



HAL
open science

Local Convex Hull density and level set estimation

Catherine Aaron, Myriam Garrido

► **To cite this version:**

Catherine Aaron, Myriam Garrido. Local Convex Hull density and level set estimation. 2023. hal-04010316v1

HAL Id: hal-04010316

<https://hal.science/hal-04010316v1>

Preprint submitted on 1 Mar 2023 (v1), last revised 23 Jan 2024 (v2)

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 density and level set estimation

March 1, 2023

Catherine Aaron^a, and Myriam Garrido^{a,b}

^a Université Clermont Auvergne, France

^b INRAE, VetAgro Sup, UMR EPIA, F-63122 Saint-Genès-Champanelle,
France

Abstract

Let $\mathbb{X}_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^D$ be an iid sample drawn with a unknown distribution \mathbb{P}_X associated to a density f and supported by S . We study the problem of estimating the density f . When the density is continuous with regard to the Lebesgue measure and under wide hypotheses, the proposed density estimator has a minimax local asymptotic behaviour when $D \leq 7$, but also offers global warranties (whatever the dimension is). It allows to estimate the level sets of the density consistently, even for level sets that intersects the boundary. It is also easy to adapt it to the case of a distribution supported by a d -dimensional, smooth enough, sub-manifold of \mathbb{R}^D with similar convergence rates (depending on d instead of D). The paper is dedicated to the study of the asymptotical behaviour of the proposed estimator based on local behaviour and is concluded by numerical experiments.

1 Introduction

The density estimation problem, got an important impulse after Parzen (1962) and Rosenblatt (1956) fundamental pioneering works and is still an important and active area of research. There have been many well known books regarding the topic. Just to mention a few of them, see Silverman (1986), Devroye (1987), Devroye and Lugosi (1987), Scott (1992). However it is still a current topic, see Chacon and Duong (2018). Nowadays there are two axis that attract particular attention: the bias reduction near the boundary of the support and the density estimation on manifolds. The aim of this paper is to propose an asymptotically unbiased (on the whole support) density estimator which can easily be adapted to the manifold case.

Let $\mathbb{X}_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^D$ be an iid sample drawn according to an (unknown) distribution \mathbb{P}_X supported by S (possibly unknown). Recall that its support is $S = \bigcap_{E \text{ closed}, \mathbb{P}_X(E)=1} E$. Our density estimator is based on Getz and Wilmer (2004) in which it is proposed to estimate the support and the density respectively by

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

and

$$\tilde{f}_{r_n}(x) = \max_{X_i, x \in B(X_i, r_n)} \frac{\#(\mathbb{X}_n \cap B(X_i, r_n))}{n|H(B(X_i, r_n)) \cap \mathbb{X}_n|_D}. \quad (2)$$

Where $B(x, r)$ is the closed ball centered at x and of radius r , $|A|_D$ denotes the D -dimensional volume of a set A and $H(A)$ the convex hull of the set A .

It as been proved in Aaron and Bodart (2016) that the proposed support estimator has many nice properties (universality, minimax and topological guarantees under some additional hypothesis).

In this paper we focus on the density estimation and we propose an inspired by (2) density estimator. Indeed we will see that the density estimator proposed in Getz and Wilmer (2004) provides a wide overestimation of the density. But with help of small accurate corrections we can obtain an asymptotically unbiased estimator, with minimax guarantees.

Bias Reduction in the full dimensional setting. Suppose that the support $S \subset \mathbb{R}^D$ is regular, that is $\bar{\overset{\circ}{S}} = S$. This is equivalent to have a density that is absolutely continuous with regard to the Lebesgue measure. We will refer to this as the full dimensional setting since the dimension d of S equals D . Suppose that the density f is continuous and positive on $\overset{\circ}{S}$. It is well known that the Lebesgue differentiation theorem entails the so-called Besicovitch condition: let $x \in \overset{\circ}{S}$

$$\lim_{r \rightarrow 0} \frac{\int_{B(x,r)} f(t) dt}{|B(x,r)|_D} = f(x), \quad (3)$$

This is at the heart of the simplest density estimation methods (since $\int_{B(x,r)} f(t) dt$ can easily be estimated by the ratio of number of observation in $B(x, r)$ on the total number of observations) but fails for points $x \in \partial S$ such that $f(x) > 0$. When $d = 1$ bias reduction in density estimation has yet been extensively studied (see for instance Ruppert and Cline (1994), Marron and Ruppert (1994), Jones et al. (1995) or Leblanc (2010)). When considering higher dimensions, most of the proposed bias correction methods rely on corrections of kernel density estimators built on the knowledge of the support. This is the case for Bertin et al (2020), Charpentier and Gallic (2016), Funkea and Kawka (2015)

or Bouezmarnia and Rombouts (2015). Up to our knowledge the only method that allows bias reduction method when the support is not known can be found in Berry and Sauer (2017).

The way to extend (3) on S and to obtain uniform convergence rates is to use that for all $x \in S$

$$\lim_{r \rightarrow 0} \frac{\int_{B(x,r)} f(t) dt}{|B(x,r) \cap S|_D} = f(x), \quad (4)$$

which generalizes Equation (3) at possible discontinuity points localized on the boundary. Thus, via a plugin method, a very simple class of density estimators should be:

$$\check{f}(x) = \frac{\#\{\mathbb{X}_n \cap B(x,r)\}}{n|B(x,r) \cap S|_D}.$$

Notice that the use of $\widehat{|B(x,r) \cap S|_D} = |H(B(x,r) \cap \mathbb{X}_n)|_D$ induces an overestimation of the density for all x such that $B(x,r) \subset S$, that is most of the points. Thus the density estimator (2) by Getz and Wilmer (2004) which is even bigger than this quantity will provide a wide overestimation of the density.

In this paper, we propose a slightly modified estimator that takes into account the number of observations that are located on the boundary of $H(B(x,r) \cap \mathbb{X}_n)$ to reduce the aforementioned bias. The importance of the number of boundary observations in the probability contents estimation as been initially noticed in Efron (1965) (see Equation (3.7)) and more recently in Baldin and Reiss (2016). Our proposal is inspired by Baldin and Reiss (2016). Now, let us give some additional notations. First ω_D denotes the volume of the D dimensional unit ball. Then, based on the sample \mathbb{X}_n , let us introduce $N_{x,r} = \#\{\mathbb{X}_n \cap B(x,r)\}$, $C_{x,r} = H(B(x,r) \cap \mathbb{X}_n)$, $N_{x,r}^\partial = \#\{\mathbb{X}_n \cap \partial C_{x,r}\}$ and $N_{x,r}^o = N_{x,r} - N_{x,r}^\partial$. We propose the following density estimator:

$$\hat{f}_{r,A}(x) = \frac{N_{x,r}^o}{(n - N_{x,r}^\partial)|C_{x,r}|} \mathbb{I}_{|C_{x,r}|_D \geq A\omega_D r^D} \mathbb{I}_{N_{x,r}^\partial \leq n/2}. \quad (5)$$

If it first looks a bit cryptic, the introduction of $N_{x,r}^o$ and $N_{x,r}^\partial$ is in fact quite intuitive due to Baldin and Reiss (2016) since $N_{x,r}^o |C_{x,r}| \sim \text{Binom}(n - N_{x,r}^\partial, \int_{C_{x,r}} f(z) dz)$.

Section 2 is dedicated to the study of this new density estimator in the full dimensional setting. After a presentation of the associated model we give the main theoretical results of our density estimator. Roughly, our proposal has the same local asymptotic in \mathring{S} when $D \leq 7$ than the classical counting density estimator $f_r^*(x) = \frac{N_{x,r}}{n|B(x,r)|_D}$. But it also offers global asymptotics in S (whatever is the dimension), and, with additional assumptions uniform convergence rates. More precisely, let $r_n = n^{-1/(D+4)}$, the classical optimal window size in kernel density estimation. Theorem 1 and Corollary 1 set that, under quite wide hypothesis (slightly more restrictive than f of class \mathcal{C}^2 and regular enough support), we have for all $x \in \mathring{S}$:

1. If $D \leq 6$ then $\hat{f}_{r_n, A}(x) - f(x) \underset{L^2}{\sim} f_{r_n}^*(x) - f(x)$.
2. If $D = 7$ then $\hat{f}_{r_n, A}(x) - f(x) = O_{L^2}(f_{r_n}^*(x) - f(x))$.
3. If $D > 7$ then $\hat{f}_{r_n, A}(x) - f(x) \xrightarrow{L^2} 0$ but $f_{r_n}^*(x) - f(x) = o_{L^2}(\hat{f}_{r_n, A}(x) - f(x))$.

Moreover, we have the existence of a constant C such that for all $x \in S$ we have $\mathbb{E}((\hat{f}_{r_n, A}(x) - f(x))^2) \leq C(r_n^2 + (nr_n^D)^{-1})$, *i.e.* we have a uniform rate of convergence.

In Theorem 2 it is proved that, under the same hypotheses, $\max_i |\hat{f}_{r_n, A}(X_i) - f(X_i)|$ converges toward 0, going more further toward uniformity in convergence rates and allowing results in level set estimation (see Corollary 2). When additionally supposing that S is compact, in theorem 3 we present convergence rates for $\sup_{x \in S} |\hat{f}_{r_n, A}(x) - f(x)|$. Remark that in the case $D > 7$ it could be more accurate to estimate the density with $f_{r_n}^*(x)$ when x is estimated far enough from the boundary and by $\hat{f}_{r_n, A}(x)$ otherwise. This proposition is not studied in this paper which focuses on the case $D \leq 7$.

Equation (5) defines a density estimator that is asymptotically unbiased on S but that may be biased outside S . Thus it may be pertinent to prefer $\hat{f}_{r, A, S}(x) = \hat{f}_{r, A}(x)\mathbb{I}_S(x)$ when the support is known and $\hat{f}_{r, A, \hat{S}}(x) = \hat{f}_{r, A}(x)\mathbb{I}_{\hat{S}}(x)$ when the support is unknown.

Density on Manifolds. Suppose now that S is a d -dimensional manifold with $d < D$ what is frequently assumed when dealing with high dimensional data sets. Under such an hypothesis an equation similar to (3) is obtained considering integration on the manifold at numerator and the d -dimensional volume at the denominator, *i.e*

$$\lim_{r \rightarrow 0} \frac{\int_{B(x, r) \cap S} f(t) dt}{|B(x, r)|_d} = f(x), \quad (6)$$

Obviously, when dealing with a manifold with boundary, a bias correction should be applied as in the "full dimensional" case. First, Pelletier (2005) generalized the classical kernel density estimator when S is a known manifold without boundary replacing the Euclidean distance by the geodesic one. In Kim and Park (2013), the same kind of idea is used but applying kernel to projections on tangent spaces. Berry and Sauer (2017) adapt the method proposed in Kim and Park (2013) to take into account the possible boundary bias. Interestingly, the boundary do not need to be preliminary known. The distance (and direction) from a point to the boundary may be estimated via a local barycenter shift. More recently, the method developed in Divol (2021) only requires the knowledge of the dimension d of the support S , while in Berenfeld and Hoffmann (2021), S is "fully" unknown (but without boundary).

To extend the results from the full dimensional case to the manifold one we classically use a local projection on estimated tangent space step. Suppose now that, for all $x \in S$,

we have \hat{T}_x estimation of $T_x S$ the tangent (to S) space at x . Denote by $\pi_{\hat{T}_x}$ the orthogonal projection onto \hat{T}_x , and introduce, for $R \in \mathbb{R}$:

1. $\mathbb{X}_{n,R} = \pi_{\hat{T}_x}(B(x, R) \cap \mathbb{X}_n)$, with as previously $N_{x,R} = \#\{\mathbb{X}_n \cap B(x, R)\}$,
2. $\hat{C}_{x,r,R} = H(\mathbb{X}_{n,R} \cap B(x, r))$,
3. $\hat{N}_{x,r,R} = \#\{\mathbb{X}_{n,R} \cap B(x, r)\}$, $\hat{N}_{x,r,R}^\partial = \#\{\mathbb{X}_{n,R} \cap \partial \hat{C}_{x,r,R}\}$ and $\hat{N}_{x,r,R}^o = \hat{N}_{x,r,R} - \hat{N}_{x,r,R}^\partial$.

We propose the following natural extension of the estimator (5) in the manifold context:

$$\hat{f}_{r,R,A}(x) = \frac{N_{x,R}}{n} \frac{\hat{N}_{x,r,R}^o}{(N_{x,R} - \hat{N}_{x,r,R}^\partial) |\hat{C}_{x,r,R}|} \mathbb{I}_{|\hat{C}_{x,r,R}| \geq A \omega_d r^d} \mathbb{I}_{\hat{N}_{x,r,R}^\partial \leq N_{x,R}/2}, \quad (7)$$

Section 3 is dedicated to provide theoretical results on this estimator that can be summed up as follows. If S is a \mathcal{C}^2 manifold with positive reach τ_S , if ∂S is either empty or is a \mathcal{C}^2 manifold with positive reach $\tau_{\partial S}$ and when the tangent space estimation is suitable (*e.g.* via local PCA), then we obtain the same convergence rates as in the full dimensional case with d instead of D . In this section, proofs are a bit technical but mostly only combine results of section 2 and technical Lemmas in Aamari et Al (2023). They are so given in appendix.

In section 4 we present some experimental results on toy examples and a real data set.

2 Full dimensional context

2.1 Main Results

As mentioned above, we will consider an iid sample $\mathbb{X}_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^D$ drawn according to a distribution, supported by S (that may be unknown), and following an unknown density f uniformly continuous with regard to the Lebesgue measure.

We will also require some assumptions on f and some smoothness condition on S .

Considering regularity assumptions on S , we will consider two types of assumptions: the first one (Definition 1 and left part of the figure 1) is much more general and allows the presence of corners, while the second one (Definition 2 and right part in the figure 1) is in terms of positive reach of the boundary and is close to a \mathcal{C}^2 regularity of the boundary.

Definition 1. Ball-standardness. *A closed set E is (r_0, δ) -ball standard if for every $x \in E$ and every $r \leq r_0$, $B(x, r)$ is path connected and there exists y such $B(y, \delta r) \subset E \cap B(x, r)$.*

Definition 2. Inside and outside rolling ball condition. *A closed set E satisfies the r_0 -inside and outside rolling ball condition if, $\overline{E} = E$ and, for every $x \in \partial E$ there exists two points O_x^{out} and O_x^{in} such that $B(O_x^{out}, r_0) \cap E = \{x\}$ and $B(O_x^{in}, r_0) \cap \overline{E^c} = \{x\}$*

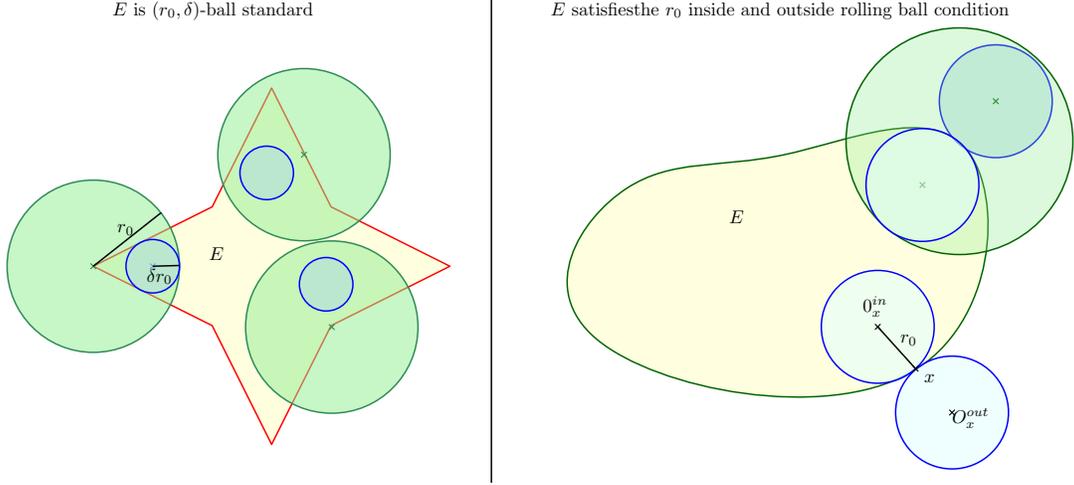


Figure 1: (r_0, δ) -ball standardness (left) and r_0 -inside an outside rolling ball condition (right)

It is easy to see that, if E satisfies the r_0 -inside and outside rolling ball condition, then E and $\overline{E^c}$ are $(2r_0, 1/2)$ -ball standard.

Our model requires distributions that have ball standard level sets for small levels, more precisely we will suppose that we have $(r_0, \delta, \varepsilon_0)$ -standard distribution.

Definition 3. $(r_0, \delta, \varepsilon_0)$ -**standardness for a distribution.** A distribution associated to a density f is $(r_0, \delta, \varepsilon_0)$ -standard if, for all $\varepsilon \in (0, \varepsilon_0)$ the level set $L_\varepsilon = \overline{\{x, f(x) > \varepsilon\}}$ is (r_0, δ) -ball standard.

Our model will consist in $(r_0, \delta, \varepsilon_0)$ -standard distributions with smooth enough densities.

Definition 4. A distribution belongs to the family model $\mathcal{M} := \mathcal{M}(r_0, \delta, \varepsilon_0, K_{\max}, f_{\max}^{(0)}, f_{\max}^{(1)}, f_{\max}^{(2)}, L_{\max})$ if:

1. The distribution is $(r_0, \delta, \varepsilon_0)$ -standard, S is path connected and for all $(x, y) \in S^2$ the geodesic distance (in S) satisfies $d_S(x, y) \leq K_{\max} \|x - y\|$.
2. The associated density f is continuous on S , of class \mathcal{C}^2 on \mathring{S} and:
 - (a) For all $x \in S$, $f(x) \leq f_{\max}^{(0)}$ and for all $x \in \mathring{S}$, $f(x) > 0$.
 - (b) For all $x \in S$, $\|\nabla f(x)\| \exp(3\|\nabla_{\ln(f)}(x)\|) \leq L_{\max}$, where ∇ stands for the gradient. Such a condition implies that $\|\nabla f(x)\| \leq f_{\max}^{(1)}$ for all x in S (see Proposition 2).

