology. We prove that, when the boundary of the support is smooth enough and when the density decreases to 0 as a power  $\alpha$  of the distance to the boundary the choice of a  $r_n$  sequence proportional to  $(\ln n/n)^{1/(d+2\alpha+1)}$  provides estimators eventually almost surely homeomorph to  $S$ .

In section 5 we present some simulations and comparisons between  $\hat{H}_r$  and  $\hat{S}_r$  for the estimation of the support and its boundary.

### Additional notations.

Here we give some additional notations used in the paper.

for a set  $A$ , the Minkowski sum (resp. difference) of  $A$  and balls of radius  $\varepsilon$  are respectively denotes  $A^\varepsilon$  and  $A^{-\varepsilon}$  with the following definitions.

$$A^\varepsilon = \bigcup_{a \in A} \mathcal{B}(a, \varepsilon) \text{ and } A^{-\varepsilon} = \{a \in A, \mathcal{B}(a, \varepsilon) \subset A\}$$

The Hausdorff distance between two sets  $A$  and  $B$  can be written  $d_H(A, B) = \inf\{\varepsilon, A \subset B^{+\varepsilon}, B \subset A^{+\varepsilon}\}$

Let  $A$  be a  $d'$  dimensional set in  $\mathbb{R}^d$ . If we denote by  $N(A, \varepsilon)$  the inner covering number (i.e. the minimal number of small balls, of radius  $\varepsilon$  and centered in  $A$  that cover  $A$ ), we clearly may guarantee the existence of positive values  $\lambda(A)$  and  $r_A$  such that:  $N(A, \varepsilon) \leq \lambda(A)\varepsilon^{-d'}$  when  $\varepsilon \leq r_A$ .

## 2 Universal consistency

Here we are interrelated in a very general result on  $\hat{H}_{r_n}$  the main assumptions are that the support is compact,  $d$ -dimensional (with  $d$  the dimension of the observation space), with a  $d'$  dimensional boundary (when the boundary is smooth enough  $d' = d - 1$  but here we also allow all the cases where  $d' \in [d - 1, d[$  such as  $C \times [0, 1]$  where  $C$  is the cantor set and  $d' = 1 + \ln(2)/\ln(3)$  or the inside of the Koch snowflake where  $d' = 4/3$ ). We also requires the existence of sequences  $r_n$  that satisfies some condition very similar to those required in [CRC04].

**Theorem 1.** *Let  $S \subset \mathbb{R}^d$  be a compact  $d$ -dimensional set such that its boundary  $\partial S$  is  $d'$ -dimensional. Let  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  be the set of independent random observations drawn from a distribution  $P_X$  with support  $S$ . Assume that:*

a)  $S \subset \hat{S}_{r_n/4}$  eventually almost surely

b)  $r_n \rightarrow 0$  almost surely

Then:

i)  $|\hat{H}_{r_n} \Delta S| \leq \lambda(\partial S) \omega_d 2^d r_n^{d-d'}$  eventually almost surely.

ii)  $d_H(\partial \hat{H}_{r_n}, \partial S) \rightarrow 0$  almost surely.

The rest of the section is dedicated to the proof of this theorem.

**Lemma 1.** *If  $S \subset \hat{S}_{r/4}$  then  $(\hat{H}_r \cap S^c) \cup (\hat{H}_r^c \cap S) \subset \partial S^{+r}$ .*

*Proof.* First let us focus on  $\hat{H}_r \cap S^c$ . It is obvious that  $\hat{H}_r \subset \hat{S}_r \subset S^{+r}$ . So  $\hat{H}_r \cap S^c \subset S^{+r} \cap S^c \subset \partial S^{+r}$ . To prove the last inclusion let us choose  $x \in S^{+r} \cap S^c$ ,  $x \in S^c$  and there exists a  $y \in S$  with  $\|x - y\| \leq r$  so  $[x, y]$  intersects  $\partial S$  at a point  $z$  with  $\|x - z\| \leq r$ .

Second we study  $\hat{H}_r^c \cap S$  and divide this set as follows:  $\hat{H}_r^c \cap S = (\hat{H}_r^c \cap S^{-r/2}) \cup (\hat{H}_r^c \cap (S \setminus S^{-r/2}))$ . Then we prove that  $\hat{H}_r^c \cap (S \setminus S^{-r/2}) \subset \partial S^{+r/2}$  and that  $\hat{H}_r^c \cap S^{-r/2} = \emptyset$ .

1)  $\hat{H}_r^c \cap (S \setminus S^{-r/2}) \subset S \setminus S^{-r/2} \subset \partial S^{+r/2}$ . To prove the last inclusion let us pick  $x \in S \setminus S^{-r/2}$  as  $\mathcal{B}(x, r/2) \cap S^c \neq \emptyset$ . there exists  $y \in \mathcal{B}(x, r/2) \cap S^c$ . Once again the  $[x, y]$  (of length inferior to  $r/2$ ) intersects the boundary.

2) To achieve the proof we now prove that  $\hat{H}_r^c \cap S^{-r/2} = \emptyset$ .

If there exists a point  $x \in \hat{H}_r^c \cap S^{-r/2}$  Let us first remark that  $x \notin \mathcal{H}(\mathcal{B}(x, r/2) \cap \mathcal{X}_n)$ . This is because if  $x \in \mathcal{H}(\mathcal{B}(x, r/2) \cap \mathcal{X}_n)$ , then there exists at least an observation (let us denote it  $X_i$ ) in  $\mathcal{B}(x, r/2)$  and as  $\mathcal{B}(x, r/2) \subset \mathcal{B}(X_i, r)$  we have  $x \in \mathcal{H}(\mathcal{B}(X_i, r) \cap \mathcal{X}_n)$  so  $x \in \hat{H}_r$ .

Now, as  $x \notin \mathcal{H}(\mathcal{B}(x, r/2) \cap \mathcal{X}_n)$  there exists a unit vector  $u$  such that:  $\forall X_i \in \mathcal{X}_n \cap \mathcal{B}(x, r/2)$  then  $\langle X_i - x, u \rangle \leq 0$ . Let us define the point  $y = x + (r/4)u$  (see Figure 1) it satisfies the three following properties:

i)  $\|x - y\| = r/4$ .

ii)  $\mathcal{B}(y, r/4) \subset \mathcal{B}(x, r/2) \subset S$  (because  $x \in S^{-r/2}$ ) and so  $y \in S$ .

iii)  $\mathcal{B}(y, r/4) \cap \mathcal{X}_n = \emptyset$  and so  $y \in \hat{S}_{r_n/4}^c$ .

Finally  $y \in S \cap \hat{S}_{r/4}^c = \emptyset$ .

□

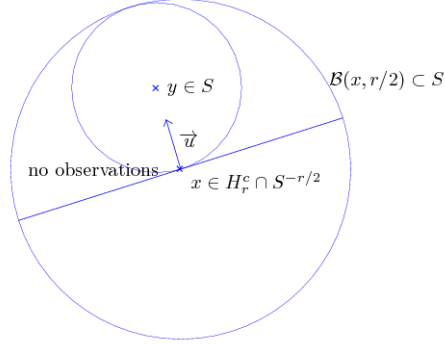


Figure 2:

**Corollary 1.** *If  $S \subset \hat{S}_{r_n/4}$  eventually almost surely then  $(\hat{H}_{r_n} \cap S^c) \cup (\hat{H}_{r_n}^c \cap S) \subset \partial S^{r_n}$  eventually almost surely*

**Lemma 2.** *If  $\partial S$  is  $d'$ -dimensional. If  $r_n$  is a sequence that converges to 0.  $|\partial S^{+r_n}| \leq \lambda(\partial S)\omega_d 2^d r_n^{d-d'}$ .*

*Proof.* Let us cover  $\partial S$  with the minimum number ( $N(\partial S, r_n)$ ) of balls of radius  $r_n$  centered in  $x_i \in \partial S$ . We have

$$\partial S \subset \bigcup_{i=1}^{N(\partial S, r_n)} \mathcal{B}(x_i, r_n)$$

So, using the triangle inequality

$$\partial S^{+r_n} \subset \bigcup_{i=1}^{N(\partial S, r_n)} \mathcal{B}(x_i, 2r_n)$$

So  $|\partial S^{+r_n}| \leq N(\partial S, r_n)\omega_d 2^d r_n^d$ .

And, as  $r_n \rightarrow 0$ , then, for  $n$  large enough  $|\partial S^{+r_n}| \leq \lambda(\partial S)\omega_d 2^d r_n^{d-d'} \rightarrow 0$ .  $\square$

**Corollary 2.** *If  $\partial S$  is  $d'$  dimensional with  $d' < d$ . If  $r_n$  is a sequence that converges to 0 almost surely.  $|\partial S^{+r_n}| \leq \lambda(\partial S)\omega_d 2^d r_n^{d-d'}$  eventually almost surely.*

The point *i*) of Theorem 1 is now a direct corollary of Corollary 1 and 2

Let us now look at the last part of the theorem, i.e. the convergence of the boundary. A direct corollary of Lemma 1 is the following.

**Corollary 3.** *Let  $S \subset \mathbb{R}^d$  be a compact set such that its boundary  $\partial S$  is  $d'$  dimensional ( $d' < d$ ). Let  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  be the set of independent random observations drawn from a distribution  $P_X$  with support  $S$ . Assume that:*

$$S \subset \hat{S}_{r_n/4} \text{ eventually almost surely}$$

$$r_n \rightarrow 0 \text{ almost surely}$$

$$\text{Then } \partial \hat{H}_{r_n} \subset \partial S^{+r_n} \text{ eventually almost surely}$$

To finish the proof of theorem 1 we now only need to prove the following Lemma.

