

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Shape and Geometry in Statistics: Support and Density Estimation

Permalink

<https://escholarship.org/uc/item/9v21t0zw>

Author

Cui, Lingfei

Publication Date

2021

Peer reviewed|Thesis/dissertation

Shape and Geometry in Statistics: Support and Density Estimation

By

LINGFEI CUI
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Wolfgang Polonik, Chair

Alexander Aue

Miles Lopes

Committee in Charge

2021

To my parents.

Contents

Abstract	iv
Acknowledgments	v
Chapter 1. Introduction	1
1.1. Density Support Estimation	1
1.2. Density Estimation	2
Chapter 2. Theoretical Analysis of the Local Convex Hull (LoCoH) Estimators of the Density Support	3
2.1. Introduction	3
2.2. Geometric Properties	7
2.3. k -Nearest Neighbor Distance	12
2.4. Convergence of the k NN LoCoH Estimator $\hat{H}_{k_n}$	19
2.5. Concentration Bounds for the r -ball LoCoH Estimator $\hat{H}_{r_n}$	21
2.6. Concentration Bounds for the k NN LoCoH Estimator $\hat{H}_{k_n}$	50
2.7. Appendix: Cited Theorems	58
Chapter 3. Generalized Switching Relation and Its Applications	59
3.1. Introduction	59
3.2. Generalized Switching Relation	60
3.3. Asymptotic Results for Nonparametric Maximum-Likelihood Estimators (NPMLEs) Under Order Restrictions	71
3.4. Pointwise Convergence Rate of NPMLE Under Block-decreasing Order Restriction	84
3.5. Appendix: Cited Theorems	86
Bibliography	89

Abstract

In this dissertation, we are tackling two estimation problems using methods with geometric arguments heavily involved. The first part of this dissertation is focused on theoretical analysis of a type of local convex hull (LoCoH) estimator of the density support, which includes the r -ball LoCoH estimator and the k NN LoCoH estimator. We present two finite-sample concentration bounds for both estimators and discuss how to use these bounds to study the large sample behaviors and the performance of these estimators in high dimensions. The second part of this dissertation is focused on establishing a generalized switching relation, which states that the behavior of the nonparametric maximum likelihood estimators (NPMLEs) of density functions under order constraints can be controlled by the behavior of the empirical measure. Some examples are shown to illustrate how to use this relation to derive asymptotic results.

Acknowledgments

First and foremost, I am extremely grateful to my advisor, Professor Wolfgang Polonik, for his continuous support and guidance throughout the years. He is always there ready to help and give advice whenever I need it on both research and life. Thank you, Wolfgang! Without your help, this dissertation would not be possible. I am also very thankful to my qualifying exam and/or dissertation committee: Professor Alexander Aue, Professor Miles Lopes, Professor James Sharpnack, Professor Prabir Burman, Professor Ethan Anderes and Professor Aditya Guntuboyina for all their time and help. In addition, I would like to thank my friends, classmates and colleagues in the Department of Statistics for the support and understanding throughout my years here. Last but not the least, I would like to thank my parents for their unconditional love and endless support. Words cannot express my gratitude enough. This is not the end of our journey together. It's just the beginning.

CHAPTER 1

Introduction

Almost all the problems in statistics relate to probability distribution estimation. For example, regression problems need to estimate the distribution of errors; testing problems are essentially distribution comparison problems; classification problems relate to probability support or mode estimations; etc. In this dissertation, we study two specific problems: one is to theoretically analyze a type of local convex hull (LoCoH) estimator of the density support; the other is to introduce a tool to study the nonparametric maximum likelihood estimators (NPMLEs) of the density functions under order constraints. For both problems, the approaches we take are heavily influenced and guided by the geometric ideas.

In this chapter, we provide a high-level introduction of both problems.

1.1. Density Support Estimation

Density support estimation is to estimate the shape of the manifold where the data points live in. Naturally, one asks the questions: what should the shape be like? Can we estimate any shapes equally well? The answer is of course not. The simplest geometric assumption on the support is probably convexity. If the convexity assumption is satisfied, the convex hull estimator has good theoretical guarantees. But in the majority of the cases, the density support is not convex. In Section 2.2, we discuss three geometric assumptions that generalize the notation of convexity. More specifically, we focus on the rolling-ball assumption. And the type of the estimator we consider is called the local convex hull (LoCoH) estimator. As the name suggests, this type of estimator is building “convex hull estimators” locally. Two different estimators of such type will be considered: the r -ball LoCoH estimator and the k -nearest neighbor (NN) LoCoH estimator. The former estimator uses balls of radius r centered at each data point as local neighborhoods. The latter uses sets of k -nearest neighbors of each data point as local neighborhoods.

We first study the connections between the two different LoCoH estimators through the k NN distance. In Section 2.3 we show a uniform bound for the k NN distance. Then in Section 2.4, we illustrate how to use this bound to get a consistency result for the k NN LoCoH estimator from a consistency result for the r -ball LoCoH estimator. In Section 2.5 we present the main results for the r -ball LoCoH estimator, namely, a finite-sample concentration bound, which is followed by the discussions of how to use this bound to study the large sample behavior and performance of this estimator in high dimensions. In Section 2.6, we present similar results for the k NN LoCoH estimator.

1.2. Density Estimation

Density estimation is to reconstruct the underlying probability density function from sampled data. We are focusing on a special class of functions, the monotone functions. In the one-dimension case, monotonicity is with respect to the usual order on the real line, and it has been well-studied in the literature. In high dimension, we hope to define the monotonicity with respect to any suitable order. So first we talk about how to generalize the notation of monotonicity in high dimensions and then introduce a tool called generalized switching relation to study the NPMLEs of these multivariate monotone functions.

In Sections 3.2.1 and 3.2.2, we review the classical switching relation that is used to study the one-dimension monotone functions and propose a key observation. Then in Sections 3.2.3 and 3.2.4 we talk about how to impose high dimension monotonicity through shape constraints and introduce the NPMLEs to estimate such multivariate monotone functions. Finally, in Section 3.2.5 we formulate the generalized switching relation for the NPMLEs with order constraints. In the following section, we use examples to illustrate how to apply the generalized switching relation to study the asymptotic behavior of the NPMLEs with order constraints.

In Section 3.4, we present a pointwise convergence rate for the NPMLE under block-decreasing order restriction for uniform distributions without using the generalized switching relation.

Theoretical Analysis of the Local Convex Hull (LoCoH)

Estimators of the Density Support

2.1. Introduction

2.1.1. Background. Probability density support estimators can be classified into two categories. First is the plug-in type, which is constructed from a probability density function estimator. The other is the direct estimation type, which constructs the estimator directly from data points. Applications of density support estimators include abnormality detection (Devroye and Wise (1980)), such as monitoring machines and medical diagnosis; marketing and yield estimation; animal home range estimation (Getz and Wilmers (2004), Getz *et al.* (2007)), etc.