(c) For all $x \in S$, $\|\mathcal{H}_f(x)\|_{op} \leq f_{\max}^{(2)}$, where $\mathcal{H}_f(x)$ stands for the Hessian and $\|\cdot\|_{op}$ for the operator norm.

Model \mathcal{M} clearly includes distributions supported by compact (r_0, δ) -ball standard support S that have a \mathcal{C}^2 density bounded below by a positive constant on S and a hessian matrix bounded above (in norm) by $f_{\max}^{(2)}$. It also includes a wide range of classical densities with non compact support such as Gaussian or exponential.

Let recall here some notations and the proposed density estimator definition. let $N_{x,r} = \#\{\mathbb{X}_n \cap B(x, r)\}$, $C_{x,r} = H(B(x, r) \cap \mathbb{X}_n)$, $N_{x,r}^\partial = \#\{\mathbb{X}_n \cap \partial C_{x,r}\}$ and $N_{x,r}^o = N_{x,r} - N_{x,r}^\partial$. As in this part there is no ambiguity on the volume that always is the D dimensional volume we will omit the D index in the volume notation.

Recall that we study the following density estimator:

$$\hat{f}_{r,A}(x) = \frac{N_{x,r}^o}{(n - N_{x,r}^\partial) |C_{x,r}|} \mathbb{I}_{|C_{x,r}| \geq A \omega_D r^D} \mathbb{I}_{N_{x,r}^\partial \leq n/2}.$$

We characterize the convergence with the use of the following notations: $o_\alpha(1)$ (resp. $O_\alpha(u_n)$) is a sequence that depend on α and the constant of the model \mathcal{M} which converges toward 0 (resp. a sequence which is a $O(u_n)$ and depend on α and the constant of the model \mathcal{M}).

In a first theorem we explicit pointwise L^2 convergence rates. The first point is that our estimator has the same pointwise L^2 convergence rate than the classical density estimator, up to an additional term of order $r_n^2 (nr_n^D)^{-\frac{2}{D+1}} \min\left(r_n, (nr_n^D)^{-\frac{2}{D+1}}\right)$. When $D \leq 7$ and $r_n = O(n^{-\frac{1}{D+4}})$ this additional term is at most $O(n^{-\frac{4}{D+4}})$ (see Corollary 1). Thus, for small enough dimensions and suitable radius sequence the proposed estimator is pointwise minimax.

Theorem 1. *Let the distribution belongs to the model \mathcal{M} . Let the window size r_n be a sequence such that $r_n \rightarrow 0$ and $nr_n^D / \ln n \rightarrow +\infty$. Let $A < \delta^D$ be a positive constant.*

There exists C an explicit constant such that :

1. For all x such that $f(x) > 0$,

$$\begin{aligned} \mathbb{E}((\hat{f}_{r_n,A}(x) - f(x))^2) &\leq \left(\frac{f(x)}{\omega_D nr_n^D} + \|\mathcal{H}_f(x)\|_{op}^2 r_n^4 \right. \\ &\quad \left. + Cr_n^2 \left(\frac{\omega_D f(x) nr_n^D}{2} \right)^{-\frac{4}{D+1}} + Cr_n^3 \left(\frac{\omega_D f(x) nr_n^D}{2} \right)^{-\frac{2}{D+1}} \right) (1 + o_x(1)). \end{aligned}$$

2. For all $t > 0$ and $x \in L_t$ such that $B(x, r) \subset \{z, f(z) > 0\}$, we have

$$\mathbb{E}((\hat{f}_{r_n,A}(x) - f(x))^2) \leq \left(\frac{2f_{\max}^{(0)}}{A\omega_D n r_n^D} + (f_{\max}^{(2)})^2 r_n^4 \right. \\ \left. C r_n^2 \left(\frac{\omega_D f_{\max}^{(0)} n r_n^D}{2} \right)^{-\frac{4}{D+1}} + C r_n^3 \left(\frac{\omega_D f_{\max}^{(0)} n r_n^D}{2} \right)^{-\frac{2}{D+1}} \right) (1 + o_t(1)).$$

3. Finally, if S is either convex or has the R_0 inside and outside property we have that, for all x with $f(x) \geq t > 0$,

$$\mathbb{E}((\hat{f}_{r_n,A}(x) - f(x))^2) \leq \left(\frac{2f_{\max}^{(0)}}{A\omega_D n r_n^D} + (C' r_n)^2 \right) (1 + o_t(1)).$$

with $C' = f_{\max}^{(1)}$ when S is convex and $C' = f_{\max}^{(1)} + \frac{f_{\max}^{(0)} \omega_{D-1}}{A R_0 \omega_D}$ when S has the inside and outside rolling ball condition (see Definition 2).

As previously announced, Corollary 1 gives the L^2 convergence rates for $r_n = cn^{-\frac{1}{D+4}}$ which is the optimal window size in the classical kernel setting. When considering the local convergence rates, when $D < 7$ the proposed estimator is equivalent to the classical constant kernel one, for the special case $D = 7$ it is still minimax, and when $D > 7$ it is still convergent but no more minimax. In all the cases we also have uniform convergence rates for all level sets, that turns to uniform convergence rates on S when the density is lower bounded by a positive constant.

Corollary 1. *Let the distribution belongs to the model \mathcal{M} . Let $f_{r_n}^*$ be the usual counting density estimator (constant kernel) and consider window size $r_n = cn^{-1/(D+4)}$. As previously x is such that $f(x) > 0$.*

1. *Local rates :*

(a) *if $D \leq 6$ then $\mathbb{E}((\hat{f}_{r_n,A}(x) - f(x))^2) \leq \mathbb{E}((f_{r_n}^*(x) - f(x))^2)(1 + o_x(1)) = O_x(n^{-\frac{4}{D+4}})$,*

(b) *if $D = 7$ then $\mathbb{E}((\hat{f}_{r_n,A}(x) - f(x))^2) = O_x(n^{-\frac{4}{D+4}})$,*

(c) *if $D > 7$ then $\mathbb{E}((\hat{f}_{r_n,A}(x) - f(x))^2) = O_x(n^{-\frac{2(D+9)}{(D+1)(D+4)}})$.*

2. *Moreover if S is either convex or has the inside and outside rolling ball property for all x with $f(x) > 0$, $\mathbb{E}((\hat{f}_{r_n,A}(x) - f(x))^2) = O_{f(x)}(n^{-\frac{2}{D+4}})$. In particular, when f is lower bounded by a positive constant $f_{\min}^{(0)}$ on S , then, for all $x \in S$, $\mathbb{E}((\hat{f}_{r_n,A}(x) - f(x))^2) = O(n^{-\frac{2}{D+4}})$.*

We are now focusing only on the case $D \leq 7$, where the proposed estimator is point-wise minimax. When $D > 7$, in the case of compact unknown support (with the R_0 inside and outside rolling ball condition) and lower bounded density, a way to obtain minimax rates should be to plug the ρ -convex hull support estimator \hat{S}_n and consider a density estimator that should be $f_{r_n}^*(x)$ (the classical constant kernel density estimator) when $B(x, r_n) \subset \hat{S}_n$, \hat{f}_{A, r_n} if $x \in \hat{S}_n$ with $d(x, \partial\hat{S}_n) \leq r_n$ and 0 otherwise. Nevertheless, the curse of dimensionality makes this approach practically unrealistic neither for the support estimation nor for the density estimation.

When $D \leq 7$ we also have uniform convergence of \hat{f}_{A, r_n} when evaluating it on the X_i 's that, in turn, allows to provide convergence rates for level set estimation.

Theorem 2. *When $D \leq 7$ and the distribution belongs to the model \mathcal{M} . Let $r_n = cn^{-1/(D+4)}$ and $A < \delta^D$.*

1. *For all $t > 0$, there exists C_t such that, with probability one, for n large enough:*

$$\max_i \left\{ \left| \hat{f}_{r_n, A}(X_i) - f(X_i) \right|, X_i \in L_t \text{ and } B(X_i, r_n) \subset S \right\} \leq C_t \sqrt{\ln nn}^{-2/(D+4)}.$$

2. *If f is bounded below by a positive constant $f_{\min}^{(0)}$, there exists C such that, with probability one, for n large enough:*

$$\max_i \left\{ \left| \hat{f}_{r_n, A}(X_i) - f(X_i) \right|, B(X_i, r_n) \subset S \right\} \leq C \sqrt{\ln nn}^{-2/(D+4)}.$$

3. *If S is either convex or satisfies the inside and outside R_0 rolling ball condition, there exists C such that, with probability one, for n large enough:*

$$\max_i \left\{ \left| \hat{f}_{r_n, A}(X_i) - f(X_i) \right| \right\} \leq C n^{-1/(D+4)}.$$

Such a result allows to plug the proposed density estimator for level set estimation using $\hat{S}(\mathbb{X}_{r_n, A}^+(t))$, where $\hat{S}(\mathbb{Y})$ is a set estimator based on a sample \mathbb{Y} and $\mathbb{X}_{r_n, A}^+(t) = \{X_i, \hat{f}_{r_n, A}(X_i) \geq t\}$. We propose to give results when the chosen set estimator is the local convex hull one. This choice has been done because it seems consistent to estimate the density and the level sets with the same tool (the local convex hull). It is also a tribute to Getz and Wilmer (2004) who first applied local convex hull to draw home range and core area maps. In the following corollary we give convergence rates for $\hat{L}_t = \bigcup_{X_i \in \mathbb{X}_{r_n, A}^+(t)} H(B(X_i, r_n) \cap \mathbb{X}_{r_n, A}^+(t))$. Notice that it is not the original one proposed in Getz and Wilmer (2004) that was $\tilde{L}_t = \bigcup_{X_i \in \mathbb{X}_{r_n, A}^+(t)} H(B(X_i, r_n) \cap \mathbb{X}_n)$ which "overestimates" the support.

Corollary 2. *Suppose that $D \leq 7$ and the distribution belongs to the model \mathcal{M} with S is either convex or have the inside and outside rolling ball property. Let $r_n = cn^{-1/(D+4)}$ and $A < \delta^D$. Assume that $\mathring{S} = \{z, f(z) > 0\}$. Introduce $\mathbb{X}_n^+(t) = \{X_i, \hat{f}_{r_n, A}(X_i) \geq t\}$ and define:*

$$\hat{L}_t := \bigcup_{i, \hat{f}_{r_n, A}(X_i) \geq t} H(B(X_i, r_n) \cap \mathbb{X}_n^+(t)).$$

1. *If $L_t \subset \mathring{S}$ and $\min\{\|\nabla f(x)\|, f(x) = t\} = m_t > 0$ with probability one, for n large enough $d_H(L_t, \hat{L}_t) = O(\ln nn^{-2/(D+4)})$.*
2. *If L_t is regular (i.e. there exists Δ_0 and β such that such that, for all $t', |t' - t| \leq \Delta_0$ we have $d_H(L_t, L_{t'}) \leq \beta|t - t'|$) and L_t is ball standard then, with probability one, for n large enough $d_H(L_t, \hat{L}_t) = O(\ln nn^{-1/(D+4)})$.*

Notice that the bias correction allows to have convergence for level sets that may intersect the boundary of the support. Most of the time, estimated level sets are supposed to be in \mathring{S} as in Rodríguez-Casal and Saavedra-Nieves (2022). Nevertheless a distribution may have no level set included in \mathring{S} . Consider for instance the 2-dimensional case of a restricted Gaussian with a density $f(x, y) = C \exp(-2(x^2 + y^2)) \mathbb{I}_{(x, y) \in [0, 1]^2}$, as used in one of the simulated examples in Section 4.1. All its level sets intersects the boundary of the support $[0, 1]^2$. Also Remark that if the \mathcal{C}^2 regularity of the density make the level lines $\ell(t) = \{z, f(z) = t\}$ also \mathcal{C}^2 (as manifold), the level sets are only ball standard in this case. For such an example the convergence rate is given by the point 2 of Corollary 2. In fact, to be "global" this convergence rate is rough and it is expected to be "locally" slightly better with the rough bound localized around the singularities of the support and around the intersections between the level lines and the boundary of the support. A more precise study of the precise behaviour of the level set estimation being let for future.

Finally, we can derive uniform convergence rates on the support under a more restrictive set of assumption.

Theorem 3. *Consider $D \leq 6$ and the distribution belongs to the model \mathcal{M} . Also assume that S satisfies the inside and outside R_0 rolling ball condition, and f is bounded bellow by a positive constant $f_{\min}^{(0)}$. Choose $r_n = cn^{-\frac{1}{D+4}}$, and $A < 1/2$. Then, with probability one, when n large enough:*

$$\sup_{x \in S} |\hat{f}_{A, r_n}(x) - f(x)| \leq O\left(n^{-\frac{1}{D+4}}\right).$$

When $D = 7$, under the same conditions, we have :

$$\sup_{x \in S} |\hat{f}_{A, r_n}(x) - f(x)| \leq O\left(n^{-\frac{1}{2(D+4)}}\right).$$

To conclude the discussion on the proposed estimator in the special case $D \leq 7$ let us notice that, when S is known, the proposed method allows to correct the bias without any computation of ∂S and any Monte-Carlo integration step. It also allows to obtain a bias correction in case of unknown support, which is a more realistic purpose. Remark that, even if we suppose that the support is known, in fact we may only know a set containing the support because there might be 0 density regions in our expected support. To illustrate: suppose we aim at estimating the population density. We have knowledge of the submerged land which is our expected support but it includes desert regions and the "true support" is only a subset of the expected one. In case of unknown support we may plug a support estimator \hat{S} and compute $\hat{f}_{r_n, A, \hat{S}} = \hat{f}_{r_n, A} \mathbb{1}_{\hat{S}}$. Obviously the uniform convergence is then lost on \mathbb{R}^D but is limited to $\mathbb{R}^D \setminus (S \Delta \hat{S}_n)$.

2.2 Proofs

2.2.1 Notations

Due to the amount of notations and to have a self contained section we list here main notations and definition used in this proof section. First, for ease of reading, we will write r instead of r_n in this proof section.

The probability content is $\Gamma_{x,r} = \int_{B(x,r)} f(z) dz$ and its approximation with the help of local convex hull is $\tilde{\Gamma}_{x,r} = \int_{B(x,r)} f(z) dz$.

Due to the indicator function in the density estimation definition we will have to condition by the event $E_{x,r}$ " $|C_{x,r}| \geq A\omega_D r^D$, $\tilde{\Gamma}_{x,r} \geq f(x)A\omega_D r^D$ and $N_{x,r}^\partial \leq n/2$ ". And, conditionally to $E_{x,r}$, the error decomposition in the density estimation is :

$$\hat{f}_{r,A}(x) - f(x) | E_{x,r} = \varepsilon_1(x) + \varepsilon_2(x),$$

where:

$$\varepsilon_1(x) = \frac{N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)}{|C_{x,r}|(n - N_{x,r}^\partial)} \mathbb{1}_{E_{x,r}} \quad \text{and} \quad \varepsilon_2(x) = \left(\frac{\tilde{\Gamma}_{x,r}}{|C_{x,r}|} - f(x) \right) \mathbb{1}_{E_{x,r}}.$$

Roughly ε_1 is the variance term of the error and ε_2 the bias one. Indeed $\mathbb{E}(\varepsilon_1)$ converges toward 0 while ε_2 can be deterministically bounded.

We will also make use of the following notations. For a set A and a real number r , the r offset of A , denoted by $A \oplus rB$, is defined as $A \oplus rB = \{x \in \mathbb{R}^D, d(x, A) \leq r\}$.

For a compact set A , we denote by π_A the projection onto A which is defined as follows: $\pi_A(x)$ is the point in A such that $\|\pi_A(x) - x\| = d(x, A)$. With regard to the possible definition of $\pi_A(x)$, there are well known results : when A is convex, π_A is defined everywhere ; when A has the r_0 inside and outside rolling ball property, π_A is defined on $A \oplus r_0B$ and $\pi_{\partial A}$ on $\partial A \oplus r_0B$.

2.2.2 Error decomposition and first bounds

We have that

$$\mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) \leq \mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2 | E_{x,r}) + f(x)\mathbb{P}(E_{x,r}^c), \quad (8)$$

$$\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t) \leq \mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t | E_{x,r}) + \mathbb{P}(E_{x,r}^c). \quad (9)$$

We then have

$$\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t_1 + t_2) \leq \mathbb{P}(|\varepsilon_1(x)| \geq t_1 | E_{x,r}) + \mathbb{P}(|\varepsilon_2(x)| \geq t_2 | E_{x,r}) + \mathbb{P}(E_{x,r}^c), \quad (10)$$

and also

$$\mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) \leq \mathbb{E}(\varepsilon_1^2(x) | E_{x,r}) + \mathbb{E}(\varepsilon_2^2(x) | E_{x,r}) + 2\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x) | E_{x,r}) + f(x)\mathbb{P}(E_{x,r}^c). \quad (11)$$

2.2.3 General bounds on $\mathbb{P}(E_{x,r}^c)$, $\varepsilon_1(x)$ and $\varepsilon_2(x)$.

Bounds on $\mathbb{P}(E_{x,r}^c)$

Lemma 1. *Suppose that distribution belongs to the model \mathcal{M} .*

For all $A < \delta^D$ and $r < \min(r_1, (4\omega_D)f_{\max}^{(0)})^{-1/D}$, there exists β_1 and β_2 positive constants such that, for all x with $f(x) > 0$, we have

$$\mathbb{P}(E_{x,r}^c) \leq \beta_1 \exp\left(-nr^D \min(f(x), \varepsilon_0)\beta_2\right) + \exp(-n/8).$$

Proof. First, tackle the condition on $N_{x,r}^\partial$. Remark that $\mathbb{P}(N_{x,r}^\partial > n/2) \leq \mathbb{P}(N_{x,r} > n/2)$. Since $N_{x,r} \sim \text{Binom}(n, \Gamma_{x,r})$, with $\Gamma_{x,r} < 1/4$ by condition on $r \leq (4\omega_D f_{\max}^{(0)})^{-1/D}$, we have that

$$\mathbb{P}(N_{x,r} > n/2) = \mathbb{P}\left(N_{x,r} - n\Gamma_{x,r} > n\left(\frac{1}{2} - \Gamma_{x,r}\right)\right) \leq \mathbb{P}(N_{x,r} - n\Gamma_{x,r} > n/4).$$

Then, by use of Hoeffding

$$\mathbb{P}(N_{x,r}^\partial > n/2) \leq \mathbb{P}(N_{x,r} > n/2) \leq \exp(-n/8). \quad (12)$$

To finish the proof, let us first establish an intermediate result. Let L_t (with $t > 0$) be a level set that has the (δ, r_1) -inside rolling ball property and set

$$\alpha = \min\left(\frac{\delta^D}{3}, \left(\frac{1}{4}(\delta - A^{1/D})\right)^D\right).$$

We aim at proving that, for all $x \in L_t$, we have that

$$\mathbb{P}\left(\frac{|C_{x,r}|}{r^D} \leq A\omega_D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D r^D A t\right) \leq (C_D + 1)C_D 4^D \left(1 - \frac{A^{1/D}}{\delta}\right)^{-D} \exp(-\omega_D \alpha n r^D t), \quad (13)$$

where C_D is the "covering constant" of a D -Dimensional Ball.