**Lemma 3.** *if  $r_n \rightarrow 0$  almost surely then:*

$$\max_{x \in \partial S} d(x, \partial \hat{H}_{r_n}) \rightarrow 0 \text{ almost surely.}$$

*Proof.* The proof is similar to the proof of the case 2 for Theorem 1 in [CRC04].

Let us proceed by contradiction and suppose that there exists a  $\varepsilon > 0$  and a subsequence  $y_n$  in  $\partial S$  such that  $d(y_n, \partial \hat{H}_{r_n}) > 2\varepsilon$ . The compactness of  $\partial S$  ensures the existence of a subsequence that converges toward  $y \in \partial S$ . As in [CRC04], for  $n$  large enough  $\mathcal{B}(y, \varepsilon) \cap \partial \hat{H}_{r_n} = \emptyset$  eventually with positive probability. As  $\mathcal{B}(y, \varepsilon) \cap \mathcal{X}_n \neq \emptyset$  for  $n$  large enough (with probability 1) (as in “On boundary estimation” Th1 case 2). we have  $\mathcal{B}(y, \varepsilon) \cap \hat{H}_{r_n} \neq \emptyset$  and so  $\mathcal{B}(x, \varepsilon) \subset \text{int}(\hat{H}_{r_n})$ . The natural inclusion  $\hat{H}_{r_n} \subset \hat{S}_{r_n}$  implies that  $\mathcal{B}(x, \varepsilon) \subset \text{int}(\hat{S}_{r_n})$  and that lead to a contradiction exactly as in (as in [CRC04] Th1 case 2).  $\square$

## 3 Convergences rates

### 3.1 Shape conditions

A common hypothesis in the topic of support estimation is that the support is standard with respect to the Lebesgue measure.

**Definition 1.** *A Borel  $S \subset \mathbb{R}^d$  is said to be standard with respect to a Borel measure  $\mu$  if there exists a  $\lambda > 0$  and a  $\delta > 0$  such that:  $\mu(\mathcal{B}(x, \varepsilon)) \geq \delta |\mathcal{B}(x, \varepsilon)|$  for all  $x \in S$  and  $0 < \varepsilon \leq \lambda$*

When we are also interested in the boundary estimation and convergence rates one can also need that the support is partly expendable.



**Definition 2.** A bounded Borel set  $S \subset \mathbb{R}^d$  is said to be partly expandable if there exist constants  $r > 0$  and  $C(S) \geq 1$  such that  $d_H(\partial S, \partial S^{+\varepsilon}) \leq C(S)\varepsilon$  for all  $0 \leq \varepsilon \leq r$ .

In, [HPT95] one can also find the following hypotheses :  $\partial S$  is of regularity  $\gamma$  and the density is  $\alpha$ -quickly decreasing (see Definition 4). Here, we replace the  $\gamma$  regularity of  $\partial S$  by the fact that balls roll freely inside and outside  $S$  (see Definition 3) that is closely related to  $\gamma = 2$  (see [Wal99]).

**Definition 3.** We will say that a ball of radius  $R_{out} > 0$  (resp.  $R_{in} > 0$ ) rolls freely outside (resp. inside)  $S$  if, for all  $x \in \partial S$  there exists a  $O_x^- \in \mathbb{R}^d$  (resp.  $O_x \in \mathbb{R}^d$ ) such that:  $x \in (O_x^-, R_{out}) \subset \overline{S^c}$  (resp.  $x \in (O_x, R_{in}) \subset \overline{S}$ ).

**Definition 4.** A density  $f$  supported by  $S$  is said to be  $\alpha$ -quickly decreasing if there exists a  $\alpha \geq 0$  and a  $f_0 > 0$  such that:  $f(x) \geq f_0 d(x, \partial S)^\alpha$  for all  $x \in S$ .

If balls rolls freely inside and outside  $S$  then  $S$  let us denote has a lot of “good” properties that are proved in [Wal99] (for Properties 1 and 2) and [CRC04] (Property 3) detailed here

**Property 1.** If balls of radius  $R_S$  rolls freely inside and outside  $S$  then  $\partial S$  is a  $(d - 1)$ -dimensional manifold

**Property 2.** If balls of radius  $R_S$  rolls freely inside and outside  $S$  then for all  $x \in \partial S$  there exists a unique inward pointing unit normal vector  $u_x$  and we also have: for all  $(x, y) \in \partial S^2$   $\|u_y - u_x\| \leq \|x - y\|/R_S$

**Property 3.** If balls of radius  $R_S$  rolls freely inside and outside  $S$  then  $S$  is partly expendable and  $C(S) = 1$

**Remark:** In all of the following we will use the notation  $O_x$ ,  $O_x^-$  and  $u_x$  as they are defined in Definitions 3 and Property 2.

## 3.2 Convergence rate under standardness assumption

**Theorem 2.** Let  $X_1, X_2, \dots$  be a sequence of i.i.d. observations drawn from a distribution  $P_X$  on  $\mathbb{R}^d$ . Let denote  $\mathcal{X}_n = \{X_1, \dots, X_n\}$  the set of the  $n$  first observations. Assume that the support  $S$  of  $P_X$  is compact partly expandable (with a constant  $C(S)$ ), standard with respect to  $P_X$  and its boundary  $\partial S$  is  $d'$ -dimensional. Let us consider the Local-Convex-Hull estimator:

$$\hat{H}_{r_n}(\mathcal{X}_n) = \bigcup_{i=1}^n \mathcal{H}(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n)$$

where

$$r_n = c \left( \frac{2 \ln(n)}{\delta \omega_d n} \right)^{1/d} \quad \text{for some } c > 4.$$

then,

$$|S \Delta \hat{H}_{r_n}(\mathcal{X}_n)| \leq \lambda(\partial S) \omega_d 2^d r_n^{d-d'} \quad \text{e.a.s.}$$

$$d_H(\partial S, \partial \hat{H}_{r_n}(\mathcal{X}_n)) \leq c(S) r_n \quad \text{e.a.s.}$$

*Proof.* Theorem 3 in [CRC04] ensures that  $S \subset \hat{S}_{r_n/4}$  eventually almost surely. So we can apply Theorem 1 and have that :

$$|S \Delta \hat{H}_{r_n}(\mathcal{X}_n)| \leq \lambda(\partial S) \omega_d 2^d r_n^{d-d'} \quad \text{e.a.s.}$$

**Remarks:**

- 1) When only focusing on the support estimation we do not need the partly expendable hypothesis.
- 2) In the most regular case when  $d' = d - 1$  we obtain the usual convergence rate  $r_n$ .

Within the application of Theorem 1 we have  $\partial H_{r_n} \subset \partial S^{r_n}$  and using the partly expendable property assumption we have  $\max_{x \in \partial \hat{H}_{r_n}} (\min_{y \in \partial S} (\|x - y\|)) \leq c(S) r_n$ .

Let us now consider  $x \in \partial S$ . If  $x \in \partial S \cap \hat{H}_{r_n}^c$ , Theorem 3 in [CRC04] ensures that there exists (eventually almost surely) an observation  $X_i$  in  $\mathcal{X}_n$  such that  $\|X_i - x\| \leq r_n/4$  and so  $d(x, \partial \hat{H}_{r_n}) \leq r_n/4$ . If  $x \in \partial S \cap \hat{H}_{r_n}$  then  $x \in \hat{H}_{r_n} \subset S^{r_n}$  and so  $d(x, \partial H_{r_n}) \leq r_n$ .

We so have  $d_H(\partial S, \partial \hat{H}_{r_n}(\mathcal{X}_n)) \leq \max(0.25, 1, c(S) r_n)$  e.a.s. and the fact that  $c(S) \geq 1$  achieve the proof.  $\square$

### 3.3 Convergence rate under stronger shape condition

In this part we will assume that balls rolls freely inside and outside  $S$  and denote  $R_S = \min(R_{\text{out}}, R_{\text{in}})$  we will also assume that the density is  $\alpha$ -quickly decreasing.

In a certain way this set of assumption is stronger on the support shape but less restrictive for the density.

Under such assumptions we have asymptotic results given Theorem 3 that shows that the specific choice of sequences  $r_n$  proportional to  $\left(\frac{\ln n}{n}\right)^{\frac{1}{d+1+2\alpha}}$  provides a support estimator  $\hat{H}_{r_n}$  that converges (for the symmetric difference measure) with a rate close to the optimal one with similar assumptions ([HPT95]).

**Theorem 3.** *Let  $X_1, X_2, \dots$  be a sequence of i.i.d. observations drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls of radius  $R_S$  roll freely inside and outside the support  $S$  of  $P_X$ . If  $r_n = \lambda(\ln n/n)^{1/(d+1+2\alpha)}$  then:*

$$|\hat{H}_{r_n} \Delta S|(n/\ln n)^{2/(d+1+2\alpha)} \text{ is e.a.s. bounded}$$

$$d_H(\partial \hat{H}_{r_n}, \partial S)(n/\ln n)^{2/(d+1+2\alpha)} \text{ is e.a.s. bounded}$$

The following of the section is dedicated to the proof of Theorem 3. As usual it is divided into two parts: the study of the asymptotic behavior of  $\hat{H}_r \cap S^c$  and the one of  $\hat{H}_r^c \cap S$ . The section 3.4 is devoted to prove the following Lemma about the asymptotic behavior of  $\hat{H}_{r_n} \cap S^c$ , it is an easy and deterministic result with no assumption on  $f$ .

**Lemma 4.** *Let  $X_1, X_2, \dots$  be a sequence of i.i.d. observations drawn from a distribution  $P_X$ . Assume that balls of radius  $R_S$  roll freely inside and outside the support  $S$  of  $P_X$ .*

*When  $r_n = \left(\lambda \frac{\ln n}{n}\right)^{\frac{1}{d+1+2\alpha}}$  For  $n$  large enough to have  $r_n < R_S/4$*

$$\hat{H}_{r_n} \cap S^c \subset \partial S + \frac{2r_n^2}{R_S}(1+o(1))$$

The section 3.5 is devoted to prove the following Lemma about the asymptotic behavior of  $\hat{H}_{r_n}^c \cap S$ , here we have an asymptotically and probabilistic result.