Let $\mathcal{X}_n = \{X_1(t_1), \dots, X_n(t_n)\}$ be a set of n points sampled from an underlying distribution F with density function f that is supported on a subset S in $\mathbb{R}^d$. For all $i \in \{1, \dots, n\}$, t_i is the time stamp when X_i is sampled. Here t_i only appears for some specific sampling strategies. For example, in the case of studying the support of the activity distribution of certain animals, a data point X_1 is recorded when an animal is captured in sight at time t_1 . The same animal might be captured and recorded again as X_2 at another time point t_2 . X_1 and X_2 might be correlated and such correlation can be measured by $|t_2 - t_1|$. Or a real-time tracker can be placed on animals. The recorded data will be trajectories indexed by time. In other cases such as an *i.i.d.* random sample, $X_i(t_i)$ can be written as X_i .

Suppose $\hat{S}$ is a density support estimator from this sample. Two distances can be used to measure the quality of this estimator.

- (symmetric difference distance) $d_\lambda(\hat{S}, S) = \lambda(\hat{S} \Delta S)$, where Δ denotes the symmetric difference, λ can be any measure on $\mathbb{R}^d$.
- (Hausdorff distance) $d_H(\hat{S}, S) = \inf\{\epsilon > 0 : \hat{S} \subset S \oplus \epsilon\mathcal{B}, S \subset \hat{S} \oplus \epsilon\mathcal{B}\}$, where $A \oplus \epsilon\mathcal{B} = \bigcup_{x \in A} B(x, \epsilon)$ for any set A .

The first metric is usually used to quantify the quality of set estimator while the second one can also be used for the boundary estimation.

The following is a brief survey of the two types of density support estimators.

I. PLUG-IN TYPE

Given a density estimator $\hat{f}_n$, the plug-in type supported estimator can be formulated as

$$(2.1.1) \quad \hat{S}_n = \{x \in \mathbb{R}^d, \hat{f}_n(x) \geq \alpha_n\}, \text{ where } \alpha_n \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Depending on whether to take time into account or not, there are two types of probability density estimation methods: the fixed-time methods and the time-varying methods. The majority of density estimators fall in the first category. For example, the classical kernel density estimator $\hat{f}_{KD,n,h}$ assumes that the data collected should be *i.i.d.* (see, e.g., Scott (2015) or Chacón and Duong (2018)).

$$(2.1.2) \quad \hat{f}_{KD,n,h}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

where $h > 0$ is a fixed bandwidth and $K(\cdot)$ is a kernel function.

In the last decade, several other methods are proposed to deal with trajectory-type data such as Brownian bridge movement model (Horne *et al.* (2007)), Dynamic Brownian bridge movement model (Kranstauber *et al.* (2012)) and time-geographic density estimator (TGDE) (Downs (2010)). The first two methods use a Brownian bridge to model the trajectories and then construct the density estimator in a maximum likelihood framework. They follow a parametric approach. The TGDE $\hat{f}_{TGDE,n,v}$, however, can be seen as a time-geographic version of the kernel density estimator.

$$(2.1.3) \quad \hat{f}_{TGDE,n,v}(x) = \frac{1}{(n-1)[(t_n - t_1)v]^2} \sum_{i=1}^{n-1} G\left(\frac{\|x - X_i\| + \|X_{i+1} - x\|}{(t_{i+1} - t_i)v}\right)$$

where v is the maximum velocity of the object and G is a function of the geoellipse. The idea is to fit an elliptical density using two temporally consecutive data points for all the data points.

The quality of the plug-in type density support estimator depends on the quality of the density estimator, their connection was studied by Cuevas *et al.* (1997).

II. DIRECT ESTIMATION TYPE.

The most simple and naive direct estimation method is probably the Devroye and Wise (1980) estimator $\hat{S}_{DW,r}$, which is the union of balls of radius r centered at the observed data points.

$$(2.1.4) \quad \hat{S}_{DW,r} = \cup_{i=1}^n B(X_i, r),$$

Under the regularity conditions on the probability distribution, its consistency and convergence rate results in terms of the symmetric difference distance have been established by Devroye and Wise (1980) and Biau *et al.* (2008).

Another type of direct estimation method is the so-called minimum volume set method (Polonik (1995b), Polonik (1997)). Let $\mathcal{S}$ be a class of sets. The minimum volume estimator $\hat{S}_{MV}$ with respect to $\mathcal{S}$ is defined as

$$(2.1.5) \quad \hat{S}_{MV} = \underset{S \in \mathcal{S}}{\operatorname{argmin}} \{\operatorname{Leb}(S) : \mathcal{X}_n \subset S\}.$$

When $\mathcal{S}$ is the class of convex sets, $\hat{S}_{MV}$ is the convex hull estimator, which has been extensively studied in the literature (see Brunel (2018)). When $\mathcal{S}$ is the class of all d-dimensional ellipsoids, $\hat{S}_{MV}$ is the minimum volume ellipsoid, which has been introduced by Rousseeuw (1985). $\hat{S}_{MV}$ can be biased when $S \notin \mathcal{S}$, namely, when $\mathcal{S}$ is misspecified, meaning that the true S is non-convex. In order to accommodate such cases, several generalizations to the notion of convexity can be adopted as constraint to $\mathcal{S}$. This includes the positive reach (Federer (1969) or Thäle (2008)), r -convexity (Rodríguez Casal (2007)) and rolling ball conditions (Cuevas *et al.* (2012)) (we will define them later in the text). For instance, when $\mathcal{S}$ is restricted to be the class of r -convex sets, the corresponding minimum volume estimator is called r -convex hull estimator. However, this minimum volume set estimator might not be well-defined when elements of $\mathcal{S}$ are assumed to have reach $> r > 0$ (see Colesanti and Manselli (2010)).

Note that the minimum volume set estimator corresponding to the class $\mathcal{S}$ coincides with the plug-in estimator when $\hat{f}_n$ in (2.1.1) is chosen to be the silhouette density estimator (see Polonik (1998)) with respect to S whose upper level sets are in $\mathcal{S}$ and $\alpha_n = 0$.

Another way to accommodate the non-convex situation using convex hull type method is by localization. The k -nearest neighbor local convex hull (k NN LoCoH) estimator $\hat{H}_k$ and the r -ball local convex hull (r -ball LoCoH) estimator $\hat{H}_r$ were proposed by Getz and Wilmers (2004) and Getz *et al.* (2007), respectively for home range estimation. Given a point set A , the convex hull of A is defined as

$$(2.1.6) \quad \mathcal{H}(A) = \left\{ \sum_{x \in A} \alpha_x x, \alpha_x \geq 0 \forall x \in A, \text{ and } \sum_{x \in A} \alpha_x = 1 \right\}$$

The two local estimators are defined as follows.