Notice that $|C_{x,r}| \geq |C_{y,\delta r}|$, where $C_{y,\delta r}$ denotes the convex hull of the observations in $B(y, \delta r)$ the ball of the "ball-standardness" assumption. For a given $\alpha \in (0, 1)$, suppose that $B(y, \alpha\delta r)$ is not included in $C_{y,\delta r}$. Thus, there exists y_1 in $B(y, \alpha\delta r)$ which is not in $C_{y,\delta r}$. Because y_1 is not in $C_{y,\delta r}$, there exists u_1 a unit vector such that, for all observation X_i in $B(y, \delta r)$, we have $\langle X_i - y_1, u_1 \rangle < 0$. Introduce now $z = y + \frac{\langle y_1 - y, u_1 \rangle + \delta r}{2} u_1$. We have that z belongs to $B(y, \delta r)$ with $B(z, (1 - \alpha)\delta r/2)$ contains no observation. We have then proved that, if $B(y, \alpha\delta r)$ is not included in $C_{y,\delta r}$, there exists $z \in B(y, \delta r)$ with $B(z, (1 - \alpha)\delta r/2) \cap \mathbb{X}_n = \emptyset$ (see Figure 2).

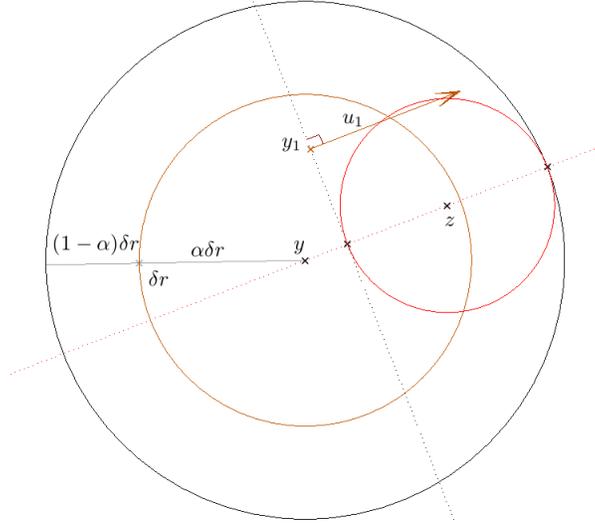


Figure 2: Construction of z

Now, deterministically cover $B(y, \delta r)$ with $\nu \leq C_D 4^D (1 - \alpha)^{-D}$ balls of radius $(1 - \alpha)\delta r/4$ centered at some z'_i . It clearly comes that, if $B(y, \alpha\delta r)$ is not included in $C_{y,\delta r}$, there exists $z'_i \in B(y, \delta r)$ with $B(z'_i, (1 - \alpha)\delta r/4) \cap \mathbb{X}_n = \emptyset$. Thus:

$$\mathbb{P}(B(y, \alpha\delta r) \not\subseteq C_{y,\delta r}) \leq \sum_{i=1}^{\nu} \mathbb{P}(B(z'_i, (1 - \alpha)\delta r/4) \cap \mathbb{X}_n = \emptyset),$$

$$\mathbb{P}(B(y, \alpha\delta r) \not\subseteq C_{y, \delta r}) \leq C_D 4^D (1 - \alpha)^{-D} \left(1 - \omega_D t ((1 - \alpha)\delta r/4)^D\right)^n,$$

$$\mathbb{P}(B(y, \alpha\delta r) \not\subseteq C_{y, \delta r}) \leq C_D 4^D (1 - \alpha)^{-D} \exp(-n\omega_D t ((1 - \alpha)\delta r/4)^D).$$

Thus,

$$\mathbb{P}\left(\frac{|C_{x,r}|}{r^D} \leq \omega_D (\delta\alpha)^D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D (\delta\alpha)^D t\right) \leq C_D 4^D (1 - \alpha)^{-D} \exp(-n\omega_D t ((1 - \alpha)\delta r/4)^D).$$

And so, if $A < \delta^D$,

$$\mathbb{P}\left(\frac{|C_{x,r}|}{r^D} \leq A\omega_D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D r^D A t\right) \leq C_D 4^D \left(1 - \frac{A^{1/D}}{\delta}\right)^{-D} \exp\left(-n\omega_D t \left(\left(1 - \frac{A^{1/D}}{\delta}\right) \frac{\delta r}{4}\right)^D\right),$$

that is (13).

Let now x be such that $f(x) > 0$.

If $f(x) < \varepsilon_0$, by application of (13), we have the existence of positive constants β_1 and β_2 such that

$$\mathbb{P}\left(\frac{|C_{y, \delta r}|}{r^D} \leq A\omega_D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D r^D A f(x)\right) \leq \beta_1 \exp(-nr^D f(x)\beta_2)$$

If $f(x) \geq \varepsilon_0$, also by application of (13), it comes that.

$$\mathbb{P}\left(\frac{|C_{y, \delta r}|}{r^D} \leq A\omega_D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D r^D \varepsilon_0 A\right) \leq \beta_1 \exp(-nr^D \varepsilon_0 \beta_2).$$

That concludes the proof. \square

Bounds on $\varepsilon_1(x)$

Lemma 2. *Suppose that distribution belongs to the model \mathcal{M} .*

We have that

$$\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|E_{x,r}) = 0. \quad (14)$$

We also roughly have a uniform bound given by:

$$\mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) \leq \frac{2 \max_{z \in C_{x,r}}(f(z))}{A\omega_D r^D n} \leq \frac{2 \max_{z \in B(x,r)}(f(z))}{A\omega_D r^D n}. \quad (15)$$

A better, but local, L_2 bound for points far from the boundary is the following : if x is such that $f(x) > 0$ and n large enough to have $\frac{\max_{B(x,r)} f(z)}{\min_{B(x,r)} f(z)} \leq 2$, then

$$\mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) \leq \frac{f(x)}{n|B(x,r)|} (1 + o_x(1)). \quad (16)$$

An explicit expression for o_x is given within the proof.

In probability we also have

$$\mathbb{P} \left(|\varepsilon_1(x)| > \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} |E_{x,r} \right) \leq 2n^{-4}(1 + o(1)). \quad (17)$$

Proof. Recall that

$$\varepsilon_1(x) = \frac{N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)}{|C_{x,r}|(n - N_{x,r}^\partial)} \mathbb{I}_{E_{x,r}}$$

When $C_{x,r}$ is known, obviously $|C_{x,r}|$ and $\tilde{\Gamma}_{x,r}$ are also known. Due to the convex polyhedron shape of $C_{x,r}$ its knowledge also implies the location of its vertices one, that, in turn fix, with probability one, $N_{x,r}^\partial$. As a conclusion, when conditioning by $C_{x,r}$ the only part of $\varepsilon_1(x)$ that remains random is $N_{x,r}^o$. See Baldin and Reiss (2016) for the possibility to define $N_{x,r}^o|C_{x,r}$. With probability one only $n - N_{x,r}^\partial$ points remain "randomly distributed" and may fall in $\hat{C}_{x,r}$ with probability $\tilde{\Gamma}_{x,r}$. It then become clear that, almost surely $N_{x,r}^o|C_{x,r} \sim \text{Binom}(n - N_{x,r}^\partial, \tilde{\Gamma}_{x,r})$.

First of all notice that, because ε_2 is fully determined by $C_{x,r}$, we have that

$$\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|C_{x,r}) = \frac{1}{|C_{x,r}|(n - N_{x,r}^\partial)} \left(\frac{\tilde{\Gamma}_{x,r}}{|C_{x,r}|} - f(x) \right) \mathbb{E}(N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)|C_{x,r}) = 0.$$

Thus

$$\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|E_{x,r}) = 0,$$

which proves (14).

We also have $\mathbb{E}(\varepsilon_1^2(x)|C_{x,r}) = \frac{\tilde{\Gamma}_{x,r}(1-\tilde{\Gamma}_{x,r})}{|C_{x,r}|^2(n-N_{x,r}^\partial)}$. Then, due to $\frac{\tilde{\Gamma}_{x,r}(1-\tilde{\Gamma}_{x,r})}{|C_{x,r}|} \leq \frac{\tilde{\Gamma}_{x,r}}{|C_{x,r}|} \leq \max_{z \in C_{x,r}}(f(z))$, it comes that

$$\mathbb{E}(\varepsilon_1^2(x)|C_{x,r}) \leq \frac{\max_{z \in B(x,r)} f(z)}{n - N_{x,r}^\partial} \frac{1}{|C_{x,r}|}.$$

Thus we clearly roughly obtain

$$\mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) \leq \frac{2 \max_{z \in B(x,r)} f(z)}{An\omega_D r^D},$$

which proves (15).

But it also can be obtained a better (but asymptotic) constant, noticing that

$$\begin{aligned}\mathbb{E}(\varepsilon_1^2(x)|C_{x,r}) &\leq \frac{\max_{z \in B(x,r)} f(z)}{n|B(x,r)|} \frac{1}{1 - \frac{N_{x,r}^\partial}{n}} \left(1 + \frac{|B(x,r) \setminus C_{x,r}|}{|C_{x,r}|}\right), \\ \mathbb{E}(\varepsilon_1^2(x)|C_{x,r}) &\leq \frac{\max_{z \in B(x,r)} f(z)}{n|B(x,r)|} \left(1 + \frac{\frac{N_{x,r}^\partial}{n}}{1 - \frac{N_{x,r}^\partial}{n}}\right) \left(1 + \frac{|B(x,r) \setminus C_{x,r}|}{|C_{x,r}|}\right).\end{aligned}$$

Because, conditionally to $E_{x,r}$, we have $\left(1 - \frac{N_{x,r}^\partial}{n}\right)^{-1} \leq 2$ and $\frac{|B(x,r) \setminus C_{x,r}|}{|C_{x,r}|} \leq \frac{|B(x,r) \setminus C_{x,r}|}{A|B(x,r)|} \leq A^{-1}$, we obtain:

$$\mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) \leq \frac{\max_{z \in B(x,r)} f(z)}{n|B(x,r)|} \left(1 + \frac{2}{n} \mathbb{E}(N_{x,r}^\partial | E_{x,r})(1 + A^{-1}) + A^{-1} \mathbb{E}\left(\frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} | E_{x,r}\right)\right).$$

1. By continuity $\max_{z \in B(x,r)} f(z) = f(x)(1 + o_x(1))$.
2. We have $\mathbb{E}(N_{x,r}^\partial | E_{x,r}) \leq \frac{\mathbb{E}(N_{x,r}^\partial)}{\mathbb{P}(E_{x,r})}$. Thus, because $N_{x,r}^\partial \leq N_{x,r}$ and due to Lemma 1, we obtain $\mathbb{E}(N_{x,r}^\partial | E_{x,r}) \leq \frac{\omega_D n r^D f_{\max}^{(0)}}{1 - \beta_1 \exp(-\beta_2 \omega_D n r^D f(x)) - \exp(-n/8)} = o_x(1)$.
3. To obtain better bounds on $|B(x,r) \setminus C_{x,r}|$ and $N_{x,r}^\partial$ we apply results in Brunel (2017) recalled in a suitable form to our purpose in Proposition 3 (see Appendix A). They are applied to a sample drawn with a density $g_{x,r}(y) = r^D \frac{f(x+ry)}{\int_{B(x,r)} f(z) dz} \mathbb{1}_{B(0,1)}(y)$ which is the density of a local re-scaling of the part of the original sample which is in $B(x,r)$. Also introduce $M_{x,r} = \max_{y \in B(0,1)} g_{x,r}(y)$.

Because n is large enough to have $\frac{\max_{B(x,r)} f}{\min_{B(x,r)} f} < 2$, that implies that $M_{x,r} \leq \frac{2}{\omega_D}$, by application of (33) in Proposition 3, it comes that:

$$\mathbb{E}\left(\frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} | N_{x,r} = k\right) \leq 2A_1 \left(1 + \frac{2}{\omega_D}\right) k^{-\frac{2}{D+1}}.$$

Thus, by application of Proposition 1 in Appendix A, with $\Gamma_{x,r} = \int_{B(x,r)} f(z) dz$ (and because, if $N_{x,r} = 0$ we clearly have $\frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} = 1$), we have

$$\mathbb{E}\left(\frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|}\right) \leq (1 - \Gamma_{x,r})^n + 2A_1 \left(1 + \frac{2}{\omega_D}\right) \left(\exp\left(-\frac{n\Gamma_{x,r}}{10}\right) + \left(\frac{n\Gamma_{x,r}}{2}\right)^{-\frac{2}{D+1}}\right),$$

$$\begin{aligned} \mathbb{E} \left(\frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \right) &\leq \left(1 - \frac{\omega_D r^D f(x)}{2} \right)^n \\ &+ 2A_1 \left(1 + \frac{2}{\omega_D} \right) \left(\exp \left(-\frac{nr^D \omega_D f(x)}{20} \right) + \left(\frac{nr^D \omega_D f(x)}{4} \right)^{-\frac{2}{D+1}} \right). \end{aligned}$$

Thus, adding a conditional step, we also have $\mathbb{E} \left(\frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \mid E_{x,r} \right) = o_x(1)$.

We then have proved Equation 16.

Now, to obtain (17), notice that, by Bennett's inequality, it comes that:

$$\begin{aligned} \mathbb{P} \left(|N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)| > \sqrt{16 \ln n (n - N_{x,r}^\partial) \tilde{\Gamma}_{x,r}} \mid C_{r,x} \right) &\leq \\ &2 \exp \left(-(n - N_{x,r}^\partial) \tilde{\Gamma}_{x,r} h \left(\sqrt{\frac{16 \ln n}{(n - N_{x,r}^\partial) \tilde{\Gamma}_{x,r}}} \right) \right), \end{aligned}$$

with $h(u) = (1+u) \ln(1+u) - u$. Because $h(u) \geq u^2/4$ when $u \leq 4$, we have that

$$\mathbb{P} \left(|\varepsilon_1(x)| > \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} \mid E_{x,r} \text{ and } \tilde{\Gamma}_{x,r} \geq \frac{8 \ln n}{n} \right) \leq 2n^{-4}. \quad (18)$$

Now, we aim at deriving a probabilistic bound conditionally to $\tilde{\Gamma}_{x,r} \leq \frac{8 \ln n}{n}$. By use of first Bernstein inequality, we have:

$$\begin{aligned} \mathbb{P} \left(|N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)| > \sqrt{a' \ln n (n - N_{x,r}^\partial) r^D} \mid C_{r,x} \right) &\leq \\ &2 \exp \left(-\frac{a' \ln n (n - N_{x,r}^\partial) r^D}{2\tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial) + 2\sqrt{a' \ln n (n - N_{x,r}^\partial) r^D}/3} \right). \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{P} \left(|\varepsilon_1(x)| > \sqrt{\frac{a' \ln n r^D}{(n - N_{x,r}^\partial) |C_{x,r}|^2}} \mid C_{x,r} \right) &\leq \\ &2 \exp \left(-\frac{a' \ln n (n - N_{x,r}^\partial) r^D}{2\tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial) + 2\sqrt{a' \ln n (n - N_{x,r}^\partial) r^D}/3} \right). \end{aligned}$$

Now, choose $a' = 16\omega_D f_{\max}^{(0)}$. It comes

$$\mathbb{P} \left(|\varepsilon_1(x)| > \sqrt{\frac{32f_{\max}^{(0)} \ln n}{A^2\omega_D n r^D}} \mid E_{x,r} \text{ and } \tilde{\Gamma}_{x,r} \leq \frac{8 \ln n}{n} \right) \leq 2 \exp \left(-\frac{a' \ln n n r^D}{16 \ln n + \frac{2}{3} \sqrt{2a' \ln n n r^D}} \right).$$

And because $n r^D / \ln n \rightarrow +\infty$,

$$\mathbb{P} \left(|\varepsilon_1(x)| > \sqrt{\frac{32f_{\max}^{(0)} \ln n}{A^2\omega_D n r^D}} \mid E_{x,r} \text{ and } \tilde{\Gamma}_{x,r} \leq \frac{8 \ln n}{n} \right) \leq 2 \exp \left(-\frac{3}{4} \sqrt{a' \ln n n r^D} \right) (1 + o(1)).$$

Apply again $\frac{n r^D}{\ln n} \rightarrow +\infty$ to have that

$$\mathbb{P} \left(|\varepsilon_1(x)| > \sqrt{\frac{32f_{\max}^{(0)} \ln n}{A^2\omega_D n r^D}} \mid E_{x,r} \text{ and } \tilde{\Gamma}_{x,r} \leq \frac{8 \ln n}{n} \right) = o(n^{-4}).$$

That, together with (18), concludes the proof of (17). \square

Bounds on $\varepsilon_2(x)$

Lemma 3. *Suppose that distribution belongs to the model \mathcal{M} .*

If S is convex, we have

$$\mathbb{P} \left(|\varepsilon_2(x)| > r f_{\max}^{(1)} \mid E_{x,r} \right) = 0. \quad (19)$$

And, if S has the R_0 -inside and outside rolling ball property, we have

$$\mathbb{P} \left(|\varepsilon_2(x)| > r \left(f_{\max}^{(1)} + \frac{f_{\max}^{(0)} \omega_{D-1}}{A R_0 \omega_D} \right) \mid E_{x,r} \right) = 0. \quad (20)$$

Proof. Recall that

$$\varepsilon_2(x) = \left(\frac{\tilde{\Gamma}_{x,r}}{|C_{x,r}|} - f(x) \right) \mathbb{I}_{E_{x,r}}$$

$$|\varepsilon_2(x)| = \frac{\left| \int_{C_{x,r}} (f(y) - f(x)) dy \right|}{|C_{x,r}|} \leq \frac{\left| \int_{C_{x,r} \cap S} (f(y) - f(x)) dy \right|}{|C_{x,r}|} + \frac{\left| \int_{C_{x,r} \setminus S} (f(y) - f(x)) dy \right|}{|C_{x,r}|}.$$

Thus,

$$|\varepsilon_2(x)| \leq r \left(\max_{z \in B(x,r) \cap S} \|\nabla f(z)\| \right) + f(x) \frac{|C_{x,r} \setminus S|}{|C_{x,r}|}.$$

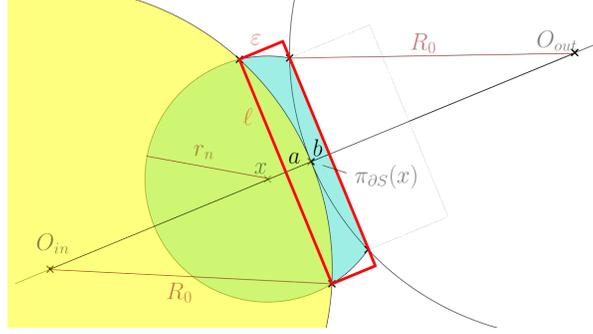


Figure 3: The volume of $H(B(x, r) \cap S) \setminus S$ is smaller than $\omega_{D-1} \ell^{D-1} \varepsilon$. Trivially $\ell \leq r_n$. By $R_0^2 = (R_0 - a)^2 + \ell^2$ and $r_n^2 = \ell^2 + (\|x - \pi_{\partial S}(x)\| - a)^2$, it comes that $a \leq \frac{r_n^2}{2R_0}$. Symmetrically we have $R_0^2 = (R_0 - b)^2 + \ell^2$ and $r_n^2 = \ell^2 + (\|x - \pi_{\partial S}(x)\| + b)^2$ and so $b \leq \frac{r_n^2}{2R_0}$. Finally $\varepsilon = a + b$.