**Lemma 5.** *Let  $X_1, X_2, \dots$  be a sequence of i.i.d. observations drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls of radius  $R_S$  roll freely inside and outside the support  $S$  of  $P_X$ .*

*When  $r_n = \left(\lambda \frac{\ln n}{n}\right)^{\frac{1}{d+1+2\alpha}}$ , there exists an explicit constant  $b(\lambda)$ :*

$$b(\lambda) = \frac{19}{32R_S} + \frac{1}{2} \left( \frac{(3d-1+2\alpha)(\alpha+1)4^{d+1+2\alpha}}{(d+1+2\alpha)\lambda f_0 \omega_d} \right)^{1/(\alpha+1)}$$

*such that:*

$$\hat{H}_{r_n}^c \cap S \subset \partial S^{b(\lambda)r_n^2} \text{ eventually almost surely}$$

And Section 3.6 is devoted to prove Theorem 3 using Lemmas 4 and 5.

### 3.4 Asymptotic behavior $\hat{H}_r \cap S^c$

Lemma 4 is a direct corollary of the following result.

**Lemma 6.** *Let  $X_1, X_2, \dots$  be a sequence of i.i.d. observations drawn from a distribution  $P_X$ . Assume that balls of radius  $R_S$  roll freely inside and outside the support  $S$  of  $P_X$ . For all  $r < R_S/4$  we have:*

$$\text{if } x \in \hat{H}_r \cap S^c \text{ then } d(x, \partial S) \leq R_S - \sqrt{R_S^2 - 4r^2}.$$

*Proof.* If  $x \in \hat{H}_r$  then there exists a  $X_i \in \mathcal{X}_n$  and  $d + 1$  observations  $X_{i_1}, \dots, X_{i_{d+1}}$  in  $\mathcal{B}(X_i, r)$  such that  $x \in \mathcal{H}(\{X_{i_1}, \dots, X_{i_{d+1}}\})$ . The distance between two  $X_{i_j}$  is less than  $2r$  (because they are in  $\mathcal{B}(X_i, r)$ ) and so the distance between  $x$  and any  $X_{i_j}$  is also bounded by  $2r$  (because  $x$  is in the convex hull of the points). So  $x \in \mathcal{H}(\mathcal{B}(x, 2r) \cap S)$  that is impossible when  $d(x, \partial S) > R_S - \sqrt{R_S^2 - 4r^2}$  (see Figure 3 to be convinced).

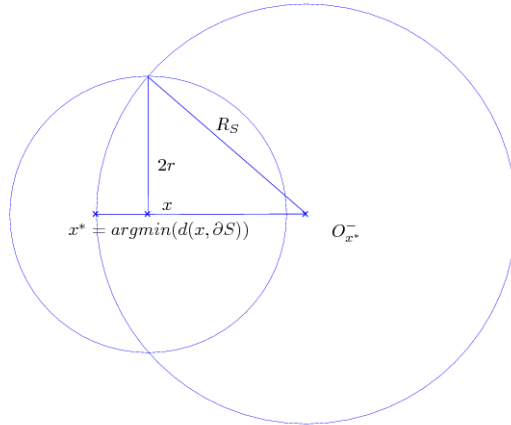


Figure 3:

□

### 3.5 Asymptotic Behavior of $S \cap \hat{H}_r^c$

In the standard case or usual proofs for the Devroye-Wise estimator, results are obtained saying that a “large” distance imply existence of a ball of “large” radius that does not contain any observation. Here we will consider empty balls when points are “far” from the boundary but replace balls by tangent cylinders when considering points close to the boundary.

### 3.5.1 Some results about empty balls

**Theorem 4.** Let  $X_1, X_2, \dots$  be a sequence of i.i.d. observations drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls rolls freely inside and outside the support  $S$  of  $P_X$  then:

$$\limsup_{n \rightarrow \infty} \left( \frac{n}{\ln n} \right)^{1/(d+\alpha)} d(\mathcal{X}_n, S) \leq \left( \frac{2}{f_0 \omega_{d-1} B(\alpha+1, d)} \right)^{1/d+\alpha} \text{ almost surely.}$$

Where  $B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$  is the Beta function.

Proof is given in appendix.

### 3.5.2 Some results about empty cylinders

**Definition 5.** Let  $u$  be a unit vector we can define the cylinder:  $\mathcal{C}^u(x, r, h) = \{y, | \langle y - x, u \rangle | \leq h, \|y - x - \langle y - x, u \rangle u\| \leq r\}$

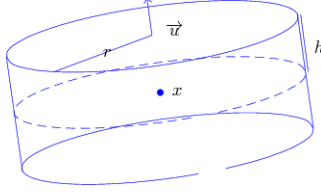


Figure 4:

**Property 4.** Let  $x$  and  $x'$  be two points and  $u$  and  $u'$  and two unit vectors. Let us denote  $\|x - x'\| = \varepsilon_x$ ,  $\|u - u'\| = \varepsilon_u$ . Let us define:

$$e_1 = \varepsilon_x + 2(\sqrt{r^2 + h^2} + \varepsilon_x)\varepsilon_u$$

$$e_2 = \varepsilon_x + \sqrt{r^2 + h^2}\varepsilon_u$$

When  $\varepsilon_x$  and  $\varepsilon_u$  are small enough to have  $e_1 \leq r$  and  $e_2 \leq h$ , we have:

$$\mathcal{C}^{u'}(x', r - e_1, h - e_2) \subset \mathcal{C}^u(x, r, h)$$

proof is let to the reader.

**Definition 6.** Tangent cylinders: For all  $x \in \partial S$  we can define  $\mathcal{C}(x, r, h) = \mathcal{C}^{u_x}(x, r, h)$

**Lemma 7.** *if  $X_1, \dots, X_n$  are drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls of radius  $R_S$  rolls freely inside and outside the support  $S$  of  $P_X$ .*

*For any  $\lambda > 0$ ,  $\mu > 1$  such that  $\frac{\lambda f_{0\omega_{d-1}}(\mu-1)^{\alpha+1}}{(\alpha+1)R_S^{\alpha+1}} - \frac{2d-2}{d+1+2\alpha} = c > 1$  let us denote  $r_n$  and  $h_n$  the sequences:  $r_n = \left(\lambda \frac{\ln n}{n}\right)^{\frac{1}{d+1+2\alpha}} (1 + o(1))$  and  $h_n = \mu \frac{r_n^2}{2R_S} (1 + o(1))$ .*

*We have:*

*“for all  $x \in \partial S$ ,  $\mathcal{C}(x, r_n, h_n) \cap \mathcal{X}_n \neq \emptyset$ ” eventually almost surely.*

Proof is given in appendix.

### 3.5.3 Proof of Lemma 5

We now have all the tools to prove the Lemma 5.

*Proof.* Let us suppose that  $\hat{H}_{r_n}^c \cap S \not\subseteq \partial S^{br_n^2}$ . There exists a point  $x \in \hat{H}_{r_n}^c \cap S$  such that  $d(x, \partial S) = d_n > br_n^2$ . Let us denote  $y^*$  the point of  $\partial S$  that realizes  $d(x, \partial S) = d_n$ . As  $x \in \hat{H}_{r_n}^c$  there exists a unit vector  $u$  such that for all  $X_i$  in  $\mathcal{B}(x, r_n/2) \cap \mathcal{X}_n$ ,  $\langle u, X_i - x \rangle \geq 0$  (see the proof of Theorem 1).

Let us denote  $\rho_n = \left(\frac{2 \ln n}{f_{0\omega_{d-1}} B(\alpha+1, d)n}\right)^{\frac{1}{d+\alpha}}$  (notice that  $\rho_n \ll r_n$ ). We are going to prove that, if  $d_n \geq 3\rho_n$  then there exists a “too large” ball that does not contains any observation, and, if  $d_n \leq 3\rho_n$  then there exists a “too large” tangent cylinder that does not contains any observation.

**First case: if  $d_n > 3\rho_n$ .** First notice that :  $d(x, \partial S) = d_n \Rightarrow \mathcal{B}(x, d_n) \subset S \Rightarrow \mathcal{B}(x, 3\rho_n) \subset S$ . Let us define  $z^* = x - (3\rho_n/2)u$ : as  $\|x - z^*\| = 3\rho_n/2$  we have  $z^* \in S$ . For  $n$  large enough to have  $3\rho_n \leq r_n/2$  we also have  $\mathcal{B}(z^*, 3\rho_n/2) \cap \mathcal{X}_n = \emptyset$ . That is impossible (eventually almost surely) according to Lemma 4 to Lemma 4.

**Second case: if  $d_n \leq 3\rho_n$**

We are going to distinguish two cases:  $u = u_{y^*}$  (see figure 5) or not (see figure 6). Each time we exhibit cylinders that does not contain any observation.