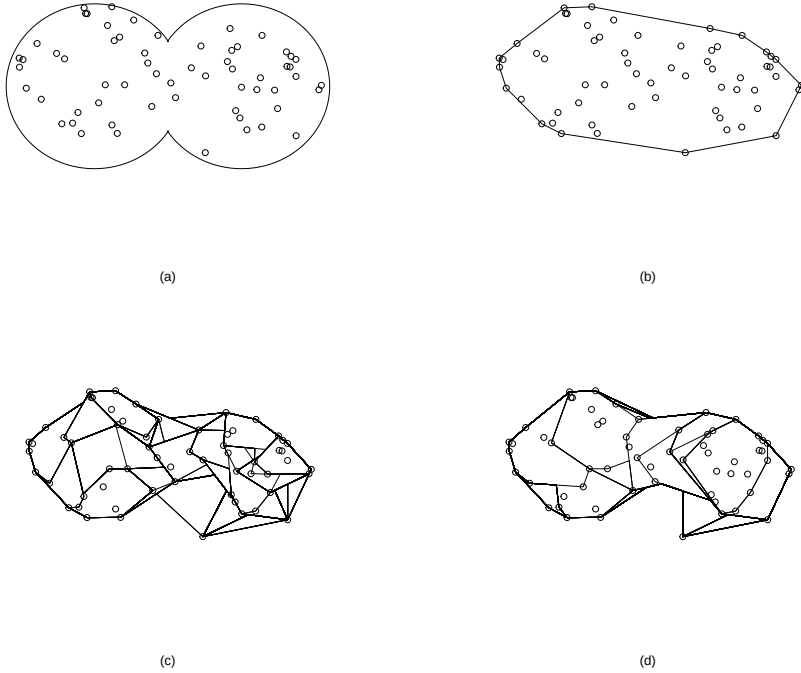


FIGURE 2.1. Convex hull type support estimates for a simulated dataset. (a) true support S ; (b) convex hull estimate; (c) k NN LoCoH estimate with $k = 10$; (d) r -ball LoCoH estimate with $r = 2$.

2.1.2. Definitions of Local Convex Hull Estimators. Let $X_{(k)}(x)$ be the k -th nearest neighbor of x in $\mathcal{X}_n$.

DEFINITION 2.1.1 (k NN LoCoH estimator). *The k -nearest neighbor local convex hull estimator of the support of f is*

$$\hat{H}_k = \bigcup_{X \in \mathcal{X}_n} \mathcal{H}(N_k(X)) = \bigcup_{X \in \mathcal{X}_n} \mathcal{H}(X_{(1)}(X), \dots, X_{(k)}(X))$$

where $N_k(X)$ is the k -nearest neighbors of X in $\mathcal{X}_n$, excluding X itself.

DEFINITION 2.1.2 (r -ball LoCoH estimator). *The r -ball local convex hull estimator of the support of f is*

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

The consistency of r -ball LoCoH estimator $\hat{H}_r$ has been studied in Aaron and Bodart (2016). In this chapter, we establish a global and a local concentration bound for $\hat{H}_r$. The consistency as well as the concentration bounds are also obtained for the k NN LoCoH estimator $\hat{H}_k$.

$\hat{H}_r$, $\hat{H}_{r_n}$ always refer to the r -ball LoCoH estimator and $\hat{H}_k$, $\hat{H}_{k_n}$ always refer to the k NN LoCoH estimator throughout the text.

2.2. Geometric Properties

2.2.1. Three Generalized Convexity Assumptions. Here we define the three geometric properties mentioned above that can be seen as generalizations to the notion of convexity and discuss their relations.

DEFINITION 2.2.1 (positive reach). *Let $\text{reach}(S, x) = \sup\{r > 0 : B^\circ(x, r) \subset \text{UP}(S)\}$ for $x \in S$, where $B^\circ(x, r)$ is the open ball centered at x with radius r , $\text{UP}(S)$ is the set of points that have a unique projection on S . Then the reach of S is*

$$\text{reach}(S) = \inf\{\text{reach}(S, x) : x \in S\}.$$

S has positive reach if $\text{reach}(S) > 0$.

DEFINITION 2.2.2 (*r*-convexity). Let $r > 0$. A set $S \subset \mathbb{R}^d$ is called *r*-convex if $S = C_r(S)$

$$C_r(S) = \bigcap_{B^\circ(y,r) \cap S = \emptyset} (B^\circ(y,r))^c.$$

r-convexity generalizes the notion of convexity in a more obvious way from its definition. Note that for a set A , $C_r(A) \subset C_s(A)$ for $0 < r \leq s$. As $r \rightarrow 0$, $C_r(A) \rightarrow \overline{A}$, where $\overline{A}$ is the closure of A ; and as $r \rightarrow \infty$, $C_r(A) \rightarrow \mathcal{H}(A)$.

DEFINITION 2.2.3 (rolling-ball condition). Let S be a d -manifold in $\mathbb{R}^d$ with non-empty boundary. If there exists $R_{out} > 0$ such that $\forall x \in \partial S$, there exists $O_x^- \in \mathbb{R}^d$ with $x \in B(O_x^-, R_{out}) \subset \overline{S}^c$, then the balls of radius $R_{out} > 0$ roll freely outside S .

If there exists $R_{in} > 0$ such that $\forall x \in \partial S$, there exists $O_x^+ \in \mathbb{R}^d$ with $x \in B(O_x^+, R_{in}) \subset \overline{S}$, then the balls of radius $R_{in} > 0$ roll freely inside S .

The rolling-ball condition is a geometric smoothness condition on the boundary. It is a special case of the more general rolling-type condition in which a convex body rolls freely inside or outside S .

The positive reach condition is the most popular among these three properties. For a compact set S with positive reach r , the Lebesgue measure of the set $B(S, \epsilon)$ can be expressed as a polynomial in ϵ with degree d for $0 < \epsilon < r$ (see Cuevas *et al.* (2012)). Note that in the case when S is a convex set, whose reach is $+\infty$, the above property holds $\forall \epsilon > 0$, which coincides with the classical Steiner formula for convex set. From this perspective, convexity can be seen as a special case of the positive reach condition.

These properties are closely related. Let $r > 0$.

- (A) S has positive reach r .
- (B) S is *r*-convex.
- (C) balls of radius r rolls freely outside S .
- (D) balls of radius r rolls freely inside S .

It can be shown that if S is compact, (A) $\implies$ (B) and (B) $\implies$ (C). But the converses are not true in general. Details can be found in Cuevas *et al.* (2012). If S is both compact and

path-connected, then $(C) \implies (B)$ (see Walther (1999)). Also $(C)&(D)$ together imply ∂S has positive reach r . And when $S = \bar{S}^o$, the converse is also true (Arias-Castro *et al.* (2017)).

2.2.2. The Rolling-ball Condition.

PROPOSITION 2.2.1 (smoothness condition). *If $S \in \mathbb{R}^d$ is a compact d -manifold that satisfies the rolling-ball condition with $R_{in} = R_{out} = R$. Then for all $x \in \partial S$, $\exists$ a unique inward pointing $u_x \in \mathbb{R}^d$ such that $\|u_x\| = 1$ and $\forall x, y \in \partial S$, the following Lipschitz condition holds $\|u_x - u_y\| \leq \frac{1}{R} \|x - y\|$.*

See Walther (1999) Theorem 1 for a proof.