Roughly bounding the second term given $E_{x,r}$, we then obtain

$$|\varepsilon_2(x)| \leq r \left(\max_{z \in B(x,r) \cap S} \|\nabla f(z)\| \right) + f(x) \frac{|H(B(x,r) \cap S) \setminus S|}{A \omega_D r^D}.$$

Obviously, if S is convex, we obtain

$$|\varepsilon_2(x)| \leq r \left(\max_{z \in B(x,r) \cap S} \|\nabla f(z)\| \right).$$

Now we are going to bound due to the outside rolling ball condition (see Figure 3 and its legend for some details):

$$|\varepsilon_2(x)| \leq r \left(\max_{z \in B(x,r) \cap S} \|\nabla f(z)\| \right) + r f(x) \frac{\omega_{D-1}}{AR_0 \omega_D}.$$

□

Lemma 4. *Suppose that distribution belongs to the model \mathcal{M} . For $x \in \mathring{S}$ and r small enough to have $B(x, r) \subset \{z, f(z) > 0\}$, introduce $H_{x,r} = \max_{z \in B(x,r)} \|\mathcal{H}_f(x)\|_{op}$, there exists an explicit constant C such that*

$$\mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) \leq \left(H_{x,r}^2 r^4 + Cr^2 \left(\left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{4}{D+1}} + r \left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2}{D+1}} \right) \right) (1 + o(1)). \quad (21)$$

And there exists explicit constants C, C' and C_2 such that

$$\mathbb{P} \left(|\varepsilon_2| \geq \frac{Cr \ln n}{nr^D \gamma(x)} + r \left(\frac{C'}{nr^D \gamma(x)} \right)^{\frac{2}{D+1}} + r^2 f_{\max}^{(2)} |E_{x,r}| \right) \leq C_2 n^{-4} + \exp \left(n(\delta r)^D \gamma(x)/10 \right), \quad (22)$$

where $\gamma(x) = \min(f(x), \varepsilon_0)$, ε_0 being the constant in the $(r_0, \delta, \varepsilon_0)$ -standardness of the distribution.

Proof. As $C_{x,r} \subset B(x, r) \subset S$, by the use of a second order Taylor expansion we have:

$$\left| \tilde{\Gamma}_{x,r} - f(x) |C_{x,r}| - \int_{C_{x,r}} (y-x) \nabla f(x)' dy \right| \leq \int_{C_{x,r}} \|y-x\|^2 \max_{z \in C_{x,r}} \|\mathcal{H}_f(z)\|_{\text{op}} dy.$$

Because $\int_{B(x,r)} (y-x) \nabla f(x)' dy = 0$, we have that:

$$\|\tilde{\Gamma}_{x,r} - f(x) |C_{x,r}|\| \leq \int_{B(x,r) \setminus C_{x,r}} r \|\nabla f(x)\| dy + |C_{x,r}| \max_{z \in C_{x,r}} \|\mathcal{H}_f(z)\|_{\text{op}} r^2,$$

from which we derive

$$\frac{\|\tilde{\Gamma}_{x,r} - f(x) |C_{x,r}|\|}{|C_{x,r}|} \leq r \|\nabla f(x)\| \frac{|B(x,r) \setminus C_{x,r}|}{|C_{x,r}|} + \max_{z \in C_{x,r}} \|\mathcal{H}_f(z)\|_{\text{op}} r^2.$$

Thus, given $|C_{x,r}| \geq A \omega_d r^D$,

$$|\varepsilon_2(x)| \leq \frac{\|\nabla f(x)\|}{A} r \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} + \max_{z \in C_{x,r}} \|\mathcal{H}_f(z)\|_{\text{op}} r^2. \quad (23)$$

To control the convergence rate of $|B(x,r) \setminus C_{x,r}|$ we will make use of Results in Brunel (2017) recalled in appendix A (see Proposition 3). To apply the aforementioned proposition let us introduce $M_{x,r} = \max_{y \in B(0,1)} \frac{r^D f(x+ry)}{\int_{B(x,r)} f(z) dz}$ (defined because $\int_{B(x,r)} f(z) dz > 0$).

Let us first consider the L^2 convergence. Because $\frac{\min_{B(x,r)} f(y)}{\max_{B(x,r)} f(y)} \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \leq \frac{\int_{B(x,r) \setminus C_{x,r}} f(z) dz}{\int_{B(x,r)} f(z) dz}$ and by application of 33, it comes that, for $k > 0$,

$$\mathbb{E} \left(\left(\frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \right)^q |N_{x,r} = k \right) \leq A_q (M_{x,r} + 1)^q \left(\frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} \right)^q k^{-\frac{2q}{D+1}},$$

Notice that $\mathbb{E} \left(\left(\frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \right)^q |N_{x,r} = 0 \right) = 1$. Then:

$$\mathbb{E} \left(\left(\frac{|B(x, r) \setminus C_{x,r}|}{|B(x, r)|} \right)^q \right) \leq A_q (M_{x,r} + 1)^q \left(\frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} \right)^q \sum_{k=1}^n \binom{n}{k} k^{-\frac{2q}{D+1}} \Gamma_{x,r}^k (1 - \Gamma_{x,r})^{n-k} + (1 - \Gamma_{x,r})^n.$$

A rough bound gives $M_{x,r} \leq \frac{1}{\omega_D} \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)}$. Introduce $B_{x,r,q} = A_q \left(\frac{1}{\omega_D} \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} + 1 \right)^q \left(\frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} \right)^q$. By application of Proposition 1, we have that

$$\mathbb{E} \left(\left(\frac{|B(x, r) \setminus C_{x,r}|}{|B(x, r)|} \right)^q \right) \leq B_{x,r,q} \left(\left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2q}{D+1}} + \exp \left(-\frac{n\Gamma_{x,r}}{10} \right) \right) + (1 - \Gamma_{x,r})^n. \quad (24)$$

And thus, because $\mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) \leq \mathbb{E}(\varepsilon_2(x)^2)$ ($\varepsilon_2(x)^2$ is a positive random variable), by use of (23), we obtain

$$\begin{aligned} \mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) \leq & \frac{\|\nabla f(x)\|^2}{A^2} r^2 \left(B_{x,r,2} \left(\left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{4}{D+1}} + \exp \left(-\frac{n\Gamma_{x,r}}{10} \right) \right) + (1 - \Gamma_{x,r})^n \right) \\ & + 2 \frac{\|\nabla f(x)\| H_{x,r}}{A} r^3 \left(B_{x,r,1} \left(\left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2}{D+1}} + \exp \left(-\frac{n\Gamma_{x,r}}{10} \right) \right) + (1 - \Gamma_{x,r})^n \right) \\ & + H_{x,r}^2 r^4. \end{aligned}$$

By application of Proposition 2 and because $(1 - \Gamma_{x,r})^n \leq \exp(-n\Gamma_{x,r})$ and $nr^D / \ln n \rightarrow +\infty$, we obtain the existence of an explicit constant C such that

$$\mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) \leq \left(H_{x,r}^2 r^4 + Cr^2 \left(\left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{4}{D+1}} + r \left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2}{D+1}} \right) \right) (1 + o(1)).$$

Let us second consider the probability convergence rate.

Introduce $B'_{x,r} = C_1 \left(\frac{1}{\omega_D} \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} + 1 \right) \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)}$. By use of Equation (32), we obtain

$$\mathbb{P} \left[k \left(\frac{|B(x, r) \setminus C_{x,r}|}{|B(x, r)|} - B'_{x,r} k^{-2/(D+1)} \right) > t \frac{\max_{B(x,r)} f(z)}{\min_{B(x,r)} f(z)} \mid N_{x,r} = k \right] \leq C_2 e^{-t}.$$

From which it comes that, by Proposition 2, there exists C and C' such that

$$\mathbb{P} \left[\left(\|\nabla f(x)\| \frac{|B(x, r) \setminus C_{x,r}|}{|B(x, r)|} \geq C \frac{t}{k} + C' k^{-\frac{2}{D+1}} \mid N_{x,r} = k \right) \mid E_{x,r} \right] \leq C_2 e^{-t}.$$

Thus, for all K

$$\mathbb{P}\left(\|\nabla f(x)\| \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \geq C \frac{t}{K} + C' K^{-\frac{2}{D+1}} |E_{x,r}\right) \leq C_2 e^{-t} + \mathbb{P}(N_{x,r} < K),$$

which, with (23), gives

$$\mathbb{P}\left(|\varepsilon_2| \geq C \frac{tr}{K} + C' r K^{-\frac{2}{D+1}} + r^2 f_{\max}^{(2)} |E_{x,r}\right) \leq C_2 e^{-t} + \mathbb{P}(N_{x,r} < K).$$

Finally, to deal with $\mathbb{P}(N_{x,r} < K)$, recall that $\gamma(x) = \min(f(x), \varepsilon_0)$ and choose $K = \frac{n\omega_D \delta^D r^D \gamma(x)}{2}$. Recall that we have $\Gamma_{x,r} \geq \omega_D (\delta r)^D \gamma(x)$ by the ball standardness assumption. Thus, by application of Bennets inequality ($h(0.5) \geq 0.1$) with the choice of $t = 4 \ln n$, it comes that the existence of (new) constants C and C' such that:

$$\mathbb{P}\left(|\varepsilon_2| \geq r \frac{C \ln n}{nr^D \gamma(x)} + r \left(\frac{C'}{nr^D \gamma(x)}\right)^{\frac{2}{D+1}} + r^2 f_{\max}^{(2)} |E_{x,r}\right) \leq C_2 n^{-4} + \exp\left(n(\delta r)^D \gamma(x)/10\right).$$

This achieves the proof of (22). \square

2.2.4 Proof of main Theorems and Corollaries

Proof of Theorem 1

Proof. By Equation (11)

$$\mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) \leq \mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) + 2\mathbb{E}(\varepsilon_1 \varepsilon_2 |E_{x,r}) + \mathbb{E}(\varepsilon_2^2(x)|E_{x,r}) + f(x)\mathbb{P}(E_{x,r}^c).$$

$$\text{By Equation (16) } \mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) \leq \frac{f(x)}{n|B(x,r)|} (1 + o_x(1)).$$

$$\text{By Equation (14) } \mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|E_{x,r}) = 0.$$

By Equation (21)

$$\mathbb{E}(\varepsilon_2(x)^2|E_{x,r}) \leq \left(H_{x,r}^2 r^4 + C r^2 \left(\left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{4}{D+1}} + r \left(\frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2}{D+1}} \right) \right) (1 + o(1)).$$

Finally, by Equation (13), if $x \in L_t$,

$$P(E_{x,r}^c) \leq (C_D + 1) C_D 4^D \left(1 - \frac{A^{1/D}}{\delta} \right)^{-D} \exp(-\omega_D \alpha n r^D t).$$

Because $x \in L_{f(x)}$, we have that

$$f(x)\mathbb{P}(E_{x,r}^c) \leq f_{\max}^{(0)} (C_D + 1) C_D 4^D \left(1 - \frac{A^{1/D}}{\delta} \right)^{-D} \exp(-\omega_D \alpha n r^D f(x)).$$

Because $f(x) > 0$ and $nr^D \rightarrow +\infty$, it comes that $f(x)\mathbb{P}(E_{x,r}^c) = o_x((nr^D)^{-1})$. Also, due to $nr^D \gg \ln n$, it comes that for all $\beta > 0$, $f(x)\mathbb{P}(E_{x,r}^c) = o_x(n^{-\beta})$. As we also have $r \geq (\ln n/n)^{1/D}$, we clearly get $f(x)\mathbb{P}(E_{x,r}^c) = o_x(n^{-4})$, and thus:

$$f(x)\mathbb{P}(E_{x,r}^c) = o_x(\max((nr^D)^{-1}, r^4)).$$

That achieves the proof of the first point.

Second point of Theorem 1 is proved the same by use of Equation (15) instead of (16).

Final point of Theorem 1 comes from the use of (15) instead of (16) and (19) or (20) instead of (21). \square

Proof of Corollary 1

Proof. By Theorem 1,

$$\begin{aligned} \mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) &\leq \left(\frac{f(x)}{n\omega_D r^D} + \|\mathcal{H}_f(x)\|_{op}^2 r^4 \right. \\ &\quad \left. + Cr^2 \left(\frac{\omega_D f(x) nr^D}{2} \right)^{-\frac{4}{D+1}} + Cr^3 \left(\frac{\omega_D f(x) nr^D}{2} \right)^{-\frac{2}{D+1}} \right) (1 + o_x(1)). \end{aligned}$$

Recall that classical computation on the uniform kernel gives $\mathbb{E}(f_r^*(x) - f(x))^2 = \left(\frac{f(x)}{n\omega_D r^D} + \|\mathcal{H}_f(x)\|_{op}^2 r^4 \right) (1 + o_x(1))$.

The choice of $r = cn^{-1/(D+4)}$ gives that $\mathbb{E}(f_r^*(x) - f(x))^2 \geq C(x)n^{-4/(D+4)}$, with $C(x) > 0$, since $f(x) > 0$. Also for $r = cn^{-1/(D+4)}$, we have:

$$\begin{cases} r^2 \left(\frac{\omega_D f(x) nr^D}{2} \right)^{-\frac{4}{D+1}} = O\left(n^{-\frac{4}{D+4} + \frac{2(D-7)}{(D+1)(D+4)}} \right) = O\left(n^{-\frac{2(D+9)}{(D+1)(D+4)}} \right), \\ r^3 \left(\frac{\omega_D f(x) nr^D}{2} \right)^{-\frac{2}{D+1}} = O\left(n^{-\frac{4}{D+4} + \frac{(D-7)}{(D+1)(D+4)}} \right) = O\left(n^{-\frac{3D+11}{(D+1)(D+4)}} \right). \end{cases}$$

That concludes the proof of the first part of Corollary 1.

Now by use of Equations (11), (14), (15) and Lemma 3, with $r = cn^{-1/(D+4)}$, it comes that there exists C such that $\mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) = Cn^{-2/(D+4)} + f(x)\mathbb{P}(E_{x,r}^c)$. Now because $f(x)\mathbb{P}(E_{x,r}^c) \leq \beta f(x) \exp(-\beta' n^{\frac{4}{D+4}} f(x))$ (due to Equation (13)), we have that $f(x)\mathbb{P}(E_{x,r}^c) = o_{f(x)}(n^{-\frac{4}{D+4}})$. And if f is lower bounded on S , we have that $f(x)\mathbb{P}(E_{x,r}^c) = o(n^{-\frac{4}{D+4}})$. That concludes the proof of the second part of Corollary 1. \square

Proof of Theorem 2

Proof. Firstly focus on the first case. Let us fix $t > 0$. For all $x \in L_t$ such that $B(x, r) \subset S$, due to Lemma 1, we have:

$$\mathbb{P}(E_{x,r}^c) \leq \beta_1 \exp\left(-\beta_2 nr^D \min(t, \varepsilon_0)\right) + \exp(-n/8) = o_t(n^{-4}).$$

Also due to Equation (17), we have

$$\mathbb{P}\left(|\varepsilon_1(x)| > \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D nr^D}} |E_{x,r}\right) \leq 2n^{-4}(1 + o(1)).$$

Finally, Equation (22) for $r = cn^{-\frac{1}{D+4}}$ and $D \leq 7$ gives

$$\mathbb{P}(|\varepsilon_2| \geq O_t(r^2) |E_{x,r}) \leq C_2 n^{-4} + \exp(nr^D \min(t, \varepsilon_0)/10).$$

Thus, due to $\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t_1 + t_2) \leq \mathbb{P}(|\varepsilon_1(x)| \geq t_1 |E_{x,r}) + \mathbb{P}(|\varepsilon_2| \geq t_2 |E_{x,r}) + \mathbb{P}(E_{x,r}^c)$,

$$\mathbb{P}\left(|\hat{f}_{r,A}(x) - f(x)| \geq \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D nr^D}} + O_t(r^2)\right) \leq Cn^{-4}(1 + o_t(1)).$$

Introduce $\mathbb{X}(t, r) = \{X_i \in \mathbb{X}_n, f(X_i) > t, B(X_i, r) \subset S\}$. Recall that $r = cn^{-\frac{1}{D+4}}$. All the above equations provide the existence of a constant C_t such that, for n large enough,

$$\mathbb{P}\left(\exists X_i \in \mathbb{X}(t, r) \text{ s.t. } |\hat{f}_{r,A}(X_i) - f(X_i)| \geq C_t \sqrt{\ln n} n^{-\frac{2}{D+4}}\right) \leq Cn^{-3}(1 + o(1)).$$

That, with Borrel Cantelli Lemma, concludes the proof of the first point of Theorem 2.