**First sub-case:  $u = u_{y^*}$ .** See Figure 5 to be convinced that:

$$\mathcal{C}(y^*, \sqrt{r_n^2/4 - 4d_n^2}, d_n) \cap \mathcal{X}_n = \emptyset$$

The upper bound  $d_n \leq 3\rho_n$  allows us to write:

$$\mathcal{C}(y^*, \sqrt{r_n^2/4 - 4d_n^2}, d_n) = \mathcal{C}(y^*, (r_n/2)(1 + o(1)), d_n)$$

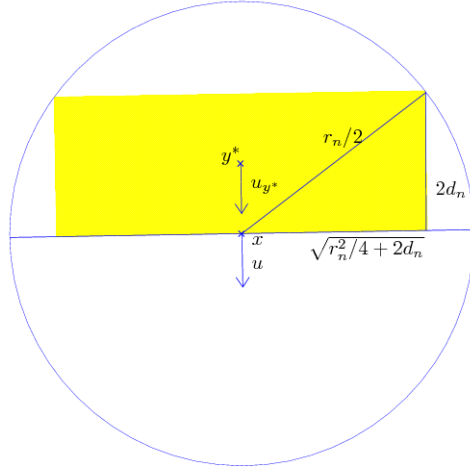


Figure 5: empty cylinder construction when  $u = u_{y^*}$

The Lower bound  $d_n > br_n^2$  allows us to write:

$$\mathcal{C}(y^*, (r_n/2)(1 + o(1)), br_n^2) \cap \mathcal{X}_n = \emptyset$$

**Second sub-case:**  $u \neq u_{y^*}$ .

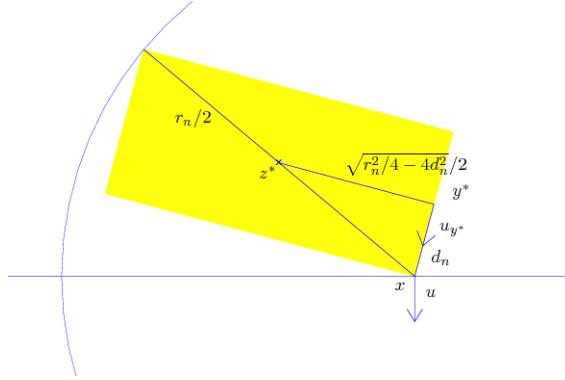


Figure 6: empty cylinder construction when  $u \neq u_{y^*}$

Let us define  $w = u - \langle u, u' \rangle u'$ ,  $v = w/||w||$  and  $z^* = y^* + (\sqrt{r_n^2/4 - 4d_n^2}/2)v$ . See Figure 6 (projection in the plan  $(x, u, u_{y^*})$ ) to be convinced that we have:

$$\mathcal{C}^{u_{y^*}}(z^*, \sqrt{r_n^2/4 - 4d_n^2}/2, d_n) \cap \mathcal{X}_n = \emptyset$$

The upper bound  $d_n \leq 3\rho_n$  allows us to write

$$\mathcal{C}^{u_{y^*}}(z^*, (r_n/4)(1 + o(1)), d_n) \cap \mathcal{X}_n = \emptyset$$

The upper bound  $d_n > br_n^2$  allows us to write

$$\mathcal{C}^{u_{y^*}}(z^*, (r_n/4)(1 + o(1)), br_n^2) \cap \mathcal{X}_n = \emptyset$$

We have not yet finish the proof because the exhibited cylinder is not tangent cylinder. The rolling balls properties implies (see Figure 7) that  $d(z^*, \partial S) \leq \alpha_n = -R_S + \sqrt{R_S^2 + r_n^2/16} \sim r_n^2/(32R_S)$

Let us chose  $z \in \partial S$  such that  $d(z, z^*) \leq \alpha_n$  we have  $d(z, y^*) \leq \alpha_n + r_n/4$

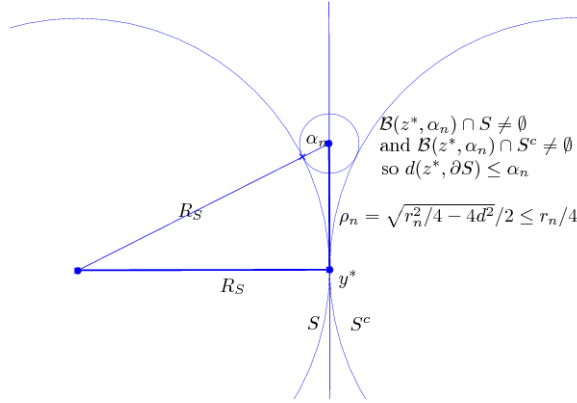


Figure 7: distance to the boundary

so  $\|u_z - u_{y^*}\| \leq (r_n/4 + \alpha_n)/R_S$  (according to Property 2). so, applying Property 4

$$\mathcal{C}\left(z, \frac{r_n}{4}(1 + o(1)), \left(b - \frac{3}{32R_S}\right)r_n^2(1 + o(1))\right) \cap \mathcal{X}_n = \emptyset$$

In both cases ( $u = u'$  or  $u \neq u'$ ), if  $d_n \in [br_n^2, 5\rho_n]$  there exists a  $z \in \partial S$  such that:

$$\mathcal{C}\left(z, \frac{r_n}{4}(1 + o(1)), \left(b - \frac{3}{32R_S}\right)r_n^2(1 + o(1))\right) \cap \mathcal{X}_n = \emptyset$$

Lemma 9 ensures that  $d_n \in [br_n^2, 5\rho_n]$  is not possible (eventually almost surely) when:

$$b > \frac{19}{32R_S} + \frac{1}{2} \left( \frac{(3d - 1 + 2\alpha)(\alpha + 1)4^{d+1+2\alpha}}{(d + 1 + 2\alpha)\lambda f_0 \omega_d} \right)^{1/(\alpha+1)}$$

□



### 3.6 proof of Theorem 3

A direct consequence of Lemmas 4 and 5 is that  $H_r \Delta S \subset \partial S^{ar_n^2(1+o(1))}$  with  $a = \max(2/R_S, b(\lambda))$ . Property 1 ensures that  $\partial S$  is a  $(d-1)$ -manifold so, exactly as in the proof of Theorem 1 we have that  $|\partial S^{ar_n^2(1+o(1))}| \leq \lambda(\partial S)\omega_d 2^d ar_n^2(1+o(1))$  that proves the first point of Theorem 3.

$H_r \Delta S \subset \partial S^{ar_n^2(1+o(1))} \Rightarrow \partial H_{r_n} \subset \partial S^{ar_n^2(1+o(1))}$  and so Property 3 allows us to have  $\max_{x \in \partial H_{r_n}}(d(x, \partial S)) \leq ar_n^2(1+o(1))$ . The rolling balls property and  $H_r \Delta S \subset \partial S^{ar_n^2(1+o(1))}$  also imply that,

- 1) for all  $x \in \partial S \cap H_{r_n}^c$ ,  $x + ar_n^2(1+o(1))u \in H_{r_n}$  (eventually almost surely),  
so for all  $x \in \partial S \cap H_{r_n}^c$ ,  $d(x, \partial H_{r_n}) \leq ar_n^2(1+o(1))$ .
- 2) for all  $x \in \partial S \cap H_{r_n}$ ,  $x - ar_n^2(1+o(1))u \in H_{r_n}^c$  (eventually almost surely),  
so for all  $x \in \partial S \cap H_{r_n}$ ,  $d(x, \partial H_{r_n}) \leq ar_n^2(1+o(1))$ .

So we also have:  $\max_{x \in \partial S}(d(x, \partial H_{r_n})) \leq ar_n^2(1+o(1))$ . And  $d_H(H_{r_n}, S) \leq ar_n^2(1+o(1))$ .

## 4 Preservation of the topology

In this section we are interest in the recognition of the topological properties of the support. In other words we are going to prove that, when balls roll freely inside and outside  $S$ , and when the density is  $\alpha$ -quickly decreasing then the choice of a radius sequence  $r_n = \lambda \left(\frac{\ln n}{n}\right)^{2/(d+1+2\alpha)}$  gives support estimator that is (eventually almost surely) homeomorph to the unknown support.

**Theorem 5.** *Let  $X_1, X_2, \dots$  be a sequence of i.i.d. observations drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls of radius  $R_S$  roll freely inside and outside the support  $S$  of  $P_X$ . When  $r_n = \lambda(\ln n/n)^{1/(d+1+2\alpha)}$ :*

*“ $\partial \hat{H}_{r_n}$  is homeomorph to  $\partial \hat{S}$ ” eventually almost surely.*

*“ $\hat{H}_{r_n}$  is homeomorph to  $\hat{S}$ ” eventually almost surely.*

*Proof.* We first prove the announced result for the boundary, then for the support.

When, focusing on the boundary we are going to prove that we can (eventually almost surely) define uniquely the function  $\varphi_{r_n} : \partial \hat{H}_{r_n} \rightarrow \partial S$  such that  $\|\varphi(x) - x\| = \min_{y \in \partial S} \|x - y\|$ . And this function is (eventually almost surely) a continuous bijection.

**Definition:**

First let us notice that balls rolls freely inside and outside  $S$  implies for any  $x$  such that  $d(x, \partial S) < R_S$  there exists a unique  $y \in \partial S$  that realizes  $\min \|x - y\|$  and as Theorem 5 ensures that for  $n$  large enough  $\partial \hat{H}_{r_n} \subset \partial S^{R_S/2}$  (eventually almost surely) then the function  $\varphi_{r_n}$  is uniquely defined (eventually almost surely).

**Continuity:**

Let us define  $\varphi : \partial S^{R_S/2} \rightarrow \partial S$  such that  $\|\varphi(x) - x\| = \min_{y \in \partial S} \|x - y\|$  and let us prove that  $\varphi$  is continuous.

For all  $x$  with  $d(x, \partial S) < R_S$ . For all  $\varepsilon \leq R_S - d(x, \partial S)$  for  $x' \in \mathcal{B}(x, \varepsilon)$ , we have :

- 1)  $\varphi(x') \in \mathcal{B}(\varphi(x), \|\varphi(x) - x'\|)$  (because  $\min d(x', \partial S) \geq \|x' - \varphi(x)\|$ )
- 2) and  $\varphi(x') \notin \mathring{\mathcal{B}}(O_{\varphi(x)}, R_S)$  (because  $\mathring{\mathcal{B}}(O_{\varphi(x)}, R_S) \subset \mathring{S}$ )

That implies that:  $\|\varphi(x') - \varphi(x)\| \leq \frac{4\varepsilon R_S}{R_S - d(x, \partial S)}$  (see Figure 4). So  $\varphi$  is continuous and so does  $\varphi_{r_n}$  its restriction on the boundary of the estimator.