LEMMA 2.2.1. *If balls of radius R roll freely inside $S \in \mathbb{R}^d$, which is a compact d -manifold, then there exists some $\gamma \geq \frac{1}{2} I_{\frac{3}{4}} \left(\frac{d+1}{2}, \frac{1}{2} \right)$ such that $\text{Leb}(S \cap B(x, r)) \geq \gamma \text{Leb}(B(x, r))$ for $0 < r \leq R$, where $I_t(\alpha, \beta)$ is the regularized incomplete beta function.*

PROOF. Let $g(r) = \frac{\text{Leb}(S \cap B(x, r))}{\text{Leb}(B(x, r))}$. The goal is to show that $g(r) \geq \gamma$ for some $\gamma \geq \frac{1}{2} I_{\frac{3}{4}} \left(\frac{d+1}{2}, \frac{1}{2} \right)$ and $0 < r < R$. The worst case scenario is that x lies on the boundary of S which is the same as the boundary of a ball with radius R . Consider two spheres of radii R and r and centered at $(0, \dots, 0)$ and $(R, 0, \dots, 0)$ respectively in $\mathbb{R}^d$. The equations of the two spheres are

$$(2.2.1) \quad \begin{aligned} x_1^2 + \dots + x_d^2 &= R^2 \\ (x_1 - R)^2 + x_2^2 + \dots + x_d^2 &= r^2. \end{aligned}$$

Solving this system of equations yields $x'_1 = \frac{2R^2 - r^2}{2R}$. Let ϕ_1 be the angle between a vector starting at $(0, \dots, 0)$ and ending at a point lying in the intersection of the two spheres and the positive x -axis. Let ϕ_2 be the angle between a vector starting at $(R, 0, \dots, 0)$ and ending at a point lying in the intersection of the two spheres and the negative x -axis. Then

$$(2.2.2) \quad \begin{aligned} \cos \phi_1 &= \frac{x_1}{R} = 1 - \frac{r^2}{2R^2}, \\ \cos \phi_2 &= \frac{R - x'_1}{r} = \frac{r}{2R}. \end{aligned}$$

So

$$\begin{aligned}
(2.2.3) \quad \sin \phi_1 &= \frac{r\sqrt{4R^2 - r^2}}{2R^2}, \\
\sin \phi_2 &= \frac{R - x'_1}{r} = \frac{\sqrt{4R^2 - r^2}}{2R}.
\end{aligned}$$

Now the volume $V(r)$ of the intersection of these two balls is the sum of the volume of two spherical caps cut by the hyperplane $x_1 = x'_1$.

$$\begin{aligned}
(2.2.4) \quad V(r) &= \frac{1}{2}V_d R^d I\left(\frac{r\sqrt{4R^2 - r^2}}{2R^2}\right)^2\left(\frac{d+1}{2}, \frac{1}{2}\right) + \frac{1}{2}V_d r^d I\left(\frac{\sqrt{4R^2 - r^2}}{2R}\right)^2\left(\frac{d+1}{2}, \frac{1}{2}\right) \\
&> \frac{1}{2}V_d r^d I\left(\frac{\sqrt{4R^2 - r^2}}{2R}\right)^2\left(\frac{d+1}{2}, \frac{1}{2}\right),
\end{aligned}$$

where V_d is the volume of the unit ball in $\mathbb{R}^d$. Therefore

$$\begin{aligned}
(2.2.5) \quad g(r) &\geq \frac{V(r)}{V_d \cdot r^d} > \frac{\frac{1}{2}V_d r^d I\left(\frac{\sqrt{4R^2 - r^2}}{2R}\right)^2\left(\frac{d+1}{2}, \frac{1}{2}\right)}{V_d \cdot r^d} \\
&= \frac{1}{2}I\left(\frac{\sqrt{4R^2 - r^2}}{2R}\right)^2\left(\frac{d+1}{2}, \frac{1}{2}\right).
\end{aligned}$$

But $0 < r \leq R$, so $\frac{3}{4} \leq \left(\frac{\sqrt{4R^2 - r^2}}{2R}\right)^2 < 1$.

$$(2.2.6) \quad g(r) > \frac{1}{2}I_{\frac{3}{4}}\left(\frac{d+1}{2}, \frac{1}{2}\right).$$

□

2.2.3. Covering Numbers. The complexity of a set in $\mathbb{R}^d$ can be measured by the Minkowski dimension. The metric we are considering is the standard Euclidean distance throughout the text. Let $\text{Vol}_d(A)$ be the d -dimensional volume of a set A and V_d be the volume of the d -dimensional unit ball.

DEFINITION 2.2.4 (covering number). *Given a set S in $\mathbb{R}^d$, its covering number $N(S, \epsilon)$ is the minimum number of (closed) ℓ_2 -balls of radius ϵ needed to cover S .*

DEFINITION 2.2.5 (Minkowski dimension). *Given a set $S \subset \mathbb{R}^d$ with covering number $N(S, \epsilon)$, the Minkowski dimension of S if exists, is*

$$(2.2.7) \quad \text{DIM}_{\text{Mink}}(S) = \lim_{\epsilon \rightarrow 0} -\frac{\log N(S, \epsilon)}{\log \epsilon}.$$

PROPOSITION 2.2.2. *A k -manifold S of class $\mathcal{C}^1$ in $\mathbb{R}^d$ has Minkowski dimension k . There exists $\lambda > 0$ such that the covering number $N(S, \epsilon) \leq \lambda_S \epsilon^{-k}$.*

A compact d -manifold in $\mathbb{R}^d$ is of class $\mathcal{C}^\infty$ (see Kühnel (2015)). This property will be used to cover S and ∂S with small balls later in the proof. The following proposition gives some explicit bounds for the covering numbers of S and ∂S when S satisfies the rolling ball condition.

PROPOSITION 2.2.3. *Let S be a compact d -manifold in $\mathbb{R}^d$. ∂S is the boundary of S . If S satisfies the rolling ball condition with radius R , then the covering number $N(S, \epsilon)$ of S and the covering number $N(\partial S, \epsilon)$ of ∂S satisfy*

$$(2.2.8) \quad N(S, \epsilon) \leq \left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon} + \frac{1}{R} \right)^d,$$

and

$$(2.2.9) \quad N(\partial S, \epsilon) \leq 2 \left[\left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon} + \frac{1}{R} \right)^d - \frac{2^d \text{Vol}_d(S) \Gamma(d/2 + 1)}{\pi^{d/2} \epsilon^d} \right].$$

PROOF. The idea is to apply the following inequality (see Vershynin (2018)), $\forall \Theta \subset \mathbb{R}^d$,

$$(2.2.10) \quad N(\Theta, \epsilon) \leq \frac{\text{Vol}_d(\Theta \oplus \frac{\epsilon}{2})}{\text{Vol}_d(B(0, \frac{\epsilon}{2}))}.$$

For $N(S, \epsilon)$, we first calculate an upper bound of the area of $\partial(S \oplus \frac{\epsilon}{2})$, then apply the isoperimetric inequality to get an upper bound of $\text{Vol}_d(S \oplus \frac{\epsilon}{2})$. We can always “flip” the concave part of S so that it becomes convex. Let S^* be the convex body constructed from S by this way. Then $S \subset S^*$ and $\text{Vol}_{d-1}(\partial S) = \text{Vol}_{d-1}(\partial S^*)$. And also S^* satisfies the rolling ball condition with radius R . We thus have

$$(2.2.11) \quad \frac{\text{Vol}_{d-1}(\partial(S^* \oplus \frac{\epsilon}{2}))}{\text{Vol}_{d-1}(\partial S^*)} \leq \frac{\text{Vol}_{d-1}(\partial B(0, R + \frac{\epsilon}{2}))}{\text{Vol}_{d-1}(\partial B(0, R))} = \left(1 + \frac{\epsilon}{2R}\right)^{d-1}.$$