The second point of Theorem 2 is the same using the lower bound of the density to obtain uniform rates and is let to the reader.

We now prove the third point of Theorem 2. First suppose that $f(x) \geq (\ln n / (nr^D))^{1/2}$. Then, by Lemma 1 and because $\frac{nr^D}{\ln n} \rightarrow +\infty$,

$$\mathbb{P}(E_{x,r}^c) \leq \beta_1 \exp\left(-\beta_2 \ln n \sqrt{\frac{nr^D}{\ln n}}\right) + \exp(-n/8) = o(n^{-4})$$

Also due to Equation (17),

$$\mathbb{P}\left(|\varepsilon_1(x)| > \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D nr^D}} |E_{x,r}\right) \leq 2n^{-4}(1 + o(1)).$$

And, under the rolling ball condition or the convexity condition, due to Lemma 3, there exists C such that

$$\mathbb{P}(|\varepsilon_2| > Cr | E_{x,r}) = 0.$$

Thus, due to $\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t_1 + t_2) \leq \mathbb{P}(|\varepsilon_1(x)| \geq t_1 | E_{x,r}) + \mathbb{P}(|\varepsilon_2| \geq t_2 | E_{x,r}) + \mathbb{P}(E_{x,r}^c)$,

$$\mathbb{P}\left(|\hat{f}_{r,A}(x) - f(x)| \geq \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} + Cr\right) \leq 2n^{-4}(1 + o(1)).$$

Second, suppose that $0 < f(x) \leq (\ln n / (nr^D))^{1/2}$. Because, conditionally to $E_{x,r}^c$, $\hat{f}_{r,A}(x) = 0$, we have $\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| > \ln n / (nr^D))^{1/2} | E_{x,r}^c) = 0$. It comes that, when $t \geq 2(\ln n / (nr^D))^{1/2}$,

$$\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t) \leq \mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t | E_{x,r}).$$

Thus, applying the same calculus as in the first case,

$$\mathbb{P}\left(|\hat{f}_{r,A}(x) - f(x)| \geq \sqrt{\frac{\ln n}{nr^D}} \max\left(\sqrt{\frac{32 f_{\max}^{(0)}}{A^2 \omega_D}}, 2\right) + Cr\right) \leq 2n^{-4}(1 + o(1)).$$

Thus, for all x in S ,

$$\mathbb{P}\left(|\hat{f}_{r,A}(x) - f(x)| \geq \sqrt{\frac{\ln n}{nr^D}} \max\left(\sqrt{\frac{32 f_{\max}^{(0)}}{A^2 \omega_D}}, 2\right) + Cr\right) \leq 2n^{-4}(1 + o(1)).$$

Thus, when $r = cn^{-\frac{1}{D+4}}$,

$$\mathbb{P}\left(\exists X_i \text{ s.t. } |\hat{f}_{r,A}(X_i) - f(X_i)| \geq Cr(1 + o(1))\right) \leq 2n^{-3}(1 + o(1)).$$

That, with Borrel Cantelli Lemma, concludes the proof of last point of the Theorem. \square

Proof of Corollary 2

Proof. Recall the local convex hull definition: for a set E ,

$$H_r(E) := \bigcup_{x \in E} H(B(x, r) \cap E).$$

Suppose first that $L_t \subset \mathring{S}$. The regularity of the level set L_t and the regularity of the density imply that L_t has the $\frac{m_t}{f_{\max}^{(2)}}$ -inside and outside rolling ball condition (see Theorem 2 in Walther (1997)). By easy Taylor majoration it also comes that for all Δ_t with $|\Delta_t| \leq \frac{m_t}{4f_{\max}^{(1)}f_{\max}^{(2)}}$, $L_{t+\Delta_t}$ has the $\frac{m_t}{2f_{\max}^{(2)}}$ -inside and outside rolling ball condition.

Notice now that there exists $\alpha_t > 0$ such that, for all $0 < \varepsilon \leq \alpha_t$, $L_{t-\varepsilon} \subset \mathring{S}$. Indeed, if there exists a sequence $\varepsilon_n \rightarrow 0$ with $\varepsilon_n < t/2$, $L_{t-\varepsilon_n} \cap \partial S \neq \emptyset$, since $L_{t-\varepsilon_n} \cap \partial S$ is a decreasing (for the inclusion) sequence of non empty compacts, we have $L_t \cap \partial S \neq \emptyset$. And that contradicts $L_t \subset \mathring{S}$.

We now prove that, for all $0 < \varepsilon < \min\left(\alpha_t, \frac{m_t d(L_{t-\alpha_t}, \partial S)}{2}, \frac{m_t^2}{2f_{\max}^{(2)}}\right)$, we have that

$$L_{t-\varepsilon} \subset L_t \oplus \frac{2\varepsilon}{m_t} B \text{ and } d(L_{t-\varepsilon}, \partial S) \geq \alpha_t - \frac{2\varepsilon}{m_t}. \quad (25)$$

Indeed, consider $x \in L_{t-\varepsilon}$ and $y = x + 2\frac{\varepsilon}{m_t} \frac{\nabla f(x)}{\|\nabla f(x)\|}$. We have that $x \in L_{t-\alpha_t}$ and $\|y - x\| < d(L_{t-\alpha_t}, \partial S)$, thus $y \in \mathring{S}$. That allows to write that

$$f(y) \geq f(x) + 2\frac{\varepsilon}{m_t} \|\nabla f(x)\| - \frac{1}{2} \left(2\frac{\varepsilon}{m_t}\right)^2 f_{\max}^{(2)}.$$

We have $f(y) \geq t - \varepsilon + 2\varepsilon - \varepsilon \frac{2\varepsilon f_{\max}^{(2)}}{m_t^2}$. Thus, since $\varepsilon < m_t^2 / (2f_{\max}^{(2)})$, we also have $f(y) \geq t$ and thus $y \in L_t$. That, together with $\|y - x\| = \frac{2\varepsilon}{m_t}$, concludes the proof of the first part of (25). It also implies that, when $\frac{2\varepsilon}{m_t} \leq \frac{m_t}{f_{\max}^{(1)}}$, we have $d(L_t, L_{t-\varepsilon}) \leq \frac{2\varepsilon}{m_t}$. Condition $\varepsilon < \alpha_t$, $d(L_t, \partial S) \geq \alpha_t$ and triangular inequality concludes the proof of (25).

By last point of Theorem 2: with probability one for n large enough, for all X_i such that $\hat{f}(X_i) \geq t$ we have $f(X_i) \geq t - Cn^{-\frac{1}{D+4}}$.

As $X_i \in L_{t - Cn^{-\frac{1}{D+4}}}$, by distance inequality in (25) we have that, with probability one for n large enough, for all X_i such that $\hat{f}(X_i) \geq t$, $B(X_i, r) \subset S$.

Also, for n large enough to have $Cn^{-\frac{1}{D+4}} < t/2$, from Theorem 2 (first point), with probability one for n large enough, for all X_i such that $\hat{f}(X_i) \geq t$, we have $|f(X_i) - \hat{f}(X_i)| \leq C_{t/2} \sqrt{\ln nn}^{-\frac{2}{D+4}}$. Thus $X_i \in L_{t-\varepsilon_n}$ with $\varepsilon_n = C_{t/2} \sqrt{\ln nn}^{-\frac{2}{D+4}}$. We then have, with probability one for n large enough:

$$\hat{L}_t \subset H_r(\mathbb{X}_n \cap L_{t-\varepsilon_n}) \subset H_r(L_{t-\varepsilon_n}).$$

Recall that $L_{t-\varepsilon_n}$ satisfies the $\frac{m_t}{2f_{\max}^{(1)}}$ -inside and outside rolling ball condition of then using a purely geometric calculus as in Figure 3 we obtain that (with probability one for

n large enough)

$$\hat{L}_t \subset L_{t-\varepsilon_n} \oplus \frac{4r^2 f_{\max}^{(1)}}{m_t} B \subset L_t \oplus \left(\frac{4r^2 f_{\max}^{(1)} + 2\varepsilon_n}{m_t} \right) B \quad (26)$$

Now we consider reverse inclusion. Introduce $\varepsilon'_n = C_t \sqrt{\ln n} n^{-\frac{2}{D+4}}$, same kind of arguments successively give that, with probability one for n large enough, $\mathbb{X}_n \cap L_{t+\varepsilon'_n} \subset \mathbb{X}_{r,A}^+(t)$. Thus for any c , with probability one for n large enough:

$$H_{c\left(\frac{\ln n}{n}\right)^{\frac{1}{D+1}}}(\mathbb{X}_n \cap L_{t+\varepsilon'_n}) \subset H_r(\mathbb{X}_n \cap L_{t+\varepsilon'_n}) \subset \hat{L}_t$$

Because $L_{t+\varepsilon'_n}$ has the $\frac{m_t}{2f_{\max}^{(1)}}$ -inside and outside rolling ball condition, and because $t > 0$, by application of Aaron and Bodart (2016) (Theorem 2 with $\alpha = 0$) it comes that, there exists a constant C'_t such that, with probability one for n large enough:

$$L_{t+\varepsilon'_n} \subset \hat{L}_t \oplus C'_t \left(\frac{\ln n}{n} \right)^{\frac{2}{D+1}} B;$$

Thus again by application of (25) we have

$$L_t \subset \hat{L}_t \oplus \left(C'_t \left(\frac{\ln n}{n} \right)^{\frac{2}{D+1}} + \frac{2\varepsilon'_n}{m_t} \right) B$$

which, together with Equation (26) achieves the proof of the first point of corollary 2.

The more general case where $L_t \cap \partial S \neq \emptyset$.

Let us here only sketch the proof, since the arguments are really similar to those of the first part. Put $\varepsilon_n = Cn^{-\frac{1}{D+4}}$, such that $f(X_i) \geq t - \varepsilon_n \Rightarrow \hat{f}_{r,A}(X_i) \geq t$ and $\hat{f}_{r,A}(X_i) \geq t \Rightarrow f(X_i) \geq t + \varepsilon_n$. We have that $\hat{L}_t \subset H_r(L_{t-\varepsilon_n}) \subset L_{t-\varepsilon_n} \oplus rB$ and regularity of the level set gives $\hat{L}_t \subset L_t \oplus ((\beta+1)r)B$. Reversely, $\mathbb{X}_n \cap L_{t+\varepsilon_n} \subset L_t$ and, since $L_{t+\varepsilon_n} \subset L_t$ has the δ -ball standardness, for all $d_H(\mathbb{X}_n \cap L_{t+\varepsilon_n}, L_{t+\varepsilon_n}) \leq O_t((\ln n/n)^{1/D})$. Then using again the regularity of the level set t we can conclude the proof. \square

Proof of Theorem 3

Proof. First, under the assumptions of Theorem 3 (i.e. the distribution belongs to the model, the support has the inside and outside rolling ball property and the density is bounded below by a positive constant), then with probability one for n large enough, for

all $x \in S$, $E_{x,r}$ is realised and as a consequence $\tilde{\Gamma}_{x,r} \geq Af_{\min}^{(0)}\omega_D r^D$. The proof is very similar to the one of Lemma 1.

For $z \in \partial S$, let denote by η_z the normal outward unit vector to S . For all $x \in S$, define

$$y = \begin{cases} x & \text{if } d(x, \partial S) \geq r, \\ \pi_{\partial S}(x) - \frac{r+d(x, \partial S)}{2}\eta_{\pi_{\partial S}(x)} & \text{if } d(x, \partial S) < r, \end{cases}$$

and

$$\rho_n = \begin{cases} r & \text{if } d(x, \partial S) \geq r, \\ \pi_{\partial S}(x) - \frac{r+d(x, \partial S)}{2}\eta_{\pi_{\partial S}(x)} & \text{if } d(x, \partial S) < r \end{cases}.$$

The regularity condition on S warranties that $B(y, \rho_n) \subset S \cap B(x, r)$. As in Lemma 1 (illustrated in Figure 2), if there exists an x such that $B(y, \rho_n - c(\ln n/n)^{1/D}) \not\subset C_{x,r}$, then there exists z with $B(z, c(\ln n/n)^{1/D}) \subset S$ and $B(z, c(\ln n/n)^{1/D}) \cap \mathbb{X}_n = \emptyset$ has a probability null for n large enough when c is a large enough constant. Since $\rho_n \geq r/2$ and $r \gg (\ln n/n)^{1/D}$, for all $A < 1/2$ with probability one, for n large enough and for all $x \in S$, we have that $C_{x,r} \geq A\omega_D r^D$ and the condition on the density also guarantees that $\tilde{\Gamma}_{x,r} \geq Af_{\min}^{(0)}\omega_D r^D$.

Now, let cover S with x_i, \dots, x_ν balls of radius r with $\nu \leq Cr^{-D}$. Suppose that there exists $x \in S$ such that $N_{x,r}^\partial > n/2$. Then $N_{x,r} > n/2 > 2^{D+1}nf_{\max}^{(0)}\omega_D r^D$ and there exists x_i with $N_{x_i,2r} > 2^{D+1}nf_{\max}^{(0)}\omega_D r^D$. Thus,

$$\mathbb{P}(\exists x \in S, N_{x,r}^\partial > n/2) \leq \sum_i \mathbb{P}(N_{x_i,2r} > 2^{D+1}f_{\max}^{(0)}\omega_D nr^D),$$

$$\mathbb{P}(\exists x \in S, N_{x,r}^\partial > n/2) \leq \sum_i \mathbb{P}(N_{x_i,2r} - n\Gamma_{x_i,2r} > n(2^{D+1}f_{\max}^{(0)}\omega_D r^D - \Gamma_{x_i,2r})),$$

$$\mathbb{P}(\exists x \in S, N_{x,r}^\partial > n/2) \leq \sum_i \mathbb{P}(N_{x_i,2r} - n\Gamma_{x_i,2r} > nf_{\max}^{(0)}\omega_D r^D(2^{D+1} - 2^D)).$$

By use of Bennett's inequality

$$\mathbb{P}(\exists x \in S, N_{x,r}^\partial > n/2) \leq \sum_i \exp\left(-n\Gamma_{x_i,2r}h\left(\frac{2^D f_{\max}^{(0)}\omega_D r^D}{\Gamma_{x_i,2r}}\right)\right).$$

And, finally, because h is increasing:

$$\mathbb{P}(\exists x \in S, N_{x,r}^\partial > n/2) \leq Cr^{-D} \exp\left(-nf_{\min}^{(0)}\omega_D 2^D r^D h(1)\right)$$

Thus, for $r = cn^{-\frac{1}{D+4}}$, by Borrel Cantelli, we have that, with probability one, for n large enough and for all $x \in S$, $E_{x,r}$ is realized. The bound on $\mathbb{P}(\exists x \in S, N_{x,r}^\partial > n/2)$ may

seem rough. But it will be re-employed since, in fact, we have shown that, with probability one, for n large enough and for all $x \in S$, $N_{x,r} \leq 2^{D+1} f_{\max}^{(0)} \omega_D n r^D$. That, in turns, allow to bound the number of different values of $\hat{f}_{r,A}(x)$ for $x \in S$. Indeed there is as many different values as there are non empty sets

$$S_I = \left(\bigcap_{i \in I} B(X_i, r) \right) \cap \left(\bigcap_{i \in I^c} B(X_i, r)^c \right) \cap S,$$

where $I \subset \{1, \dots, n\}$ and $I^c = \{1, \dots, n\} \setminus I$, because if $x \in S_I$ then $\mathbb{X}_n \cap B(x, r) = \{X_i, i \in I\}$. Now, with probability one, for n large enough and for all $x \in S$, there exists i such that $x \in B(X_i, r)$ (because $r \gg (\ln n/n)^{1/D}$) and $B(X_i, r)$ may intersect at most $2^{2^{2D+1} f_{\max}^{(0)} \omega_D r^D} B(X_j, r)$ (following the previous calculus). Thus with probability one, for n large enough,

$$\#S_I \leq n 2^{2^{2D+1} f_{\max}^{(0)} \omega_D n r^D}.$$

Define Y_1, \dots, Y_K arbitrary points in each non empty S_I . We have that:

$$\max_{x \in S} |\hat{f}_{r,A}(x) - f(x)| = \max_{i, x \in S_{I_i}} |\hat{f}_{r,A}(Y_i) - f(x)| \leq \max_i |\hat{f}_{r,A}(Y_i) - f(Y_i)| + \max_{i, x \in S_{I_i}} |f(x) - f(X_i)|.$$

Because S_I as a diameter bounded by $2r$, due to f regularity,

$$\max_{x \in S} |\hat{f}_{r,A}(x) - f(x)| = \max_{i, x \in S_{I_i}} |\hat{f}_{r,A}(Y_i) - f(x)| \leq \max_i |\hat{f}_{r,A}(Y_i) - f(Y_i)| + 2r f_{\max}^{(1)}. \quad (27)$$

Now, by Hoeffding inequality,

$$\mathbb{P} \left(|N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^o)| > \sqrt{C' n^{\frac{D}{2(D+4)}}} |C_{x,r} \right) \leq 2 \exp \left(-C' n^{\frac{4}{D+4}} \right),$$

which, adapted to our case ($r = cn^{-\frac{1}{D+4}}$), gives:

$$\mathbb{P} \left(|\varepsilon_1(x)| > \frac{2\sqrt{C'}}{Ac^D \omega_D} n^{\frac{D-8}{2(D+4)}} |E_{x,r} \right) \leq 2 \exp \left(-C' n^{\frac{4}{D+4}} \right).$$

Finally, as we can apply Lemma 3, it comes that:

$$\mathbb{P} \left(|\hat{f}_{r,A}(Y_i) - f(Y_i)| \geq \frac{2\sqrt{C'}}{Ac^D \omega_D} n^{\frac{D-8}{2(D+4)}} + Ccn^{-\frac{1}{D+4}} |E_{Y_i,r} \right) \leq 2 \exp \left(-C' n^{\frac{4}{D+4}} \right).$$

Which, with Equation (27) and the given r sequence, gives:

$$\mathbb{P} \left(\max_{x \in S} |\hat{f}_{r,A}(x) - f(x)| \geq \frac{2\sqrt{C'}}{Ac^D \omega_D} n^{\frac{D-8}{2(D+4)}} + (C + 2f_{\max}^{(1)})cn^{-\frac{1}{D+4}} \right) \leq 2n \exp \left(-n^{\frac{4}{D+4}} (C' - 2^{D+1} f_{\max}^{(0)} c \ln 2) \right) + \mathbb{P}(\exists x \in S, E_{x,r}^c).$$

□

3 Density on manifold

Suppose now that, if the observations are in \mathbb{R}^D , the distribution is in fact supported by S a d -dimensional manifold with $d < D$. Suppose that, for all $x \in S$, we know \hat{T}_x an estimation of $T_x S$ the tangent space at x . Recall the notations given in introduction: $\pi_{\hat{T}_x}$ is the orthogonal projection onto \hat{T}_x , $\mathbb{X}_{n,R} = \pi_{\hat{T}_x}(B(x, R) \cap \mathbb{X}_n)$, $N_{x,R} = \#\{\mathbb{X}_n \cap B(x, R)\}$, $\widehat{C}_{x,r,R} = H(\mathbb{X}_{n,R} \cap B(x, r))$, $\widehat{N}_{x,r,R} = \#\{\mathbb{X}_{n,R} \cap B(x, r)\}$, $\widehat{N}_{x,r,R}^\partial = \#\{\mathbb{X}_{n,R} \cap \partial \widehat{C}_{x,r,R}\}$ and $\widehat{N}_{x,r,R}^o = \widehat{N}_{x,r,R} - \widehat{N}_{x,r,R}^\partial$.