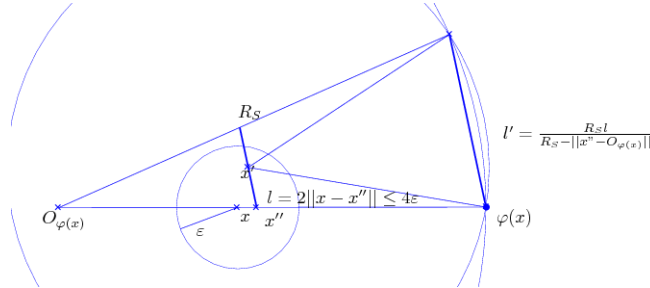


Figure 8:

**Surjectivity:**

Theorem 3 implies that there exists a  $a$  such that  $H_{r_n} \Delta S \subset \partial S^{+ar_n^2}$  (eventually almost surely). For any  $x \in \partial S \cap \hat{H}_{r_n}^c$  let us define  $y = x + ar_n^2 u_x$  we have  $y \in \hat{H}_{r_n}$  eventually almost surely and so  $[x, y]$  crosses  $\partial \hat{H}_{r_n}$  at a point  $x^*$  that satisfies  $\varphi_{r_n}(x^*) = x$  For any  $x \in \partial S \cap \hat{H}_{r_n}$  let us define  $y = x - ar_n^2 u_x$  we have  $y \in \hat{H}_{r_n}^c$  eventually almost surely and so  $[x, y]$  crosses  $\partial \hat{H}_{r_n}$  at a point  $x^*$  that satisfies  $\varphi_{r_n}(x^*) = x$ .

**Injectivity:**

The injectivity is the most difficult point.

Here we denote:  $\rho_n = 2 \left( \frac{2 \ln n}{f_0 \omega_{d-1} B(\alpha+1, d)n} \right)^{\frac{1}{d+\alpha}}$  and  $\varepsilon_n = ar_n^2$  (with  $a$  the suitable value such that  $\partial H_{r_n} \subset \partial S^{ar_n^2(1+o(1))}$ ).

Let us first notice that  $\partial \hat{H}_{r_n}$  is a random polytope with a finite number of faces. Let us suppose that  $\varphi_{r_n}$  is not injective and exhibit  $x$  and  $y$  such that  $\varphi(x) = \varphi(y) = z$  and  $\langle x - y, u_z \rangle \gg 0$ .

The first case is the following:  $[x, y] \cap \partial \hat{H}_r \neq [x, y]$ , as  $\partial \hat{H}_{r_n}$  is a random polytope with a finite number of faces one can suppose that  $[x, y] \cap \partial \hat{H}_{r_n} = \{x, y\}$ . Let us define  $x^*$  as follows: if  $]x, y[ \subset \hat{H}_r$  let us choose  $x^* = x$ , If  $]x, y[ \subset \hat{H}_r^c$  let us choose  $x^* = y$ .  $x^* \in \partial \hat{H}_{r_n}$  so there exists an observation  $X_i$  such that  $x^* \in \partial \mathcal{H}(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n)$  and, the way we define  $x^*$  implies that there exists a vector  $u$ , such that  $\langle u, u_z \rangle \gg 0$  and such that for all observation  $X$  in  $(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n) \cap \{x^* - X, u \rangle \leq 0$ .

The second case is  $[x, y] \cap \partial \hat{H}_r = [x, y]$  and there exists a vector  $u$  orthogonal to  $y - x$  such that for all observation  $X$  in  $(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n) \cap \{x - X, u \rangle \leq 0$ .

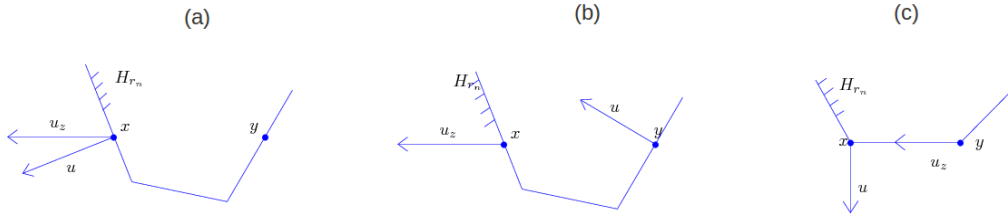


Figure 9:

For both cases there exists  $x^* \in \partial \hat{H}_{r_n}$ ,  $X_i \in \mathcal{X}_n$  and  $u$  with  $\langle u, u_z \rangle \gg 0$  such that  $x^* \in \partial \mathcal{H}(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n)$  and for all observation  $X$  in  $(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n) \cap \{x^* - X, u \rangle \leq 0$ . As  $x^* \in \partial \hat{H}_{r_n}$  we also have that for all  $X \in \mathcal{X}_n$   $x^* \notin \overset{\circ}{\mathcal{H}}(X, r_n)$ .

Let  $X_i$  be an observation such that  $x^* \in \partial \mathcal{H}(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n)$ . If  $\langle x^* - X_i, u \rangle \leq r_n - 2\rho_n$ , let us define  $y^* = X_i + (r_n - \rho_n)u$ , we have  $\mathcal{B}(y, \rho_n) \subset \{z \in \mathcal{B}(X_i, r_n), \langle x^* - z, u \rangle \gg 0\}$  and so  $\mathcal{B}(y, \rho_n) \cap \mathcal{X}_n = \emptyset$

According to Theorem 3  $\|x^* - z\| \leq \varepsilon_n$  (e.a.s.), so  $\|X_i - z\| \leq r_n + \varepsilon_n$ . We also have:

- 1)  $X_i \in \mathcal{B}^c(O_z^-, R_S)$  (because  $X_i \in S$ )
- 2)  $\|X_i - y^*\| = r_n - 2\rho_n \leq r_n$

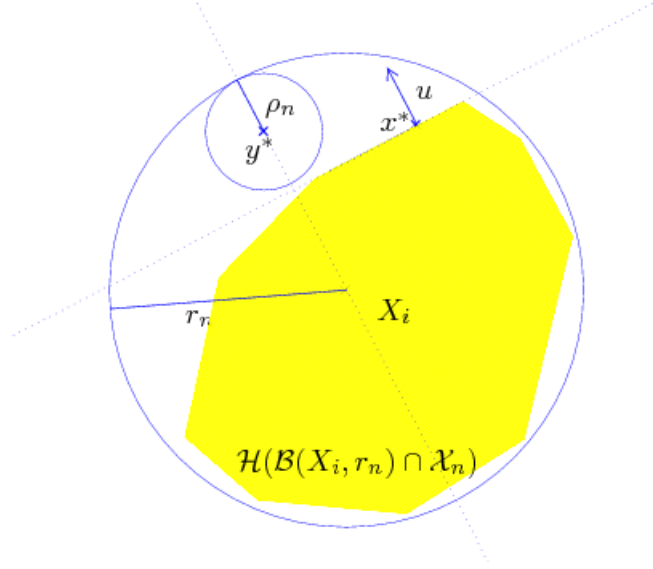


Figure 10:

- 3) For  $n$  large enough  $\langle y^* - X_i, u_z \rangle \geq 0$  (because  $y^* - X_i = (r_n - \rho_n)u$ , for  $n$  large enough  $r_n - \rho_n > 0$  and  $\langle u, u_z \rangle \geq 0$ ).

This three points together imply that  $\|y^* - O_z\| \leq e_n \sim \frac{5r_n^2}{2R_S}$  (e.a.s.) (see Figure 12) so  $d(y^*, S) \leq \sim \frac{5r_n^2}{2R_S}$  (e.a.s.) and finally we can (e.a.s.) find a point  $\hat{y} \in S$  such that  $\mathcal{B}(\hat{y}, \rho_n - e_n) \cap \mathcal{X}_n = \emptyset$  that is (e.a.s.) impossible according to Lemma 4 for the given values of  $r_n$  and  $\rho_n$ .

As a first conclusion we now have: For every observation  $X_i$  such that  $x^* \in \partial\mathcal{H}(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n)$ ,  $\langle x^* - X_i, u \rangle \geq r_n - 2\rho$  (e.a.s.) (\*).

Let us choose a  $X_i$  such that  $x^* \in \partial\mathcal{H}(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n)$ . As  $x^*$  is in  $\partial\mathcal{H}(\mathcal{B}(X_i, r_n) \cap \mathcal{X}_n)$  there exists  $k$  observations ( $1 \leq k \leq d$ )  $X_{(1)}, \dots, X_{(k)}$  in  $\mathcal{B}(X_i, r)$  such that  $x^* \in \mathcal{H}(\{X_{(1)}, \dots, X_{(k)}\})$ . As we also have, for all observation  $X$  in  $\mathcal{B}(X_i, r_n)$ ,  $\langle X - x^*, u \rangle \leq 0$  we have that  $X_{(1)}, \dots, X_{(k)}$  also belong to the plan  $\{z, \langle x^* - z, u \rangle = 0\}$  so  $\|x^* - X_{(i)}\| \leq 2\sqrt{\rho r_n - \rho_n^2}$  and for all  $i, j$   $\|X_{(i)} - X_{(j)}\| \leq 2\sqrt{\rho_n r_n - \rho_n^2}$ . For  $n$  large enough such that  $2\sqrt{\rho r_n - \rho_n^2} \leq r_n - 2\rho_n$  we have  $x^* \in \mathcal{H}(\mathcal{B}(X_{(1)}, r_n) \cap \mathcal{X}_n)$  with  $\|x^* - X_{(1)}\| < r_n - 2\rho_n$  that is impossible according to (\*).