Therefore,

$$(2.2.12) \quad \text{Vol}_{d-1}(\partial(S \oplus \frac{\epsilon}{2})) \leq \text{Vol}_{d-1}(\partial(S^* \oplus \frac{\epsilon}{2})) \leq \left(1 + \frac{\epsilon}{2R}\right)^{d-1} \text{Vol}_{d-1}(\partial S).$$

By the isoperimetric inequality,

$$(2.2.13) \quad \text{Vol}_d(S \oplus \frac{\epsilon}{2}) \leq \text{Vol}_d(S^* \oplus \frac{\epsilon}{2}) \leq \frac{\left(1 + \frac{\epsilon}{2R}\right)^d \text{Vol}_{d-1}(\partial S)^{\frac{d}{d-1}}}{(d^d V_d)^{\frac{1}{d-1}}}.$$

So

$$(2.2.14) \quad \begin{aligned} N(S, \epsilon) &\leq \frac{\text{Vol}_d(S \oplus \frac{\epsilon}{2})}{\text{Vol}_d(B(0, \frac{\epsilon}{2}))} \\ &\leq \frac{\left(1 + \frac{\epsilon}{2R}\right)^d \text{Vol}_{d-1}(\partial S)^{\frac{d}{d-1}}}{(d^d V_d)^{\frac{1}{d-1}} V_d \frac{\epsilon^d}{2^d}} \\ &= \left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(\frac{d}{2} + 1)}{d \pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon} + \frac{1}{R} \right)^d. \end{aligned}$$

For $N(\partial S, \epsilon)$, we use $2(\max\{\text{Vol}_d(S^* \oplus \frac{\epsilon}{2})\} - \text{Vol}_d(S))$ to bound $\text{Vol}_d(\partial S \oplus \frac{\epsilon}{2})$.

$$(2.2.15) \quad \begin{aligned} N(\partial S, \epsilon) &\leq \frac{2(\max\{\text{Vol}_d(S^* \oplus \frac{\epsilon}{2})\} - \text{Vol}_d(S))}{\text{Vol}_d(B(0, \frac{\epsilon}{2}))} \\ &\leq 2 \left[\left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{d \pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon} + \frac{1}{R} \right)^d - \frac{2^d \text{Vol}_d(S) \Gamma(d/2 + 1)}{\pi^{d/2} \epsilon^d} \right]. \end{aligned}$$

□

2.3. k -Nearest Neighbor Distance

2.3.1. Definition.

DEFINITION 2.3.1 (k -nearest neighbor distance). $R_k(x) = \|x - X_{(k)}(x)\|$ is the k -nearest neighbor distance of x with respect to the set $\mathcal{X}_n$.

The following theorem states an upper bound for $\sup_{x \in S} R_k(x)$ and a lower bound for $\inf_{x \in S} R_k(x)$. Under some regularity conditions, there exist constants C_1 and C_2 such that $C_1 \left(\frac{k}{n}\right)^{\frac{1}{d+\alpha}} \leq R_k(x) \leq C_2 \left(\frac{k}{n}\right)^{\frac{1}{d+\alpha}}$ for all $x \in S$, with probability at least $1 - \delta_n$, where $\delta_n \rightarrow 0$ as $n \rightarrow \infty$. Actually, when the number of nearest neighbors k is chosen appropriately, δ_n vanishes fast enough such that the bounds hold almost surely. It is later used to establish the connections between $\hat{H}_{k_n}$ and $\hat{H}_{r_n}$ for appropriate choices of k_n and r_n .

DEFINITION 2.3.2 (eventually almost surely (e.a.s.)). *A sequence (E_n) of events in the same probability space will happen eventually almost surely if $\mathbb{1}_{E_n}$ is such that $\mathbb{1}_{E_n} \rightarrow 1$ a.s. as $n \rightarrow \infty$.*

2.3.2. Concentration Type Results on the Uniform Bounds. Let X be a random variable with a distribution function F . Throughout the text, we use $F(B(x, r))$ and $P_X(B(x, r))$ interchangeably to denote the probability of X falls in the ball $B(x, r)$, i.e., $P(X \in B(x, r))$.

THEOREM 2.3.1 (uniform bounds on k -nearest neighbor distance). *Let $a \in (1, \infty)$, $b \in (0, 1)$.*

I. *If*

- (a) $\frac{2^8}{(a-1)^2} \log^2 \frac{2}{\delta} d \log n \leq k \leq \frac{C_0 R^{d+\alpha}}{a} n$ for some $\delta \in (0, 1)$,
- (b) $\exists R > 0$, $\alpha \geq 0$ and $C_0 > 0$ such that $\forall x \in S$, $F(B(x, r)) \geq C_0 r^{d+\alpha}$ for all $0 \leq r \leq R$.

then with probability at least $1 - \delta$,

$$(2.3.1) \quad \sup_{x \in S} R_k(x) \leq \left(\frac{a}{C_0} \frac{k}{n} \right)^{\frac{1}{d+\alpha}}.$$

Moreover, when $\frac{2^8}{(a-1)^2} \log^2(2n^2) d \log n \leq k \leq \frac{C_0 R^{d+\alpha}}{a} n$,

$$(2.3.2) \quad \sup_{x \in S} R_k(x) \leq \left(\frac{a}{C_0} \frac{k}{n} \right)^{\frac{1}{d+\alpha}} \text{ e.a.s. as } n \rightarrow \infty.$$

II. *If*

- (c) $\frac{2^8}{(1-b)^2} \log^2 \frac{2}{\delta} d \log n \leq k \leq n$ for some $\delta \in (0, 1)$,
- (d) *there exists $\beta \geq 0$ and $C_\infty > 0$ such that $\forall x \in S$, $F(B(x, r)) \leq C_\infty r^{d-\beta}$,*

then with probability at least $1 - \delta$,

$$(2.3.3) \quad \inf_{x \in S} R_k(x) \geq \left(\frac{b}{C_\infty} \frac{k}{n} \right)^{\frac{1}{d-\beta}}.$$

Moreover, when $k \geq \frac{2^8}{(1-b)^2} \log^2(2n^2) d \log n$,

$$(2.3.4) \quad \inf_{x \in S} R_k(x) \geq \left(\frac{b}{C_\infty} \frac{k}{n} \right)^{\frac{1}{d-\beta}} \quad e.a.s. \text{ as } n \rightarrow \infty.$$

PROOF. **I.** Let $r_1 = \left(\frac{a}{C_0} \frac{k}{n} \right)^{\frac{1}{d+\alpha}}$. (a) implies that $0 \leq r_1 \leq R$. By (b), $\forall x \in S$,

$$(2.3.5) \quad \begin{aligned} F(B(x, r_1)) &\geq C_0 r_1^{d+\alpha} \\ &= C_0 \left(\frac{a}{C_0} \frac{k}{n} \right) \\ &= a \frac{k}{n} \\ &= \frac{k}{n} + (a-1) \sqrt{k} \frac{\sqrt{k}}{n} \\ &\geq \frac{k}{n} + C_{\delta,n} \frac{\sqrt{k}}{n}. \end{aligned}$$

The last inequality is due to the fact that (a) implies that $\sqrt{k} \geq C_{\delta,n} = 2^4 \log(2/\delta) \sqrt{d \log n} / (a-1)$.