We propose the natural extensions of the density estimator in this manifold context which is given by:

$$\hat{f}_{r,R,A}(x) = \frac{N_{x,R}}{n} \frac{\widehat{N}_{x,r,R}^o}{(N_{x,R} - \widehat{N}_{x,r,R}^\partial) |\widehat{C}_{x,r,R}|} \mathbb{I}_{|\widehat{C}_{x,r,R}| \geq A\omega_d r^d} \mathbb{I}_{\widehat{N}_{x,r,R}^\partial \leq N_{x,R}/2}.$$

As in the full dimensional case we need to make assumptions on the support and the density. Regarding support we will, classically, characterize its smoothness by its reach.

Definition 5. Let S be a compact set. Its medial axis is defined by

$$\text{Med}(S) = \{z \in \mathbb{R}^D, \exists x \neq y \in S, \|x - z\| = \|y - z\| = d(z, S)\}.$$

Its reach τ_S is

$$\tau_S = \min_{x \in S} d(x, \text{Med}(S)).$$

Manifolds with positive reach are \mathcal{C}_1^1 and, due to Federer (1959) and Federer (1959), positive reach allow to make calculus on manifolds with few need of differential geometry. Our model in the manifold setting will only focus on supports that are manifolds with positive reach. We may deal with manifolds with or without boundary. Nevertheless, we will restrict the study to boundary with positive reach.

Assumptions on the density will be a bit more restrictive than in the full dimensional case. Beyond the \mathcal{C}^2 regularity on the interior which is expressed in (28), we will restrict the study to lower bounded densities.

Definition 6. A distribution belongs to the family model $\mathcal{M}'_d = \mathcal{M}'_d(\tau_{\min}, \tau_{\partial, \min}, f_{\min}^{(0)}, f_{\max}^{(0)}, \mathcal{G}_{S, f}, H_{S, f})$, where $\tau_{\min}, \tau_{\partial, \min}, f_{\min}^{(0)}, f_{\max}^{(0)}, \mathcal{G}_{S, f}$ and $H_{S, f}$ are positive constants, if:

1. S is a \mathcal{C}^2 compact d -dimensional manifold with positive reach $\tau_S \geq \tau_{\min}$. If exists, its boundary ∂S is a $(d-1)$ -manifold without boundary and with reach $\tau_{\partial S} \geq \tau_{\partial, \min}$.
2. Its density f satisfies that $f_{\min}^{(0)} \leq f(x) \leq f_{\max}^{(0)}$ for all $x \in S$. Also, there exist constants $\mathcal{G}_{S, f}$ and $H_{S, f}$, and there exist vectors $G_x \in T_x S$ such that, for all $(x, y) \in S^2$

$$\|f(y) - f(x) - G_x \cdot (y - x)\| \leq H_{S, f} \|y - x\|^2, \text{ and } \sup_{x \in S} \|G_x\| \leq \mathcal{G}_{S, f} \quad (28)$$

The following Theorem gives the convergence rates of the density estimator. Notice first that the O does not depend on x (as in point 2. and 3. in the lower bounded density case given in Theorem 1). The manifold case just add an error term that depends on the tangent space approximation quality Θ_n , classically characterized by $\angle \hat{T}_x, T_x S$, the angle between the tangent space and its approximation, which is also defined as the operator norm of $\pi_{T_x S} - \pi_{\hat{T}_x}$.

Theorem 4. *If the distribution belongs to the model \mathcal{M}'_d . Let the window size r_n be a sequence such that $r_n \rightarrow 0$ and $nr_n^d / \ln n \rightarrow +\infty$. Let $A < 4^{-d}$ be a positive constant. Suppose that for all x , we have $\angle T_x S, \hat{T}_x \leq \Theta_n$ with $\Theta_n \rightarrow 0$.*

1. For all x such that $d(x, \partial S) \geq 5r_n/4$, we have

$$\begin{aligned} \mathbb{E}((\hat{f}_{r, R, A}(x) - f(x))^2) = \\ O\left((nr_n^d)^{-1} + \left((r_n + \Theta_n)^2 + (r_n + \Theta_n^2)(nr_n^d)^{-\frac{2}{d+1}}\right)^2 + \max(n^{-1}, \Theta_n^4)\right). \end{aligned}$$

2. For all $x \in S$,

$$\mathbb{E}((\hat{f}_{r, R, A}(x) - f(x))^2) = O\left((nr_n^d)^{-1} + (r_n + \Theta_n^2)^2 + \max(n^{-1}, \Theta_n^4)\right).$$

A nice corollary of previous theorem is that, for optimal windows size $r_n = cn^{-1/(d+4)}$ and the usual error rates in tangent space estimation with local PCA which are of order $(\ln n/n)^{1/d}$ (see Aamari et Al (2023)), the Θ_n dependant part of the convergence rates can be neglected.

Corollary 3. *If the distribution belongs to the model \mathcal{M}'_d . Let the window size $r_n = cn^{-1/(d+4)}$. Let $A < 4^{-d}$ be a positive constant. Suppose that $\Theta_n \leq (\ln n/n)^{1/d}$ and $d \leq 7$, then:*

1. For all x such that $d(x, \partial S) \geq 5r_n/4$, we have $\mathbb{E}((\hat{f}_{r,R,A}(x) - f(x))^2) = O(n^{-4/(d+4)})$.

2. For all $x \in S$, we have $\mathbb{E}((\hat{f}_{r,R,A}(x) - f(x))^2) = O(n^{-2/(d+4)})$.

We also have a uniform (on the observation) convergence result :

Theorem 5. *If the distribution belongs to the model \mathcal{M}'_d . Let the window size $r_n = cn^{-1/(d+4)}$. Let $A < 4^{-d}$ be a positive constant. Suppose that $\Theta_n \leq (\ln n/n)^{1/d}$ and $d \leq 7$. Then, there exists C such that, with probability one, for n large enough,*

$$\max_i \left\{ \left| \hat{f}_{r,R,A}(X_i) - f(X_i) \right|, d(X_i, \partial S) \geq 5r_n/4 \right\} \leq C\sqrt{\ln nn}^{-2/(d+4)},$$

and

$$\max_i \left\{ \left| \hat{f}_{r,R,A}(X_i) - f(X_i) \right| \right\} \leq Cn^{-1/(d+4)}.$$

4 Numerical Experimentation's

4.1 On simulated data

First, we propose to test the local convex hull density estimator on a sample drawn according to a Gaussian restricted to the square $[0, 1]^2$. Namely, in our first example the density is $f(x, y) = c \exp(-2(x^2 + y^2))$. In Figure 4, we present a comparison of the local convex hull density estimator \hat{f} and f^* the classical kernel one on a sample of size $n = 5000$. On the left (a) and (b) panels, the results associated to the local convex hull density estimator ($r = 0.18$ and $A = 0.1$) are presented. The top figure (a) present the sample cloud with a color that is related to the rank of the estimated density value (isopleth presentation). The bottom figure (b) present the associated level sets estimated with the local convex hull method (also $r = 0.18$) for levels in $\{0, 0.5, 1, 1.5, 2, 2.5\}$. On the same figure, (c) and (d) illustrates the results with the density estimator f^* (and the same window size).

On figure 4, we can observe that, as expected, the local convex hull correction improves the density estimation near the boundary of the support.

To investigate a bit further this improvement we also tested the proposed method on a uniform drawn on the unit disk. Results are presented in Figure 5. We tried different sample sizes $n \in \{1000, 2000, 3000, 5000, 10000\}$. Tested parameters are $r_n = 0.5n^{-1/6}$ and $A = 1/8$. For each sample sizes we ran each time 100 replications and, for each replication k compute $Q_\alpha(k)$, the α -quantile (on i) of $|\hat{f}_{r_n,A}(X_i) - f(X_i)|$. We then present in blue plain line the median (on k) of $Q_\alpha(k)$, in blue dashed line we present the 25 and 75 percent quantiles (on k) of $Q_\alpha(k)$ and blue points are the max and min (on k) of $Q_\alpha(k)$. The red curves present the same quantities computed on $f^*_{r_n}$ (classical uniform kernel method with the same radius). Because the local convex hull corrections acts close to the boundary i.e. on few data points) we can observe difference between blue and red curves for high values

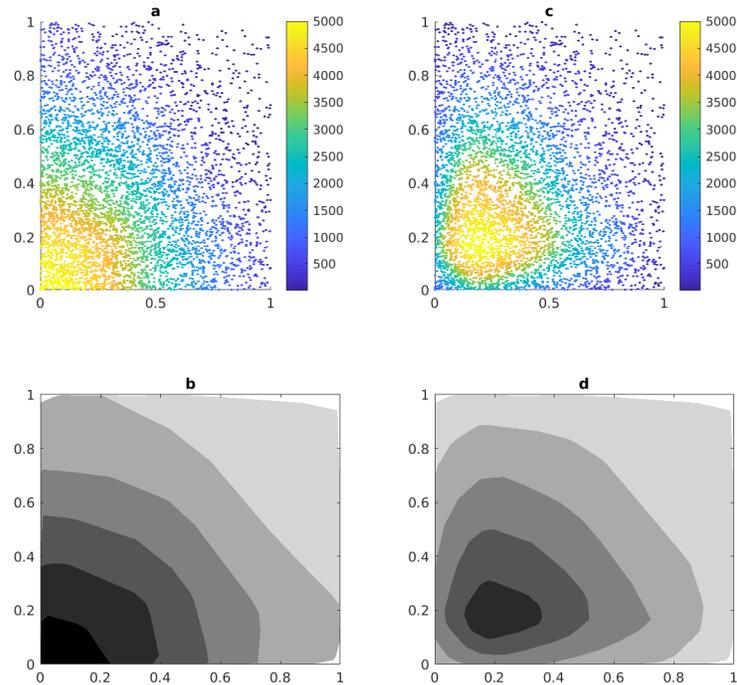


Figure 4: Comparison of density estimation methods for a gaussian restricted to a square. (a) and (b) our estimator compared to the usual one in (c) and (d). Top graphs : sample cloud colored with the estimated density rank value ; bottom graphs : level sets estimated with the local convex hull method for levels in 0, 0.5, 1, 1.5, 2, 2.5.

of α (100%, 99% and 95%) and, quickly (in n) the corrected estimator is better than the classical kernel one. For "small" quantiles (75% and 50%) we can observe the similarity between the two estimators that reflects the similarity between the two estimators in \hat{S} .

4.2 Example on real data

4.2.1 Comparison with the Original Paper

In this part we aim at making a comparison between the original Getz and Wilmer (2004) method and our proposed correction. As a tribute to Getz and Wilmer, we test the two approaches on core-area estimation of Panther Jitter (data base downloaded from PennState Department of Ecosystem Science and Management). We present in Figure 6 the $x\%$ isopleth ($x \in \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90\}$) on the left with application of the

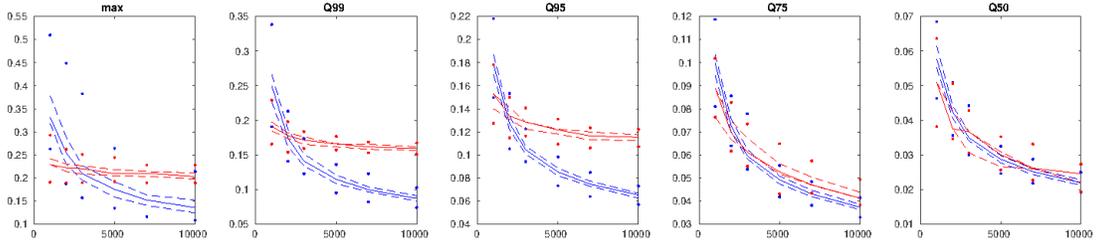


Figure 5: Comparison of density estimation methods for samples uniformly drawn on a disk. Computation of different quantiles of $|\hat{f}_{r_n, A}(X_i) - f(X_i)|$ with the estimator we propose in blue and the usual one in red.

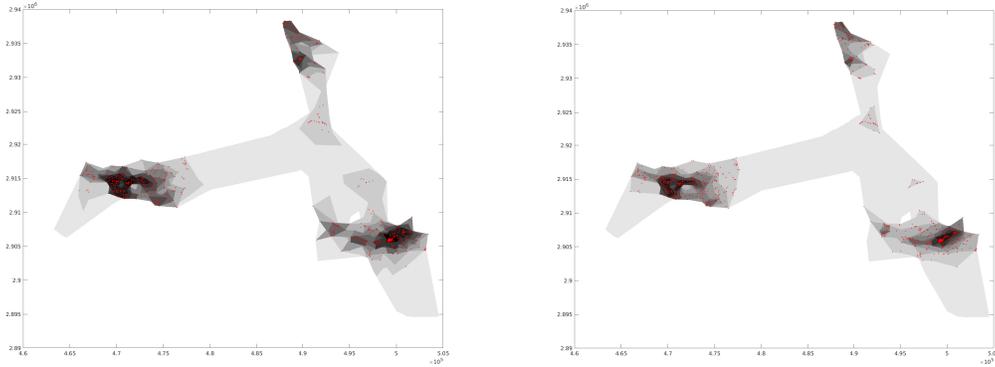


Figure 6: Isopleth for Panter Jitter, left original method from Getz and Wilmer (2004) and right the proposed corrected method

original method and on the right with our proposed correction. If the two plots look quite similar, they are a bit more "smooth" (less local maxima) with the proposed correction.

4.2.2 A manifold example

We finally illustrate in Figure 7 the manifold method. Here data are localizations on the earth via latitude and longitude (a 2 dimensional known manifold) of earthquakes (magnitude greater than 3). Computations have been done using the method proposed in section 3 with known manifold and real tangent space. The two graphs present a projection on a planisphere of the data sample with color depending on the rank of the density. The top one illustrates the local convex hull method, while the second one is just the constant kernel for manifold. The two densities look quite similar but there still exists some differences. For instance our correction allow to put more mass on the Mid-Atlantic Ridge, where the density is underestimated in the classical kernel method due to proximity

with a lower dimensional structure. Also in the circled area we can observe a maximum of the density localized in a corner of the support that should be impossible with no bias correction.

References

- Aamari, E. and Levrard, C. (2019) Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47, 177-204.
- Aamari, E. and Aaron, C. and Levrard, C. (2023) Minimax Boundary Estimation and Estimation with Boundary. *To appear in Bernoulli*
- Aaron C. and Bodart O. (2016) Local convex hull support and boundary estimation *Journal of Multivariate Analysis*, 147(C), 82-101.
- Baldin N. and Reiss M. (2016) Unbiased estimation of the volume of a convex body *Stochastic Processes and their Applications*, 126(12), 3716-3732.
- Bárány I., Fodor F. and Vígh V. (2016) Intrinsic volumes of inscribed random polytopes in smooth convex bodies *Advances in Applied Probability*, 42 , 3, 605 - 619.
- Bárány I., Larman D. (1988) Convex bodies, economic cap coverings, random polytopes *Matematika*, 35 , 2, 274 - 291.
- Beermann M. and Reitzner M. (2015) Beyond the Efron-Butcha identities: Distributional results for Poisson polytopes *Discrete and Computational Geometry*, 53 , 1, 226 - 244.
- Clément Berenfeld and Marc Hoffmann (2021) Density estimation on an unknown submanifold. *hal-02329703*
- T. Berry and T. Sauer. (2017) Density estimation on manifolds with boundary. *Computational Statistics and Data Analysis*, 107,1-17.
- Karine Bertin, Nicolas Klutchnikoff, José León, Clémentine Prieur (2020) Adaptive density estimation on bounded domains under mixing conditions *Electronic journal of Stat*, 14 , 1, 2198-2237.
- Jean-Daniel Boissonnat, Andre Lieutier, and Mathijs Wintraecken. (2019) The reach, metric dis-tortion, geodesic convexity and the variation of tangent spaces. *J. Appl. Comput. Topol.*, 3, 29-58.
- Taoufik Bouezmarnia, Jeroen V.K. Rombouts (2015) Nonparametric density estimation for multivariate bounded data. *Journal of Statistical Planning and Inference*, 140 , 139-152.