We now have proved the announced result for the boundary.

We can now construct a function  $\psi : S \rightarrow H_{r_n}$  as follows:

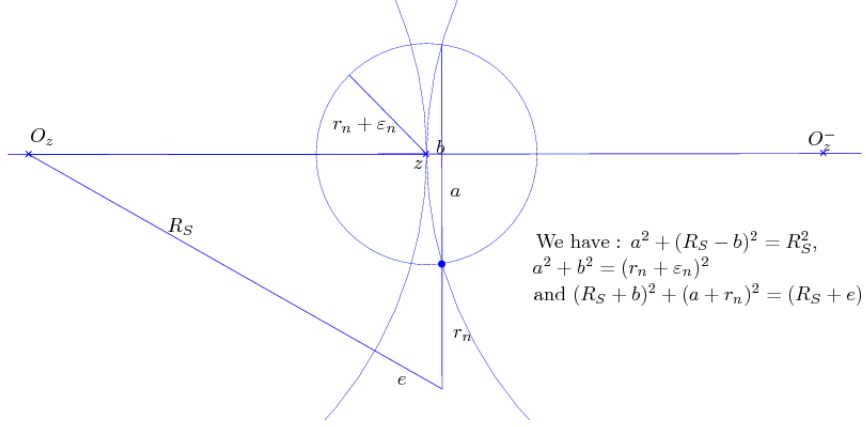


Figure 11:

- 1) if  $x \in S^{-R_S/2}$  then  $\psi(x) = x$
- 2) if  $x \in S \setminus S^{-R_S/2}$  let us define:
  - a)  $g(x) = \operatorname{argmin}(d(x, \partial S))$
  - b)  $x' = g(x) - \frac{R_S}{2}u_{g(x)}$  (let us remark that the rolling ball property implies that  $x' \in \partial(S^{-R_S/2})$ )
  - c)  $z = \varphi_{r_n}^{-1}(g(x))$
$$\psi(x) = x' - \frac{2\|x'-z\|\cdot\|x'-x\|}{R_S}u_{g(x)}$$

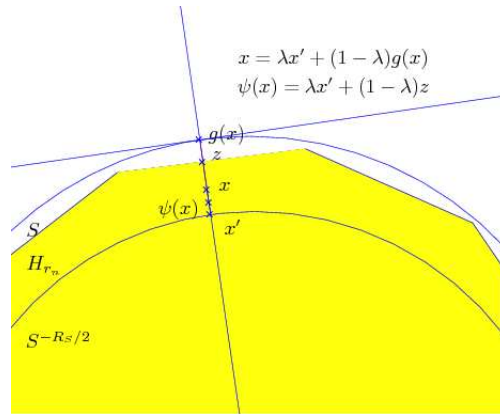


Figure 12: Construction for  $\varphi$

The function  $\psi$  is continuous on  $S \setminus S^{-R_S/2}$  (eventually almost surely) because  $\varphi_{r_n}^{-1}$  is continuous (eventually almost surely),  $g(x)$  is continuous and  $u_x$  is continuous (Property 2). The continuity on  $S$  is easily obtained (basically if  $x_n \in S \setminus S^{-R_S/2}$  is such that  $x_n \rightarrow x \in \partial(S^{-R_S/2})$  then  $x'_n \rightarrow x$  and so  $\psi(x_n) \rightarrow x$ )

It is also easy to see that its is bijective because its reverse function is the following:

- 1) if  $x \in S^{-R_S/2}$  then  $\psi^{-1}(x) = x$
- 2) if  $x \in \hat{H}_{r_n} \setminus S^{-R_S/2}$  let us define:
  - a)  $h(x) = \operatorname{argmin}(d(x, \partial\hat{H}_{r_n}))$
  - b)  $x'' = \varphi_{r_n}(h(x))$
  - c)  $x' = x'' - \frac{R_S}{2}u_{x''}$
$$\psi(x) = x' - \frac{\|x-x'\|R_S}{\|x''-x'\|^2}u_{x''}$$

And so we build an homeomorphism from  $S$  to  $H_{r_n}$ .

□

## 5 some simulations

4 simulated simulations on “toy” examples drawn as follows.

**Figure 13:** 500 realizations in  $\mathcal{B}(0, 1)$  of  $X = (r \cos \theta, r \sin \theta)$  with  $\theta \rightsquigarrow \mathcal{U}([0, 2\pi])$  and  $r \rightsquigarrow \mathcal{U}([0, 1])$ .

**Figure 14:** 500 realizations in  $\mathcal{B}(0, 1) \setminus \mathcal{B}(0, 0.5)$  of  $X = (r \cos \theta, r \sin \theta)$  with  $\theta \rightsquigarrow \mathcal{U}([0, 2\pi])$  and  $r \rightsquigarrow \mathcal{U}([0.5, 1])$ .

**Figure 15:** 500 realizations in  $\mathcal{B}(0, 1) \setminus \mathcal{B}(0, 0.9)$  of  $X = (r \cos \theta, r \sin \theta)$  with  $\theta \rightsquigarrow \mathcal{U}([0, 2\pi])$  and  $r \rightsquigarrow \mathcal{U}([0.9, 1])$ .

**Figure 16:** 500 realizations of an uniform distribution on the star shape  $S = [-1, 1]^2 \setminus (\cup_i \mathcal{B}(C_i, 1))$ ,  $C_1 = (-1, -1)$ ,  $C_2 = (-1, 1)$ ,  $C_3 = (1, -1)$ ,  $C_4 = (1, 1)$ .

**Figure 18:** 500 realizations of distribution on a asterisk shape with the following distribution  $k \in \{0, 1, 2, 3\}$   $P(\theta = k\pi/4) = 1/4$ ,  $(x, y) \rightsquigarrow \mathcal{U}([-1, 1] \times [-0.05, 0.05])$ ,  $X = (\cos(\theta)x + \sin(\theta)y, \cos(\theta)y - \sin(\theta)x)$ .

For each simulation it is first presented the estimated (Monte Carlo) measure of the symmetric difference for various values of the radius  $r$  (the asterisked point on this plot locates the best radius for the result for  $r^* = \max_{x \in S} d(x, \mathcal{X}_n)$ ). Then we plot the Devroye Wise (DW) estimator for the best radius according to the measure of the symmetric difference (Monte Carlo) and for the best radius according to the boundary estimation i.e.  $r^*$  ([CRC04]). Finally we plot the Local convex hull estimator, once again for the best radius value.

It can be notice that, has announced in the introduction, the Devroye Wises estimator with the best radius according to the support estimation is not accurate to estimate the boundary and the topological properties, and the the Devroye Wises estimator with the best radius according to the boundary estimation overfills a lot the support. The Local Convex Hull estimator, with the best radius according to the support estimation over performs the support estimation and the boundary estimation. Even when the support does not fullfills the rolling balls property we always obtain estimated support homeomorph to the support.

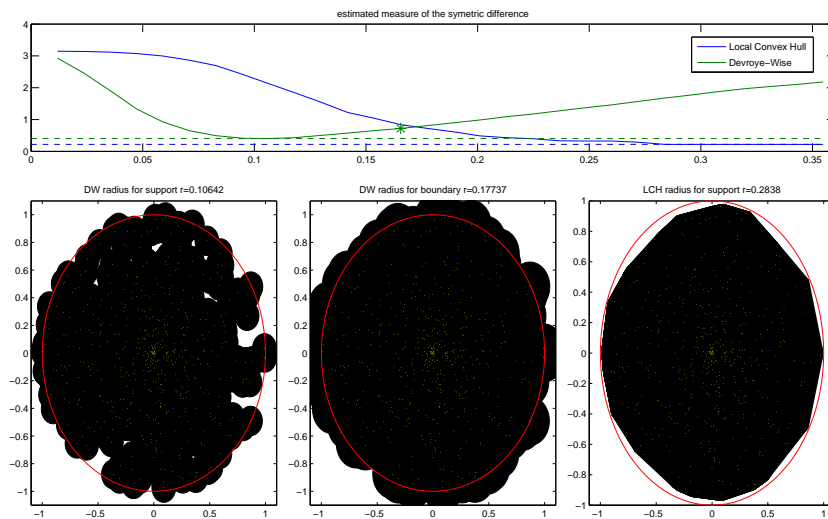


Figure 13: Results for a disc

Finally we applied the method to a set of 5323 locations of epicenters of earthquakes of magnitudes superior to 6. We changed a little the way to fix the neighborhood in a matter to take into account the spherical coordinates and obtain the following map for the earthquakes.