Therefore, by Lemma (2.7.1) in the Appendix, with probability at least $1 - \delta$, $F_n(B(x, r_1)) \geq \frac{k}{n}$ for all x . This means $\sup_{x \in S} R_k(x) \leq r_1$.

For the almost sure convergence, let k be such that $\frac{2^8}{(a-1)^2} \log^2(2n^2) d \log n \leq k \leq \frac{C_0 R^{d+\alpha}}{a} n$. Let

$$(2.3.6) \quad D_n = \left\{ \sup_{x \in S} R_k(x) \leq r_1 \right\}.$$

So $P(D_n) \geq 1 - 1/n^2$. Then

$$(2.3.7) \quad \sum_n P(D_n^c) \leq \sum_n \frac{1}{n^2} < \infty.$$

Therefore by Borel-Cantelli lemma,

$$(2.3.8) \quad P(D_n^c \text{ happens i.o.}) = 0.$$

Let $Y_n = \sup_x R_k(x) - r_1$. Then $D_n = \{Y_n \leq 0\}$. So

$$(2.3.9) \quad P(Y_n \leq 0 \text{ happens all but finitely many } n) = 1.$$

But

$$(2.3.10) \quad \begin{aligned} Y_n \leq 0 \text{ happens all but finitely many } n &\iff \exists m \geq 1 \text{ such that } \forall n \geq m, Y_n \leq 0 \\ &\iff \inf_{m \geq 1} \sup_{n \geq m} Y_n \leq 0 \\ &\iff \limsup_{n \rightarrow \infty} Y_n \leq 0 \end{aligned}$$

So

$$(2.3.11) \quad P(\limsup_{n \rightarrow \infty} Y_n \leq 0) = 1.$$

II. Let $r_2 = \left(\frac{a}{C_\infty} \frac{k}{n}\right)^{\frac{1}{d-\beta}}$. Then by (d) $\forall x \in S$,

$$(2.3.12) \quad \begin{aligned} F(B(x, r_2)) &\leq C_\infty r_2^{d-\beta} \\ &= C_\infty \left(\frac{b}{C_\infty} \frac{k}{n}\right) \\ &= \frac{k}{n} - (1-b)\sqrt{k} \frac{\sqrt{k}}{n} \\ &\leq \frac{k}{n} - C_{\delta,n} \frac{\sqrt{k}}{n} \end{aligned}$$

The last inequality is due to that (c) implies that $\sqrt{k} \geq C_{\delta,n} = 2^4 \log(2/\delta) \sqrt{d \log n} / (1-b)$.

Therefore, by Lemma (2.7.1) in the Appendix, with probability $1 - \delta$, $F_n(B(x, r_2)) < \frac{k}{n}$ for all x . This means $\inf_{x \in S} R_k(x) \geq r_2$.

For the almost sure convergence, let $k \geq \frac{2^8}{(1-b)^2} \log^2(2n^2)d \log n$. Let

$$(2.3.13) \quad E_n = \left\{ \inf_{x \in S} R_k(x) \geq r_2 \right\}.$$

So $P(E_n) \geq 1 - 1/n^2$. Then

$$(2.3.14) \quad \sum_n P(E_n^c) \leq \sum_n \frac{1}{n^2} < \infty.$$

Therefore by Borel-Cantelli lemma,

$$(2.3.15) \quad P(E_n^c \text{ happens i.o.}) = 0.$$

Let $W_n = \inf_x R_k(x) - r_1$. Then $E_n = \{W_n \geq 0\}$. So

$$(2.3.16) \quad P(W_n \geq 0 \text{ happens all but finitely many } n) = 1.$$

But

$$(2.3.17) \quad \begin{aligned} W_n \geq 0 \text{ happens all but finitely many } n &\iff \exists m \geq 1 \text{ such that } \forall n \geq m, W_n \geq 0 \\ &\iff \sup_{m \geq 1} \inf_{n \geq m} W_n \geq 0 \\ &\iff \liminf_{n \rightarrow \infty} W_n \geq 0 \end{aligned}$$

So

$$(2.3.18) \quad P(\liminf_{n \rightarrow \infty} W_n \geq 0) = 1.$$

□

Conditions (b) and (d) in Theorem 2.3.1 can be satisfied under some additional regularity conditions on the density functions and the shape of S . For example, consider the case when the density $f(x) \geq f_0 > 0$ is bounded away from zero, and $\forall x \in S, \exists \gamma > 0, R > 0$ such that $\text{Leb}(S \cap B(x, r)) \geq \gamma \text{Leb}(B(x, r))$ for $0 < r \leq R$. Then $F(B(x, r)) \geq \gamma f_0 V_d r^d$ for $0 \leq r \leq R$. The

corollary below (the first part of which is the same as Lemma 3 in Jiang (2019) when $a = 2$) follows immediately.

COROLLARY 2.3.1. *Let $a \in (1, \infty)$. If*

- (a) $f_0 = \inf_{x \in S} f(x) > 0$,
- (b) $\forall x \in S, \exists \gamma > 0, R > 0$ such that $\text{Leb}(S \cap B(x, r)) \geq \gamma \text{Leb}(B(x, r))$ for $0 < r \leq R$.

Then $F(B(x, r)) \geq \gamma f_0 V_d r^d$ and condition (a) in Theorem 2.3.1 is satisfied with $C_0 = \gamma f_0 V_d$, then with probability at least $1 - \delta$,

$$(2.3.19) \quad \sup_{x \in S} R_k(x) \leq \left(\frac{a}{\gamma f_0 V_d} \frac{k}{n} \right)^{\frac{1}{d}}.$$

Moreover,

$$(2.3.20) \quad \sup_{x \in S} R_k(x) \leq \left(\frac{a}{\gamma f_0 V_d} \frac{k}{n} \right)^{\frac{1}{d}} \quad \text{e.a.s. as } n \rightarrow \infty,$$

$$\text{when } \frac{2^8}{(a-1)^2} \log^2(2n^2) d \log n \leq k \leq \frac{\gamma f_0 V_d R^{d+\alpha}}{a} n.$$

DEFINITION 2.3.3 (α -quickly decreasing). *Let f be a density function with support $S \subset \mathbb{R}^d$. f is α -quickly decreasing if there exists $\alpha \geq 0$ and $C_f > 0$ such that*

$$\forall x \in S, f(x) \geq C_f d(x, \partial S)^\alpha.$$

When f is α -quickly decreasing and S satisfies the rolling-ball condition, the following holds.