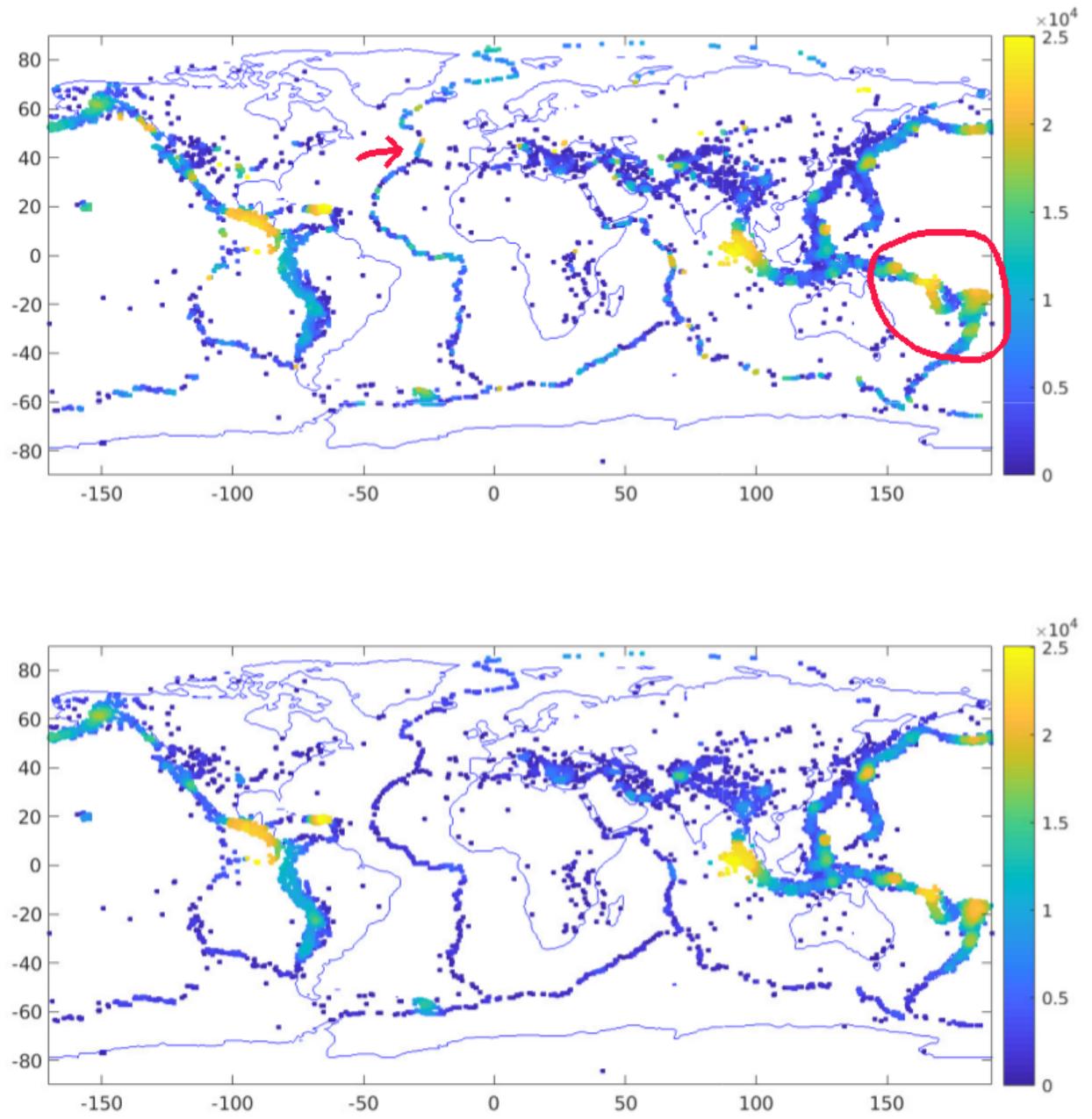


Figure 7: Seism density with two methods : the local convex hull for known manifold (top) and classical kernel for known manifold (down).

- Brunel V.E. (2013) A universal deviation inequality for random polytopes <https://hal.archives-ouvertes.fr/hal-00903687v2/document>,
- Brunel V.E. (2017) Uniform deviation and moment inequalities for random polytopes with general densities in arbitrary convex Body <https://arxiv.org/pdf/1704.01620.pdf>,
- Chacón, J.E., and Duong, T. (2018) Multivariate Kernel Smoothing and Its Applications. *Chapman and Hall*, CRC.
- Arthur Charpentier, Ewen Gallic (2016) Kernel density estimation based on Ripley's correction. *GeoInformatica*, 20 , 1, 95-116.
- Devroye, L. (1987) A Course in Density Estimation. *Birkhauser, Boston*
- Devroye, L. and Lugosi G. (1987) Combinatorial Methods in Density Estimation. *Springer Series in Statistics*
- Vincent Divol (2021) Reconstructing measures on manifolds: an optimal transport approach. *arXiv:2102.0759*
- Efron, B. (1965) The convex hull of a random set of points. *Biometrika*, 52:331–343.
- Federer, H. (1959) Curvature measures. *Transactions of the American Mathematical Society*, 93:418–491.
- Federer, H. (2014) Geometric measure theory. *Springer*.
- Benedikt Funkea Rafael Kawka (2015) Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods. *Computational Statistics and Data Analysis*, 92 , 148-162.
- Getz W.M. and Wilmers CC (2004) A local nearest-neighbor convex-hull construction of home ranges and utilization distributions *Ecography*, 27(4)489-505
- Giné E. and Guillou A. (2002) Parzen Rates of strong uniform consistency for multivariate kernel density estimators *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 38 , 6, 907-921.
- M. C. Jones, O. Linton and J. P. Nielsen (1995) A Simple Bias Reduction Method for Density Estimation *Biometrika*, 82 , 2, 327-338.
- R.J. Karunamunia and S.Zhangb (2008) Some improvements on a boundary corrected kernel density estimator *Statistics and Probability Letters*, 78 , 5, 499-507.
- Kim, Yoon Tae, Park, Hyun Suk, (2013) Geometric structures arising from Kernel density estimation on Riemannian manifolds. *J. Multivariate Anal.*, 114,112–126.

- Alexandre Leblanc (2010) A bias-reduced approach to density estimation using Bernstein polynomials *Parzen Advances in Applied Probability*, 22 , 4, 459-475.
- Peter Malec and Melanie Schienleb (2014) Nonparametric kernel density estimation near the boundary *Computational Statistics and Data Analysis*, 72,57-76
- Marron J. S. and Ruppert D. (1994) Transformations to Reduce Boundary Bias in Kernel Density Estimation *Journal of the Royal Statistical Society. Series B* , 56 , 4, 653 - 671. Parzen
- Parzen, E. (1962) On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*,33(3), 1065–1076.
- Pelletier, B. (2005) Kernel density estimation on Riemannian manifolds. *Statistics and Probability Letters*, 73, 3 , 297-304.
- Rodríguez-Casal A. and Saavedra-Nieves P. (2022) A data-adaptive method for estimating density level sets under shape conditions. *The Annals of Statistics*, 50, 3 , 1653-1668. Alberto Rodríguez-Casal, Paula Saavedra-Nieves
- Rosenblatt, M. (1956) Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27(3), 832–837.
- Ruppert D. and Cline D.B.H. (1994) Bias Reduction in Kernel Density Estimation by Smoothed Empirical Transformations *Ann. Statist.*, 22 , 1, 185-2010.
- Scott D.W. (1992) Multivariate Density Estimation: Theory, Practice, and Visualization. *Wiley Series in Probability and Statistics* Parzen
- Silverman, B.W (1986) Density estimation for statistics and data analysis. *Monograph on Statistics and Applied Probability*, 26, Chapman and Hall.
- Guenther Walther (1997) Granulometric smoothing *The Annals of Statistics*, 25(6),2273–2299

A Technical Lemmas in the full dimensional setting

Proposition 1. Let $p \in]0, 1[$ and $(n, q, D) \in (\mathbb{N}^*)^3$

$$\sum_{k=1}^n \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k} \leq \exp\left(-\frac{np}{10}\right) + \left(-\frac{np}{2}\right)^{-\frac{2q}{D+1}}.$$

Proof.

$$\sum_{k=1}^n \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k} \leq \sum_{k=1}^{\lfloor np/2 \rfloor} \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k} + \sum_{k=\lceil np/2 \rceil}^n \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k}.$$

Due to Bennet's Inequality that implies that $\mathbb{P}(K \leq np/2) \leq \exp(-np/10)$, we have that

$$\sum_{k=1}^n \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k} \leq \exp\left(-\frac{np}{10}\right) + \left(\frac{np}{2}\right)^{-\frac{2q}{D+1}}.$$

That concludes the proof of the proposition. \square

Proposition 2. *Let f be a \mathcal{C}^2 density supported by S such that, for all $x \in S$,*

$$g(x) = \|\nabla_f(x)\| \exp(3\|\nabla_{\ln(f)}(x)\|) \leq L_{\max}.$$

Additionally assume that, for all x , we have that $B(x, r) \cap S$ is path connected and that, for all $(x, y) \in S^2$ which are path connected, $d_S(x, y) \leq K_{\max}\|x - y\|$.

First we have that $f_{\max}^{(1)} := \sup_S \|\nabla_f(x)\|$ is well define.

Second, for all $r \leq \frac{1}{2K_{\max}}$, for all $x \in S$ such that $B(x, r) \cap S \subset \{z, f(z) > 0\}$, we have:

$$G_r(x) := \|\nabla_f(x)\| \frac{\max_{z \in B(x, r) \cap S} f(z)}{\min_{z \in B(x, r) \cap S} f(z)} \left(\frac{1}{\omega_d} \frac{\max_{z \in B(x, r) \cap S} f(z)}{\min_{z \in B(x, r) \cap S} f(z)} + 1 \right) \leq \max\left(\frac{2L_{\max}}{\omega_d}, f_{\max}^{(1)}\right).$$

Proof. Preliminary introduce $S_- = \{x \in S, f(x) < 1\}$ and $S_+ = \{x \in S, f(x) \geq 1\}$. Because S_+ is a compact, regularity condition on f implies that $\max_{S_+} \|\nabla_f(x)\|$ exists. Suppose now that $x \in S_-$. We have that :

$$\|\nabla_f(x)\| \exp(3\|\nabla_f(x)\|) \leq \|\nabla_f(x)\| \exp(3\|\nabla_f(x)/f(x)\|) \leq L_{\max}.$$

Because $\varphi : \mathbb{R}^+ \mapsto \mathbb{R}^+$ defined by $\varphi(x) = xe^{3x}$ is an increasing one to one function we have $\|\nabla_f(x)\| \leq \varphi^{-1}(L_{\max})$ and thus $\sup_{x \in S_-} \|\nabla_f(x)\|$ is well define. Therefore, $f_{\max}^{(1)} = \sup_{x \in S} \|\nabla_f(x)\|$ is well define.

Suppose that $r \leq (2K_{\max})^{-1}$ is fixed and consider x such that $B(x, r) \subset S$.

First, notice that:

$$\text{If } \frac{\max_{z \in B(x, r)} f(z)}{\min_{z \in B(x, r)} f(z)} \leq \omega_d \text{ then } G_r(x) \leq 2\|\nabla_f(x)\|. \quad (29)$$

Second, suppose that x is such that

$$\frac{\max_{z \in B(x,r)} f(z)}{\min_{z \in B(x,r)} f(z)} > \omega_d,$$

then $G_r(x) \leq \frac{2}{\omega_d} \|\nabla f(x)\| \left(\frac{\max_{z \in B(x,r)} f(z)}{\min_{z \in B(x,r)} f(z)} \right)^2$.

Denote by z_+ (resp z_-) the point in $B(x,r) \cap S$ where $\max_{B(x,r) \cap S} f(z)$ (resp. $\min_{B(x,r) \cap S} f(z)$) is realized. Because z_+ and z_- are path connected, by application of the mean value theorem,

$$\frac{\ln(f(z_+)) - \ln(f(z_-))}{d_S(z_+, z_-)} \leq \max_{B(x,r)} \|\nabla_{\ln(f)}(y)\|.$$

Because $r \leq \frac{1}{2K_{\max}}$ and the geodesic condition, we have

$$\frac{\max_{z \in B(x,r)} f(z)}{\min_{z \in B(x,r)} f(z)} \leq \exp \left(\max_{B(x,r)} \|\nabla_{\ln(f)}(y)\| \right). \quad (30)$$

Thus, $G_r(x) \leq \frac{2}{\omega_d} \|\nabla f(x)\| \exp(2 \max_{B(x,r)} \|\nabla_{\ln(f)}(y)\|)$, from which we have $G_r(x) \leq \frac{2}{\omega_d} f(x) \frac{\|\nabla f(x)\|}{f(x)} \exp(2 \max_{B(x,r)} \|\nabla_{\ln(f)}(y)\|)$ and finally:

$$G_r(x) \leq \frac{2}{\omega_d} f(x) \|\nabla_{\ln(f)}(x)\| \exp \left(2 \max_{B(x,r)} \|\nabla_{\ln(f)}(y)\| \right).$$

In the sequel z is a point of $B(x,r)$ where $\|\nabla_{\ln(f)}(z)\| = \max_{y \in B(x,r)} \|\nabla_{\ln(f)}(y)\|$. We clearly have $G_r(x) \leq \frac{2}{\omega_d} f(x) \|\nabla_{\ln(f)}(z)\| \exp(2 \|\nabla_{\ln(f)}(z)\|)$. Thus:

$$G_r(x) \leq \frac{2}{\omega_d} \frac{f(x)}{f(z)} \|\nabla f(z)\| \exp(2 \|\nabla_{\ln(f)}(z)\|),$$

and by new application of (30) we have $G_r(x) \leq \frac{2}{\omega_d} \|\nabla f(z)\| \exp(3 \|\nabla_{\ln(f)}(z)\|)$. Thus,

$$\text{if } \frac{\max_{z \in B(x,r)} f(z)}{\min_{z \in B(x,r)} f(z)} \geq \omega_d, \text{ then } G_r(x) \leq 2 \|\nabla f(z)\| \exp(3 \|\nabla_{\ln(f)}(z)\|) \leq \frac{2L_{\max}}{\omega_d}. \quad (31)$$

This and Equation (29) conclude the proof of the second part of the proposition. \square

Introduce, for x a point where $f(x) > 0$, $g_{x,r}(y) = r^D \frac{f(x+ry)}{\int_{B(x,r)} f(z) dz} \mathbb{I}_{B(0,1)}(y)$ which is the density of a local re-scaling of the sample. When exists, also introduce $M_{x,r} = \max_{y \in B(0,1)} g_{x,r}(y)$.

From the results in Theorem 1, Corollary 1 and Theorem 3 in Brunel (2017), we have:

Proposition 3.

$$\mathbb{P} \left[k \left(\frac{\int_{B(x,r) \setminus C_{x,r}} f(z) dz}{\int_{B(x,r)} f(z) dz} - C_1 (M_{x,r} + 1) k^{-2/(D+1)} \right) > t | N_{x,r} = k \right] \leq C_2 e^{-t}, \quad (32)$$

$$\mathbb{E} \left[\left(\frac{\int_{B(x,r) \setminus C_{x,r}} f(z) dz}{\int_{B(x,r)} f(z) dz} \right)^q | N_{x,r} = k \right] \leq A_q (M_{x,r} + 1)^q k^{-2q/(D+1)}, \quad (33)$$

$$\mathbb{E} \left[(N_{x,r}^\partial)^q | N_{x,r} = k \right] \leq B_q (M_{x,r} + 1)^q k^{q(D-1)/(D+1)}, \quad (34)$$

where C_1 and C_2 only depend on the dimension D , A_q and B_q depend on the dimension D and $q \in \mathbb{N}^*$.

B Proofs in the manifold setting

B.1 preliminary formulas on manifolds with positive reach

We recall here some properties of manifolds with positive reach that we will use in the following of the proof. First, by Lemma 3 in Boissonnat et al (2019), we have that

$$\|x - y\| \leq d_S(x, y) \leq 2\tau_S \arcsin \left(\frac{\|x - y\|}{2\tau_S} \right). \quad (35)$$

From which, we trivially obtain that, when $\|x - y\| \leq 2\tau_S$, we have

$$\|x - y\| \leq d_S(x, y) \leq \frac{\pi}{2} \|x - y\|. \quad (36)$$

Also, by prop 5.2 (arxiv version) in Aamari et Al (2023),

$$\angle T_x S, T_y S \leq \frac{d_S(x, y)}{\tau_S}. \quad (37)$$

And by prop 5.4 (arxiv version) in Aamari et Al (2023),

$$\|\pi_{\hat{T}_x}(y - x)\| \geq \|y - x\| \left(1 - \angle \hat{T}_x, T_x S - \frac{\|x - y\|}{2\tau_S} \right), \quad (38)$$

and

$$\|\pi_{\hat{T}_x^\perp}(y - x)\| \leq \|y - x\| \left(\angle \hat{T}_x, T_x S + \frac{\|x - y\|}{2\tau_S} \right). \quad (39)$$

In particular, when $\|\pi_{\hat{T}_x}(y - x)\| < \frac{\tau_S}{2(1 - \angle \hat{T}_x, T_x S)^2}$, by (38) it comes that either

$$\|y - x\| \leq \tau_S (1 - \angle \hat{T}_x, T_x S) \left(1 - \sqrt{1 - \frac{2\|\pi_{\hat{T}_x}(y - x)\|}{\tau_S(1 - \angle \hat{T}_x, T_x S)^2}} \right) \leq \frac{2\|\pi_{\hat{T}_x}(y - x)\|}{(1 - \angle \hat{T}_x, T_x S)}, \quad (40)$$

or

$$\|y - x\| \geq \tau_S(1 - \angle \hat{T}_x, T_x S) \left(1 + \sqrt{1 - \frac{2\|\pi_{\hat{T}_x}(y - x)\|}{\tau_S(1 - \angle \hat{T}_x, T_x S)^2}} \right) \geq \tau_S(1 - \angle \hat{T}_x, T_x S). \quad (41)$$

This last equation implies, again by (38), that we also have $\|\pi_{\hat{T}_x}(y - x)\| > \frac{\tau_S(1 - \angle \hat{T}_x, T_x S)^2}{2}$ and thus only (40) is true.