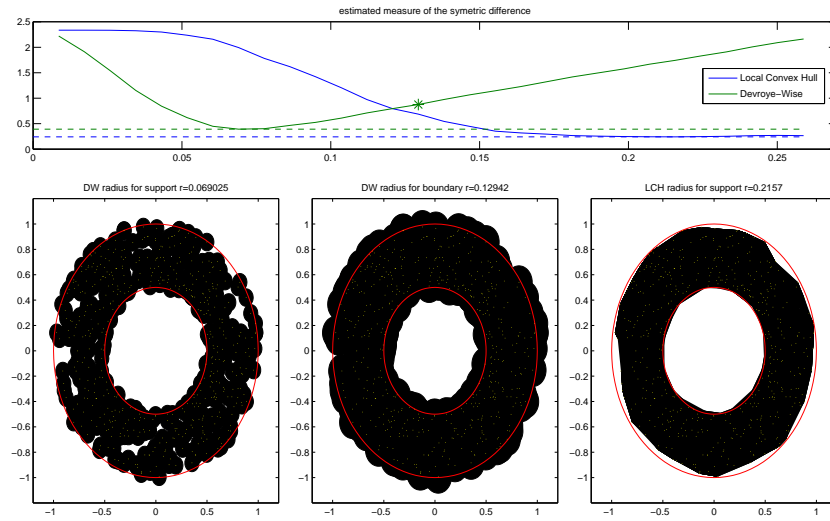


Figure 14: Result for a  $CD$  shape

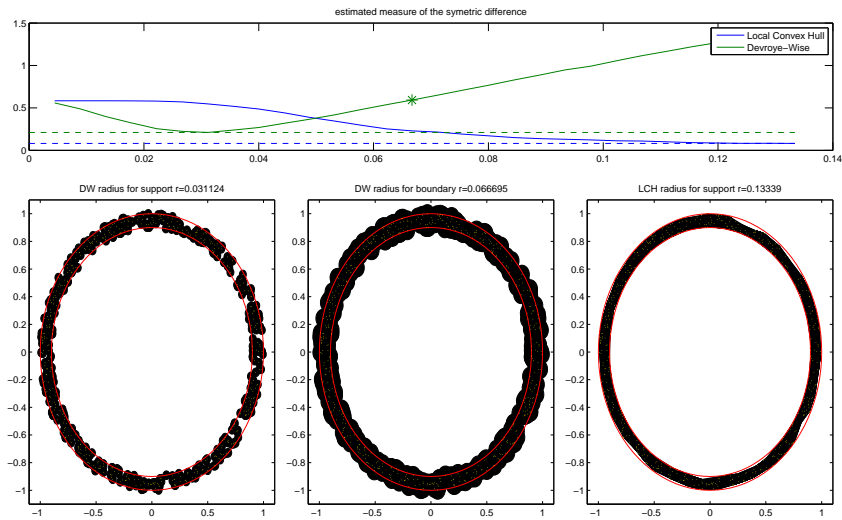


Figure 15: Result for a  $CD$  shape

## 6 conclusion and further works

The local convex hull method allows us to estimate the support, its boundary (with a rate close to the optimal one), and to recover its topology with



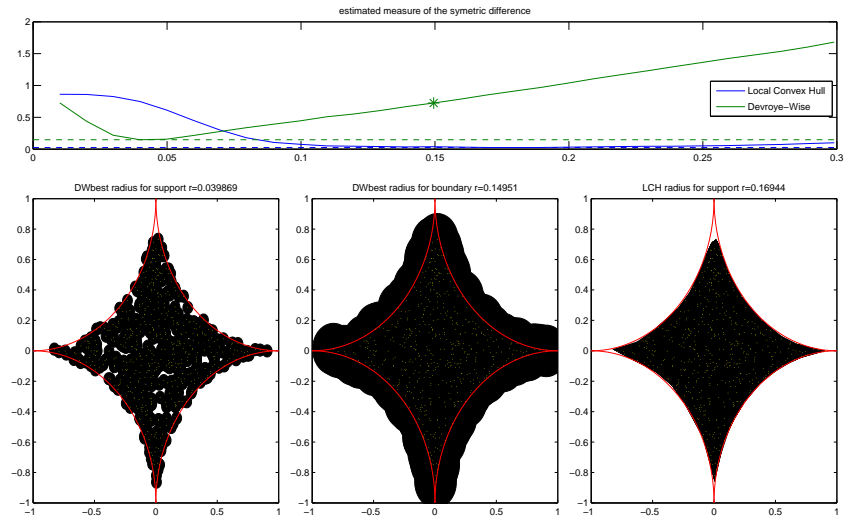


Figure 16: Result for a star shape

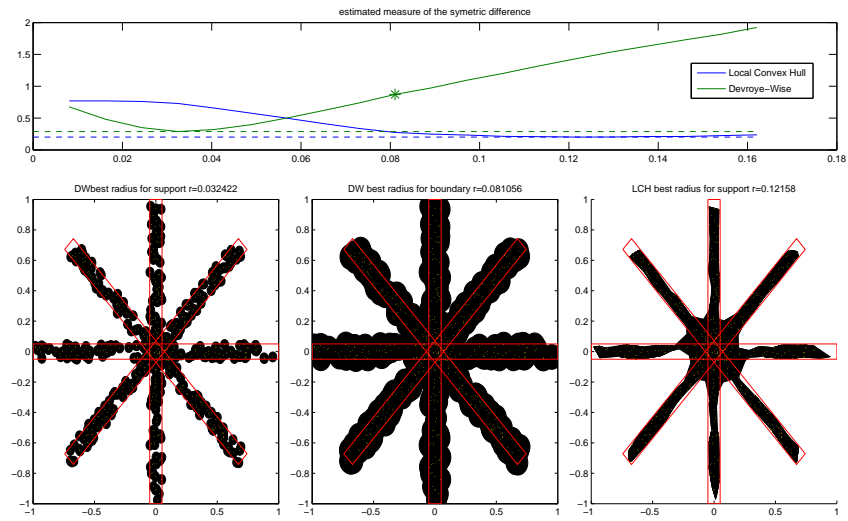


Figure 17: Result for an asterisk shape

only one parameter  $r$  and a method less sensitive to the choice of the parameter than the Devroye-Wise estimation. That are encouraging results but we

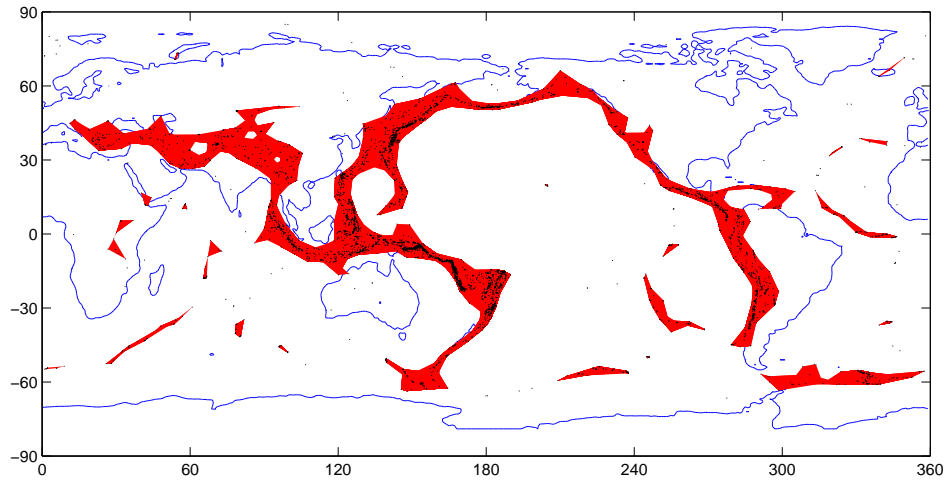


Figure 18: Result for epicenters

need to find a practical way to choose suitable value for the radius for a given problem. In addition to this practical problem some theoretical issues can also be pointed out: We would like to study the properties of this estimator when the support is less than  $d$ -dimensional (with  $d$  the dimension of the observation space). We strongly believe that it behaves very well but need to prove it. What happen when the density go to infinity when points gets close to the boundary ? Is our estimator linked with the  $\alpha$ -convexity ?

Moreover, in [GW04] the authors used the local-convex hull idea not only to estimate the support but also to estimate the level line of the density. We strongly believe that their idea is very well adapt to the case where the support is compact because it helps to decrease the bias of the density estimation when considering points near to the boundary.

# A Appendix (or supplementary material : proofs of Lemmas ? and ?

## A.1 proof of Lemma 4

*Proof.* Let  $X$  be drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls rolls freely inside and outside the support  $S$  of  $P_X$ .

We are first going to bound the probability  $P(X \in \mathcal{B}(x, r_n))$  for any sequence  $r_n \rightarrow 0$  and any  $x \in S$ .

For  $x \in S$  with  $d(x, \partial S) = z_0 < r$ .

for any  $z \in [0, z_0 + r]$  let us denote  $\mathcal{A}(z) = \mathcal{S}(O_y, R_S) \cap \mathcal{B}(x, r)$  and  $A(z)$  its  $d - 1$ dimensional measure.

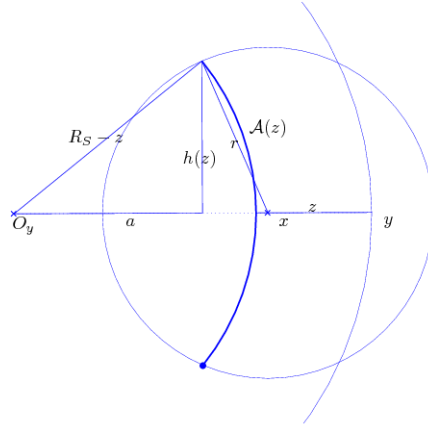


Figure 19: Probability that an observation belong to a given ball

$$P(X \in \mathcal{B}(x, r)) \geq \int_{z=0}^{z_0+r} z^\alpha f_0 A(z) dz$$

$$P(X \in \mathcal{B}(x, r)) \geq \int_{z=z_0}^{z_0+r} z^\alpha f_0 A(z) dz$$

When  $r = r_n = o(1)$ , we have  $z = O(r_n)$ ,  $z_0 = O(r_n)$  and  $A(z) \geq \omega_{d-1} h(z)^{d-1}$  with  $h(z) = z_0^2 + 2zz_0 - z^2 + r_n^2 + o(r_n^2)$  and so  $h(z) = r_n^2 - (z - z_0)^2 + 2z_0^2 + o(r_n^2)$  and so

$$P(X \in \mathcal{B}(x, r_n)) \geq \int_{z=z_0}^{z_0+r} z^\alpha f_0 \omega_{d-1} (r_n^2 - (z - z_0)^2)^{\frac{d-1}{2}} dz$$