COROLLARY 2.3.2. *Let $a \in (1, \infty)$. If*

- (a) $2^8 \log^2 \frac{2}{\delta} d \log n \leq k \leq \frac{C_f V_{d-1} B(\frac{\alpha+1}{2}, \frac{d+1}{2}) R^{d+\alpha}}{3a} n$, where $B(., .)$ is the beta function,
- (b) f is α -quickly decreasing, i.e., there exists $\alpha \geq 0$ and $C_f > 0$ such that $\forall x \in S, f(x) \geq C_f d(x, \partial S)^\alpha$,
- (c) balls of radius R roll freely inside S ,

then with probability at least $1 - \delta$, when $0 \leq r_n \leq ((3/2)^{1/(\alpha+d)} - 1)R$,

$$(2.3.21) \quad \sup_{x \in S} R_k(x) \leq \left(\frac{3a}{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)} \frac{k}{n} \right)^{\frac{1}{d+\alpha}}.$$

Moreover, when $2^8 \log^2(2n^2) d \log n \leq k \leq \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) R^{d+\alpha}}{3a} n$,

$$(2.3.22) \quad \sup_{x \in S} R_k(x) \leq \left(\frac{3a}{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)} \frac{k}{n} \right)^{\frac{1}{d+\alpha}} \quad e.a.s. \text{ as } n \rightarrow \infty.$$

PROOF. Inequality (2.5.6) implies that $\forall x \in S$

$$(2.3.23) \quad F(B(x, r)) \geq \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)}{3} r^{d+\alpha}$$

for all $0 \leq r \leq ((3/2)^{1/(\alpha+d)} - 1)R$. Then let $C_0 = \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)}{3}$, the conclusion follows by Theorem 2.3.1.

□

Also note that $V_d = V_{d-1} B\left(\frac{1}{2}, \frac{d+1}{2}\right)$ and Lemma 2.2.1. Corollary 2.3.1 agrees with Corollary 2.3.2.

The following corollary is when f is bounded from above.

COROLLARY 2.3.3. *Let $b \in (0, 1)$.*

$$(a) \quad \frac{2^8}{(1-b)^2} \log^2 \frac{2}{\delta} d \log n \leq k \leq n,$$

$$(b) \quad f_\infty = \sup_{s \in S} f(s) < \infty.$$

then with probability $1 - \delta$,

$$(2.3.24) \quad \inf_{x \in S} R_k(x) \geq \left(\frac{b}{f_\infty V_d} \frac{k}{n} \right)^{\frac{1}{d}}.$$

Moreover,

$$(2.3.25) \quad \inf_{x \in S} R_k(x) \geq \left(\frac{b}{f_\infty V_d} \frac{k}{n} \right)^{\frac{1}{d}} \quad e.a.s. \text{ as } n \rightarrow \infty,$$

when $k \geq \frac{2^8}{(1-b)^2} \log^2(2n^2)d \log n$.

The proof is a straight-forward application of Theorem 2.3.1 since $F(B(x, r)) \leq f_\infty V_d r^d$.

2.4. Convergence of the k NN LoCoH Estimator $\hat{H}_{k_n}$

2.4.1. Consistency. The following theorem states that under some regularity conditions on S and f , when k_n is carefully chosen, $\text{Leb}(\hat{H}_{k_n} \Delta S) = O((\log n/n)^{\frac{2}{d+1}})$.

THEOREM 2.4.1 (consistency). *Let $\mathcal{X}_n$ be an i.i.d. sample from f . Let $S \subset \mathbb{R}^d$ be the support of f . If*

- (a) S is compact d -manifold,
- (b) balls of radius R roll freely inside and outside of S for some $R > 0$,
- (c) $f_\infty = \sup_{x \in S} f(x) < \infty$,
- (d) $f_0 = \inf_{x \in S} f(x) > 0$,
- (e) $k_n = \eta(\log n)^{\frac{d}{d+1}} n^{\frac{1}{d+1}}$ for some $\eta > 0$,

then the k -nearest neighbor local convex hull estimator $\hat{H}_k$ of S is such that

$$\text{Leb}(\hat{H}_{k_n} \Delta S)(n/\log n)^{\frac{2}{d+1}} \text{ is e.a.s. bounded.}$$

PROOF. Fix $b \in (0, 1)$, $a \in (1, \infty)$. For convenience, write $k = k_n$.

By Corollary 2.3.3, (c) and (e) imply

$$(2.4.1) \quad \inf_{x \in S} R_k(x) \geq \left(\frac{b}{f_\infty V_d} \frac{k}{n} \right)^{1/d} \text{ e.a.s. as } n \rightarrow \infty.$$

By Lemma 2.2.1, (b) implies that there exists some $\gamma \geq \frac{1}{2} I_{\frac{3}{4}}(\frac{d+1}{2}, \frac{1}{2})$ such that $\text{Leb}(S \cap B(x, r)) \geq \gamma \text{Leb}(B(x, r))$ for $0 < r \leq R$. This together with (d) and (e) implies

$$(2.4.2) \quad \sup_{x \in S} R_k(x) \leq \left(\frac{a}{\gamma f_0 V_d} \frac{k}{n} \right)^{1/d} \text{ e.a.s. as } n \rightarrow \infty,$$

by Corollary 2.3.1.

Now pick

$$r_{n_1} = \left(\frac{bk}{f_\infty V_d n} \right)^{\frac{1}{d}} = \left(\frac{b\eta}{f_\infty V_d} \right)^{\frac{1}{d}} \left(\frac{\log n}{n} \right)^{\frac{1}{d+1}},$$

$$r_{n_2} = \left(\frac{ak}{\gamma f_0 V_d n} \right)^{\frac{1}{d}} = \left(\frac{a\eta}{\gamma f_0 V_d} \right)^{\frac{1}{d}} \left(\frac{\log n}{n} \right)^{\frac{1}{d+1}}.$$

So when $n \rightarrow \infty$,

$$\hat{H}_{r_{n_1}} \subset \hat{H}_k \subset \hat{H}_{r_{n_2}} \text{ e.a.s.}$$

This will give us

$$S \setminus \hat{H}_k \subset S \setminus \hat{H}_{r_{n_1}} \subset H_{r_{n_1}} \Delta S,$$

$$\hat{H}_k \setminus S \subset \hat{H}_{r_{n_2}} \setminus S \subset H_{r_{n_1}} \Delta S.$$

By Theorem 2 in Aaron and Bodart (2016) (see appendix 2.7.1), (abd) implies

$$\text{Leb}(\hat{H}_{r_{n_1}} \Delta S)(n/\log n)^{\frac{2}{d+1}} \text{ is e.a.s. bounded,}$$

$$\text{Leb}(\hat{H}_{r_{n_2}} \Delta S)(n/\log n)^{\frac{2}{d+1}} \text{ is e.a.s. bounded.}$$

Therefore,

$$\text{Leb}(\hat{H}_k \Delta S)(n/\log n)^{\frac{2}{d+1}} \leq \text{Leb}(\hat{H}_{r_{n_1}} \Delta S)(n/\log n)^{\frac{2}{d+1}} + \text{Leb}(\hat{H}_{r_{n_2}} \Delta S)(n/\log n)^{\frac{2}{d+1}}$$

is eventually almost surely bounded.

□

REMARK 2.4.1. *The consistency also hold if we relax the assumption on f to be α -quickly decreasing because of Corollary 2.3.2.*

2.4.2. Discussion of Rates of Convergence. Dümbgen and Walther (1996) has shown that under moderate regularity conditions, including a smoothness condition similar to Proposition 2.2.1, the rate of convergence of the Hausdorff distance for a convex hull estimator is $O((\log n/n)^{2/(d+1)})$

when the underlying set to estimate is a convex compact body. Under a stronger regularity conditions on the boundary, the expectation of the Hausdorff distance is exactly of order $(\log n/n)^{2/(d+1)}$. In the above, we have shown that the rate of convergence of the symmetric difference distance is also $O((\log n/n)^{2/(d+1)})$.