Proposition 4. *If for all $x \in S$, $\angle T_x S, \hat{T}_x \leq \theta < 1/12$ then for all $x \in S$, $\pi_{\hat{T}_x}$ is injective on $B(x, \tau_S/3)$*

Proof. For y and y' in $B(x, \tau_S/3)$, first $\angle T_y S, \hat{T}_x \leq \angle T_y S, T_x S + \theta$ thus by (37) and (36) $\angle T_y S, \hat{T}_x \leq \frac{d_S(x, y)}{\tau_S} + \theta \leq \frac{\pi\|x - y\|}{2\tau_S} + \theta \leq \frac{\pi}{6} + \theta$. From that Equation (38)

$$\|\pi_{\hat{T}_x}(y' - y)\| \geq \|y' - y\| \left(1 - \angle \hat{T}_x, T_y S - \frac{\|y' - y\|}{2\tau_S} \right)$$

With that, equation (38) gives :

$$\|\pi_{\hat{T}_x}(y' - y)\| \geq \|y' - y\| \left(\frac{4 - \pi}{6} - \theta \right) \geq \|y' - y\| \frac{7 - 2\pi}{12}$$

from which we obtain the injectivity. \square

B.2 Sketch of proof of Theorem 4 and 5

Take $R \leq \min\left(\frac{\tau_S}{3}, \frac{1}{4(D-d)}\right)$ and n large enough to have $\Theta_n \leq \min\left(\frac{1}{12}, \frac{1}{2(D-d)} - 2\frac{R}{\tau_S}\right)$.

Due to Proposition 4, we have that $\pi_{\hat{T}_x} : S \cap B(x, R) \rightarrow S_{x, R} = \pi_{\hat{T}_x}(S \cap B(x, R))$ is one to one. Let $(u_{x,1}, \dots, u_{x,d})$ be an orthonormalized basis of \hat{T}_x , completed by $(u_{x,d+1}, \dots, u_{x,D})$ to have an orthonormalized basis of \mathbb{R}^D , so that

$$\begin{aligned} \pi_{\hat{T}_x}^{-1} : \quad S_{x,R} &\rightarrow S \cap B(x, R) \\ z = \sum_{i=1}^d z_i u_{x,i} &\mapsto x + z + \sum_{i=d+1}^D \varphi_{x,i-d}(z) u_{x,i} \end{aligned}$$

Define $\phi_x(z) = (\varphi_{x,1}(z), \dots, \varphi_{x,D-d+1}(z))$ and denote by $J_{x,R}(z)$ the Jacobian matrix of $\phi_x(z)$ i.e.

$$J_x(z) = \begin{pmatrix} \frac{\partial \varphi_{x,1}}{\partial z_1}(z) & \dots & \frac{\partial \varphi_{x,1}}{\partial z_d}(z) \\ \vdots & & \vdots \\ \frac{\partial \varphi_{x,D-d}}{\partial z_1}(z) & \dots & \frac{\partial \varphi_{x,D-d}}{\partial z_d}(z) \end{pmatrix}$$

On the one hand, when $i \geq 1$, then $u_{x,d+i}$ is in \hat{T}_x^\perp . On the other hand for all $j \in \{1, \dots, d\}$ we have that $v_{x,j} := u_{x,j} + \sum_{k=1}^{D-d} \frac{\partial \varphi_{x,k}}{\partial z_j}(z) u_{x,k+d} \in T_{x+z+\phi_x(z)} S$. Thus

$|\langle u_{x,d+i}, v_{x,j} \rangle| \leq (\Theta_n + \angle T_x S, T_{x+z+\phi_x(z)} S) \|v_{x,j}\|$, from which it comes that, since $\|z + \phi_x(z)\| \leq R \leq \tau_S/3$, we have that $|\langle u_{x,d+i}, v_{x,j} \rangle| \leq \left(\Theta_n + \frac{\pi\|z+\phi_x(z)\|}{2\tau_S}\right) \|v_{x,j}\|$. Thus:

$$\left(\frac{\partial \varphi_{x,i}}{\partial z_j}(z)\right)^2 \leq \left(\Theta_n + \frac{\pi\|z + \phi_x(z)\|}{2\tau_S}\right)^2 \left(1 + \sum_{k=1}^{D-d} \left(\frac{\partial \varphi_{x,k}}{\partial z_j}(z)\right)^2\right).$$

Now, summing on i gives that, if $z \in S_{x,R}$,

$$\sum_{k=1}^{D-d} \left(\frac{\partial \varphi_{x,k}}{\partial z_j}(z)\right)^2 \leq \frac{\left(\Theta_n + \frac{\pi\|z+\phi_x(z)\|}{2\tau_S}\right)^2 (D-d)}{1 - \left(\Theta_n + \frac{\pi\|z+\phi_x(z)\|}{2\tau_S}\right)^2 (D-d)}.$$

That directly implies, by Cauchy Schwartz, that

$$\|J'_x(z)J_x(z)\|_\infty \leq \frac{\left(\Theta_n + \frac{\pi\|z+\phi_x(z)\|}{2\tau_S}\right)^2 (D-d)}{1 - \left(\Theta_n + \frac{\pi\|z+\phi_x(z)\|}{2\tau_S}\right)^2 (D-d)}.$$

Then, because $J'_x J_x$ is a d -dimensional symmetric positive matrix that has a rank upper bounded by $\min(d, D-d)$ and eigen values upper bounded by $d\|J'_x(z)J_x(z)\|_\infty$, it comes that

$$1 \leq \det(I + J'_x(z)J_x(z)) \leq \left(1 + \frac{\left(\Theta_n + \frac{\pi\|z+\phi_x(z)\|}{2\tau_S}\right)^2 (D-d)}{1 - \left(\Theta_n + \frac{\pi\|z+\phi_x(z)\|}{2\tau_S}\right)^2 (D-d)}\right)^{\min(d, D-d)}. \quad (42)$$

Introduce $g_{x,R}$ the density of $\pi_{\hat{T}_x}(X)$ given $\|X - x\| \leq R$:

$$g_{x,R}(z) = \sqrt{\det(I + J_{x,R}(z)'J_{x,R}(z))} \frac{f(x+z+\phi_x(z))}{\mathbb{P}_X(B(x,R))} \mathbb{1}_{S_{x,R}}(z).$$

Consider now $\mathbb{Y}_{n,R,x} = \pi_{\hat{T}_x}(\mathbb{X} \cap B(x,R)) = \{Y_1, \dots, Y_{N_{x,R}}\}$, where Y_i is an iid sample that has the same law than a sample drawn as follows : draw $N_{x,R}$ according to $\text{Binom}(n, \mathbb{P}(B(x,R)))$ then draw $N_{x,R}$ points in \mathbb{R}^d with a density $g_{x,R}$.

As previously $E'_{x,r,R}$ denotes the event " $|\widehat{C}_{x,r,R}| \geq A\omega_d r^d$ and $\widehat{N}_{x,r,R}^\partial \leq N_{x,R}/2$ ". Now introduce :

$$\begin{aligned} \widehat{\Gamma}_{x,r,R} &= \int_{\widehat{C}_{x,r,R}} g_{x,R}(z) dz \quad , \quad \widehat{\varepsilon}_1(x) = \frac{\widehat{N}_{x,r,R}^o - \tilde{\Gamma}_{x,r}(N_{x,R} - \widehat{N}_{x,r,R}^\partial)}{|\widehat{C}_{x,r,R}|(N_{x,R} - \widehat{N}_{x,r,R}^\partial)} \\ \widehat{\varepsilon}_2(x) &= \frac{\tilde{\Gamma}_{x,r}}{|\widehat{C}_{x,r,R}|} - g_{x,R}(0) \quad \text{and} \quad \widehat{\varepsilon}_3(x)(x) = \frac{N_{x,R}}{n} g_{x,R}(0) - f(x), \end{aligned}$$

so that, given $E'_{x,r,R}$, $\hat{f}_{r,R,A}(x) - f(x) = \frac{N_{x,R}}{n} (\widehat{\varepsilon}_1(x) + \widehat{\varepsilon}_2(x)) + \widehat{\varepsilon}_3(x)(x)$.

1. We have, for all $z \in S_{x,R}$, $f_{\min}^{(0)} \leq g_{x,R}(z) \leq f_{\max}^{(0)}(1+d)^{\min(d,D-d)}$. Also, by Corollary B.6 in Aamari et Al (2023), we have that there exists $x' \in B(x, 3r_n/4)$ with $B(x', r_n/4) \cap \partial S = \emptyset$. Thus, by Lemma B.11 in Aamari et Al (2023), $\pi_{\hat{T}_x}(B(x', r_n/4)) \subset S_{x,R}$ and, by reach condition, $B_{\hat{T}_x}(\pi_{\hat{T}_x}(x'), \frac{r_n}{4}(1 - \theta_n + \frac{r_n}{2\tau_S})) \subset S_{x,R}$. Thus, exactly as in proof of Lemma 1 and by use of the lower bound of f , we easily obtain that, for all $A < 4^{-d}$, there exists β_1 and β_2 such that

$$\mathbb{P}((E'_{x,r,R})^c | N_{x,R} = k) \leq \beta_1 \exp(-nr_n^d f_{\min}^{(0)} \beta_2) + \exp(-k/8).$$

2. Concerning $\hat{\varepsilon}_1(x)$, we easily obtain, exactly as in the proof of Lemma 2, that

$$\mathbb{E}(\hat{\varepsilon}_1(x) \varepsilon_2(x) | E_{x,r}) = 0. \quad (43)$$

We roughly have a uniform bound given by:

$$\mathbb{E}(\hat{\varepsilon}_1(x)^2 | E_{x,r} \text{ and } N_{x,R} = k) \leq \frac{2f_{\max}^{(0)}(1+d)^{\min(d,D-d)}}{A\omega_d r^d k}, \quad (44)$$

and

$$\mathbb{P}\left(|\hat{\varepsilon}_1(x)| > \sqrt{\frac{32 \ln k f_{\max}^{(0)}(1+d)^{\min(d,D-d)}}{A^2 \omega_d k r_n^d}} | E_{x,r} \text{ and } N_{x,R} = k\right) \leq 2k^{-4}(1+o(1)). \quad (45)$$

3. Concerning $\hat{\varepsilon}_2(x)$, when n large enough to have $r_n < \tau_S \frac{1}{2} (\frac{11}{12})^2$ so that (41) is impossible for all $z \in \pi_{\hat{T}_x}(S \cap B(x,R)) \cap B(0, r_n)$, we have $\|z + \phi_x(z)\| \leq \frac{2r_n}{1-\Theta_n}$. Thus, by Equations (28) and (42), it comes the existence of constants $C_{S,f}$ such that, for all $z \in \pi_{\hat{T}_x}(S \cap B(x,R)) \cap B(0, r_n)$,

$$|g_{x,R}(z) - g_{x,R}(0)| \leq C_{S,f}(r_n + \Theta_n^2) \quad (46)$$

$$|g_{x,R}(z) - g_{x,R}(0) - \frac{G_{x,z}}{\mathbb{P}(B(x,R))}| \leq C_{S,f}(r_n + \Theta_n)^2 \quad (47)$$

- (a) A rough bound $\hat{\varepsilon}_2(x) \leq C_{S,f}(r_n + \Theta_n^2)$ is obtained, as in Lemma 3, by:

$$|\hat{\varepsilon}_2(x)| \leq C_{S,f}(r_n + \Theta_n^2) + f_{\max}^{(0)}(1 + O(\Theta_n + r_n)^2) \frac{H(B_{\hat{T}_x}(0, r_n) \cap S_{x,R}) \setminus S_{x,R}}{A\omega_d r_n^d}.$$

To obtain a bound on $\frac{H(B(0,r_n) \cap S_{x,R}) \setminus S_{x,R}}{A\omega_d r_n^d}$, we use Lemmas A.6 and A.7 in Aamari et Al (2023). Namely:

i. By lemma A.6: if $d(x, \partial S) \geq 5r_n/4$, then $B_{\hat{T}_x}(0, r_n) \subset S_{x,R}$ and thus

$$\frac{H(B(0, r_n) \cap S_{x,R}) \setminus S_{x,R}}{A\omega_d r_n^d} = 0.$$

ii. If $d(x, \partial S) < 5r_n/4$, introduce $x^* = \pi_{\partial S}(x)$. We have $\angle T_{x^*} S, \hat{T}_x \leq \Theta_n + \frac{d_S(x, x^*)}{\tau_S}$ and then $d_S(x, x^*) \leq 2\tau_S \arcsin\left(\frac{\|x-x^*\|}{2\tau_S}\right) \leq \pi \frac{\|x-x^*\|}{2} \leq 3r_n$. Since $\Theta_n + 3r_n \leq 1/8$, we can apply Lemma A.7 and there exists a unit vector $\hat{\eta}^*$ such that

$$B_{\hat{T}_x}(\pi_{\hat{T}_x}(x^*) - r_0\hat{\eta}^*, r_0) \subset S_{x,R} \subset \mathring{B}_{\hat{T}_x}^c(\pi_{\hat{T}_x}(x^*) + r_0\hat{\eta}^*, r_0).$$

This is a local version of the rolling ball property used in the proof of Lemma 3, which is sufficient to conclude that, for n large enough,

$$|\hat{\varepsilon}_2(x)| \leq 2 \frac{C_{S,f}}{A} (r_n + \Theta_n^2). \quad (48)$$

(b) Suppose now that x is far from the boundary, i.e. $d(x, \partial S) \geq 5r_n/4$. That guaranties, due to Lemma 5.6. in Aamari et Al (2023) (arxiv version), that $B(O, r_n) \subset S_{x,R}$. Better bounds on $\hat{\varepsilon}_2(x)$, are obtained as in Equation 21. Precisely, it comes that

$$\hat{\varepsilon}_2(x) = \frac{\int_{B(x, r_n)} (g_{x,R}(z) - g_{x,R}(0))}{|\widehat{C}_{x,r,R}|} - \frac{\int_{B(x, r_n) \setminus \widehat{C}_{x,r,R}} (g_{x,R}(z) - g_{x,R}(0))}{|\widehat{C}_{x,r,R}|}.$$

By use of Equation (47), we have that, conditionally to $E'_{x,r,R}$,

$$\left| \frac{\int_{B(x, r_n)} (g_{x,R}(z) - g_{x,R}(0))}{|\widehat{C}_{x,r,R}|} \right| \leq \frac{C_{S,f}}{A} (r_n + \Theta_n)^2.$$

Also, when n large enough,

$$\left| \frac{\int_{B(x, r_n) \setminus \widehat{C}_{x,r,R}} (g_{x,R}(z) - g_{x,R}(0))}{|\widehat{C}_{x,r,R}|} \right| \leq \frac{2C_{S,f}}{A} (r_n + \Theta_n^2) \frac{|B(x, r_n) \setminus \widehat{C}_{x,r,R}|}{|B(x, r_n)|}.$$

Then, because the density $g_{x,R}$ is lower and upper bounded by positive constants, we can apply results in Brunel (2017) and formulaes (33) to obtain the existence of a constant C such that

$$\mathbb{E}(\hat{\varepsilon}_2(x)^2 | E_{x,r} \text{ and } N_{x,R} = k) \leq C \left((r_n + \Theta_n)^2 + (r_n + \Theta_n^2) (kr_n^d)^{-\frac{2}{d+1}} \right)^2. \quad (49)$$

Also, by use of (32), there exists C and C' such that:

$$\mathbb{P} \left(|\widehat{\varepsilon}_2(x)| \geq \frac{Cr_n \ln k}{kr_n^d} + r_n \left(\frac{C'}{kr_n^d} \right)^{\frac{2}{d+1}} + r_n^2 f_{\max}^{(2)} |E_{x,r} \text{ and } N_{x,R} = k \right) \leq C_2 k^{-4} + \exp \left(-kr_n^D f_{\min}^{(0)} / 10 \right). \quad (50)$$

4. By definition of $\widehat{\varepsilon}_3(x)$, we have

$$\mathbb{V}(\widehat{\varepsilon}_3(x)) = \frac{(1 - \mathbb{P}(B(x, R))) \det(I + J_{x,R}(0)' J_{x,R}(0)) f(x)^2}{n \mathbb{P}(B(x, R))},$$

from which $\mathbb{V}(\widehat{\varepsilon}_3(x)) \leq \frac{f(x)^2}{n \omega_d R^d f_{\min}^{(0)}} (1 + O(\Theta_n^2))$.

Now, by use of the decomposition

$$\widehat{\varepsilon}_3(x) = f(x) \left(\frac{N_{x,R}}{n \mathbb{P}(B(x, R))} - 1 \right) + f(x) \frac{N_{x,R}}{n \mathbb{P}(B(x, R))} \left(1 - \sqrt{\det(I + J_{x,R}(0)' J_{x,R}(0))} \right),$$

it comes that $|\mathbb{E}(\widehat{\varepsilon}_3(x))| \leq f(x) C \Theta_n^2 (1 + o(1))$ and so $\mathbb{E}(\widehat{\varepsilon}_3(x)^2) \leq O(\max(n^{-1}, \Theta_n^4))$.

Also, we clearly have

$$\left| f(x) \frac{N_{x,R}}{n \mathbb{P}(B(x, R))} \left(1 - \sqrt{\det(I + J_{x,R}(0)' J_{x,R}(0))} \right) \right| \leq \frac{f_{\max}^{(0)}}{f_{\min}^{(0)} \omega_d R^d} O_{d,D}(\Theta_n)^2.$$

With this equation and by use of Hoeffding inequality, it comes that

$$\mathbb{P} \left(|\widehat{\varepsilon}_3(x)| \geq \sqrt{\frac{2 \ln n}{n \omega_d R^d f_{\max}^{(0)}}} + O(\Theta_n^2) \right) \leq 2n^{-4}.$$

Theorems 4 and 5 are consequences of previous formulas, getting rid of the condition on $N_{x,R}$ by use of Hoeffding inequality which gives the uniform bound:

$$\mathbb{P}(N_{x,R} \leq \frac{n}{2} f_{\min}^{(0)} \omega_d R^d) \leq \exp \left(-\frac{n}{2} f_{\min}^{(0)} \omega_d R^d \right).$$