Let us substitute  $z$  by  $u = (z - z_0)^2/r_n^2$  we have:

$$P(X \in \mathcal{B}(x, r_n)) \geq \frac{f_0 \omega_{d-1}}{2} r_n^{\alpha+d} (1 + o(r_n)) \int_{z=0}^1 u^{(\alpha-1)/2} (1-u)^{(d-1)/2} du$$

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

For  $x \in S$   $d(x, \partial S) = z_0 < r$ .

$$P(X \in \mathcal{B}(x, r)) \geq \int_{z=0}^r z^\alpha f_0 \omega_{d-1} r^{d-1} dz$$

So

$$P(X \in \mathcal{B}(x, r)) \geq \int_{z=0}^r z^\alpha f_0 \omega_{d-1} z^{d-1} dz$$

$$P(X \in \mathcal{B}(x, r)) \geq \frac{f_0 \omega_{d-1} r^{d+\alpha}}{d+\alpha}$$

And we have: For all  $r_n \rightarrow 0$ , for all  $x \in S$

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

So, we obtain (as in ‘‘On boundary estimation’’)

$$\limsup_{n \rightarrow +\infty} \left( \frac{n}{\ln n} \right)^{1/(d+\alpha)} d_h(\mathcal{X}_n, S) \leq \left( \frac{2}{\omega_{d-1} f_0 B(\frac{\alpha+1}{2}, \frac{d+1}{2})} \right)^{1/(d+\alpha)}$$

□

## A.2 proof of lemma 9

**Lemma 8.** *if  $X$  is drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls of radius  $R_S$  rolls freely inside and outside the support  $S$  of  $P_X$  then: for all  $x \in \partial S$ , for all  $r > 0$ , for all  $h \geq R_S - \sqrt{R_S^2 - r^2}$  we have:*

$$P(X \in \mathcal{C}(x, r, h)) \geq f_0 \omega_{d-1} \int_0^{r_0} z^\alpha \left(1 - \frac{z}{R_S}\right)^{d-1} dz$$

with

$$r_0 = R_S \frac{h - R_S + \sqrt{R_S^2 - r^2}}{\sqrt{R_S^2 - r^2}}$$

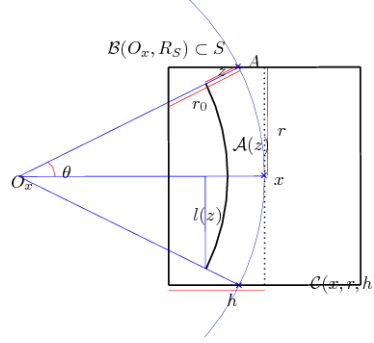


Figure 20:

**Corollary 4.** *if  $X$  is drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls of radius  $R_S$  rolls freely inside and outside the support  $S$  of  $P_X$  then: for all  $x \in \partial S$ , for  $r_n = \left(\lambda \frac{\ln n}{n}\right)^{\frac{1}{d+1+2\alpha}} (1 + o(1))$  and  $h_n = \mu \frac{r_n^2}{2R_S} (1 + o(1))$  with  $\mu > 1$  we have*

$$P(X \in \mathcal{C}(x, r, h)) \geq \frac{f_0 \omega_{d-1} (\mu - 1)^{\alpha+1} \lambda \ln n}{(\alpha + 1) (2R_S)^{\alpha+1} n} (1 + o(1))$$

**Lemma 9.** *if  $X_1, \dots, X_n$  are drawn from a distribution  $P_X$  associated to a  $\alpha$ -quickly decreasing density. Assume that balls of radius  $R_S$  rolls freely inside and outside the support  $S$  of  $P_X$ .  $r_n = \left(\lambda \frac{\ln n}{n}\right)^{\frac{1}{d+1+2\alpha}} (1 + o(1))$  and  $h_n = \mu \frac{r_n^2}{2R_S} (1 + o(1))$  with  $\mu > 1$  and  $\frac{\lambda f_0 \omega_{d-1} (\mu - 1)^{\alpha+1}}{(\alpha + 1) R_S^{\alpha+1}} - \frac{2d-2}{d+1+2\alpha} = c > 1$  Then: “for all  $x \in \partial S$   $\mathcal{C}(x, r_n, h_n) \cap \mathcal{X}_n \neq \emptyset$ ” eventually almost surely.*

*Proof.* We are going to prove that:

$$\sum_{n=1}^{+\infty} P(\exists x \in \partial S, \mathcal{C}(x, r_n, h_n) \cap \mathcal{X}_n = \emptyset) < \infty$$

Let us denote  $\varepsilon_n = (\ln n)^{-2}$ . Let us cover  $\partial S$  with  $N(\partial S, \varepsilon_n r_n^2)$  small deterministic balls, centered in points  $x_i \in \partial S$  and that have a radius  $\varepsilon_n r_n^2$ .

If  $\exists x \in \partial S, \mathcal{C}(x, r_n, h_n) \cap \mathcal{X}_n = \emptyset$  then exists a  $x_i$  such that  $x \in \mathcal{B}(x_i, \varepsilon_n r_n^2)$ . and  $\|u_x - u_{x_i}\| \leq R_S^{-1} \varepsilon_n r_n^2$  (according to Walther 99 Th1). So, according to Property ?? we can find explicit values for  $r'_n$  and  $h'_n$  such that:  $\mathcal{C}(x_i, r'_n, h'_n) \subset \mathcal{C}(x, r_n, h_n)$ ,  $r'_n = r_n(1 + o(1))$  and  $h'_n = h_n(1 + o(1))$ .

So:

$$p_n = P(\exists x \in \partial S, \mathcal{C}(x, r_n, h_n) \cap \mathcal{X}_n = \emptyset) \leq P(\exists x_i, \mathcal{C}(x_i, r'_n, h'_n) \cap \mathcal{X}_n = \emptyset)$$

According to corollary 4

$$p_n \leq N(\partial S, \varepsilon_n r_n^2) \left( 1 - \frac{f_0 \omega_{d-1} (\mu - 1)^{\alpha+1} \lambda \ln n}{(\alpha + 1)(2R_S)^{\alpha+1} n} (1 + o(1)) \right)^n$$

$$p_n \leq \lambda(\partial S) \varepsilon_n^{-d+1} (\ln n)^{-2d+2} \lambda^{-2} \frac{d-1}{d+1+2\alpha} n^{\frac{2d-2}{d+1+2\alpha}} n^{-\frac{f_0 \omega_{d-1} (\mu-1)^{\alpha+1} \lambda}{(\alpha+1)(2R_S)^{\alpha+1}} + o(1)}$$

With the chosen value for  $\varepsilon_n$

$$p_n \leq \lambda(\partial S) \lambda^{-\frac{2d-2}{d+1+2\alpha}} n^{-c+o(1)}$$

As  $c > 1$  we have  $\sum p_n < \infty$  and we can apply the Borrel-Cantelli Lemma to conclude.  $\square$

## References

- [B82] I. Bárány. Random polytopes in smooth convex bodie. *Mathematika*, 39:81–92, 1982.
- [BCMP09] G. Biau, B. Cadre, D.M. Mason, and B. Pelletier. Asymptotic normality in density support estimation. *Electronic Journal of Probability*, pages 2617–2635, 2009.
- [BCP08] G. Biau, B. Cadre, and B. Pelletier. Exact rates in density support estimation. *Journal of Multivariate Analysis*, 99:2185–2207, 2008.

- [Car05] 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.
- [CCSL09] 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.
- [CCSM11] F. Chazal, D. Cohen-Steiner, and Q. Mrigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11:733–751, 2011.
- [CF97] A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *The Annals of Statistics*, 25:2300–2312, 1997.
- [CF09] A. Cuevas and R. Fraiman. *New perspectives on Stochastic Geometry*. Oxford press, 2009.
- [CGOS11] F. Chazal, L. J. Guibas, S.Y. Odot, and P. Skraba. Scalar field analysis over point cloud data. *Discrete & Computational Geometry*, 46:746–775, 2011.
- [Che76] Chevalier. Estimation du support et du contour du support d’une loi de probabilit. *Ann. Inst. H. Poincaré sec. B*, 12:339–364, 1976.
- [CRC04] A. Cuevas and A. Rodríguez-Casal. On boundary estimation. *Advanced in Applied Probability*, 36:340–354, 2004.
- [CZCG05] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11:149–187, 2005.
- [DW80] 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.
- [DW96] L. Dumbgen and G. Walther. rates of convergence for random approximations of convex sets. *advanced in Applied Probability*, 28:384–393, 1996.
- [Efr65] B. Efron. The convex hull of a random set of points. *biometrika*, 15:331–343, 1965.

- [EKS83] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory*, 29:551–559, 1983.
- [ES97] H. Edelsbrunner and N. R. Shah. Triangulating topological spaces. *Internat. J. Comput. Geom. Appl.*, 7:365–378, 1997.
- [GW04] W. M. Getz and C.C. Wilmers. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, 27:489–505, 2004.
- [HPT95] W. Hardle, B.U. Park, and A.B. Tsybakov. Estimation of non-sharp support boundary. *Journal of Multivariate Analysis*, 55:205–218, 1995.
- [Kle04] J. Klemelä. Complexity penalized support estimation. *Journal of Multivariate Analysis*, 88:274–297, 2004.
- [Mei06] A. Meister. Support estimation via moment estimation in presence of noise. *Statistics: A Journal of Theoretical and Applied Statistics*, 40:259–275, 2006.
- [Rei03] M. Reitzner. Random polytopes and the efronstein jackknife inequality,. *Ann. Probab.*, 31:21362166., 2003.
- [Sch88] R. Schneider. Random approximation of convex sets. *Jornal of Microscopy*, 151:211–227, 1988.
- [Wal99] G. Walther. On a generalization of blaschke’s rolling theorem and the smoothing of surfaces. *Mathematical Methods in the Applied Sciences*, 22:301–316, 1999.
- [ZC05] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:247–274, 2005.