2.5. Concentration Bounds for the r -ball LoCoH Estimator $\hat{H}_{r_n}$

In this section, we follow the main idea in the Aaron and Bodart (2016). Instead of proving the consistency, we derive a concentration bound for $\hat{H}_{r_n}$. Unless specified, the following notations and assumptions hold throughout this section.

Let $\mathcal{X}_n$ be an i.i.d. sample in $\mathbb{R}^d$ associated with a density function f .

ASSUMPTION A1. S is compact d -manifold.

ASSUMPTION A2. Balls of radius R roll freely inside S for some $R > 0$.

ASSUMPTION A3. Balls of radius R roll freely outside S for some $R > 0$.

ASSUMPTION A4. f is α -quickly decreasing.

2.5.1. Three Supporting Lemmas. We first prove three supporting lemmas. The first one (Lemma 2.5.1) is to bound the Hausdorff distance between S and the data cloud $\mathcal{X}_n$. The second one (Lemma 2.5.2) provides a way to construct a smaller cylinder inside a given cylinder under some conditions on the centroids and their directions, which will be used to construct tangent cylinders. And the last one (Lemma 2.5.3) is to bound the probability of having tangent cylinders non-intersecting with $\mathcal{X}_n$.

LEMMA 2.5.1. Let r_n be a small positive number and $0 < \epsilon_n < r_n < R$.

$$\begin{aligned}
& P(d_H(S, \mathcal{X}_n) \geq r_n) \\
(2.5.1) \quad & \leq N(S, \epsilon_n) \left(1 - \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)}{2} (r_n - \epsilon_n)^{\alpha+d} \left(1 - \min(1, (\alpha+d) \frac{r_n - \epsilon_n}{R}) \right) \right)^n \\
& \leq e^{\log N(S, \epsilon_n) - \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) n r_n^{\alpha+d}} \left(1 - (\alpha+d) \frac{\epsilon_n}{r_n} - \min(1, (\alpha+d) \frac{r_n - \epsilon_n}{R}) \right),
\end{aligned}$$

where $B(\cdot, \cdot)$ is the Beta function and $N(S, \epsilon_n)$ is the covering number of S .

DISCUSSION. r_n is the radius of the balls used to build the r_n -ball LoCoH estimator. ϵ_n is the radius of the small balls used to cover S . We can see that the ratio $\frac{\epsilon_n}{r_n}$ needs to be small enough so that $(\alpha + d)\frac{\epsilon_n}{r_n}$ is small. Particularly, when the dimension d is not fixed and goes to ∞ , this ratio needs to vanish faster enough so that the right-hand side of the inequality can go to 0.

For example, if r_n and ϵ_n are chosen such that the following conditions are satisfied, the right-hand side will converge to 0 as $n \rightarrow \infty$.

$$\begin{aligned}
 (2.5.2) \quad & \log N(S, \epsilon_n) - \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) n r_n^{\alpha+d}}{2} (1 - \delta) \rightarrow -\infty \\
 & \frac{(\alpha + d)\epsilon_n}{r_n} < \delta/2 < 1 \\
 & \frac{(\alpha + d)(r_n - \epsilon_n)}{R} < \delta/2 < 1
 \end{aligned}$$

for some $0 < \delta < 1$.

□

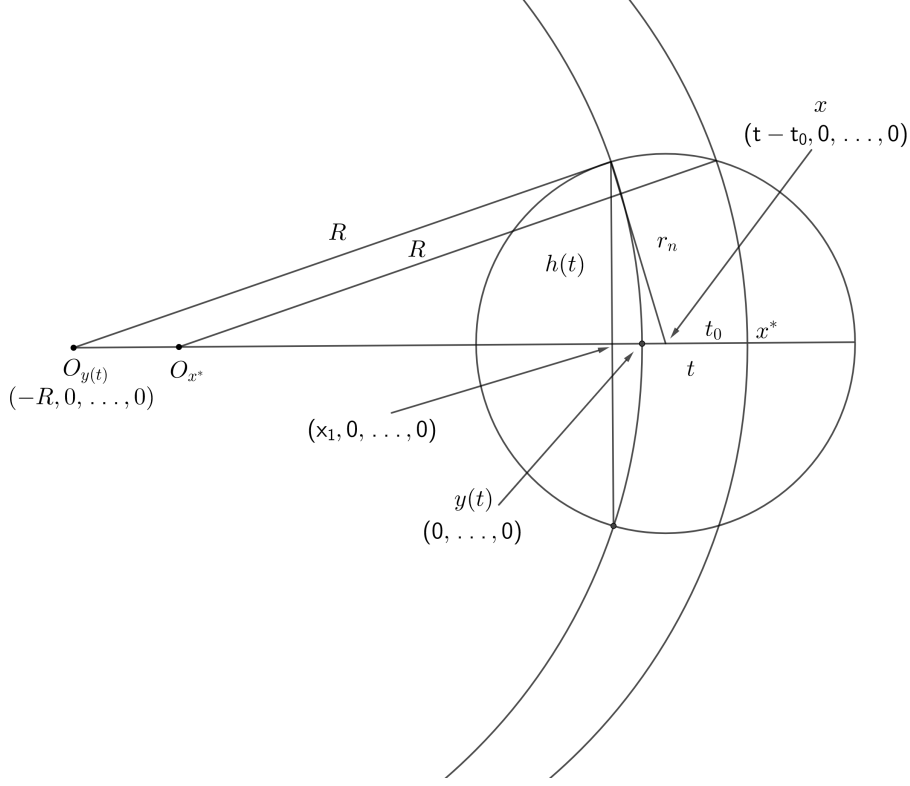


FIGURE 2.2. Lemma 2.5.1. $x \in S$ is such that $B(x, r_n) \cap \partial S \neq \emptyset$. x^* is the projection of x onto ∂S . $t_0 = d(x, x^*)$. $y(t)$ is a point such that $y(t)$, x and x^* are collinear. And $y(t) \in B(x, r_n)$. $t = d(y(t), x^*)$. So $t_0 \leq t \leq t_0 + r_n$.

PROOF. Let $x \in S$. The worst case scenario is when x is such that $B(x, r_n) \not\subset S$. We now derive a bound for $P(X \in B(x, r_n))$, where $X \sim f$. For $t_0 \leq t \leq t_0 + r_n$, consider the two spheres depicted in Figure 2.2 given by the equations

$$(2.5.3) \quad \begin{cases} (X_1 + R)^2 + \cdots + X_d^2 = R^2 \\ (X_1 - (t - t_0))^2 + \cdots + X_d^2 = r_n^2, \end{cases}$$

$$x_1 = \frac{1}{2} \frac{(t - t_0)^2 - r_n^2}{R + t - t_0}.$$

Denote $\mathcal{A}(t) = (\partial B(O_{y(t)}, R)) \cap B(x, r_n)$. Pick any point $z \in \mathcal{A}(t)$ then

$$\begin{aligned}
d(z, \partial S) &\geq R - d(z, O_{x^*}) = R - \sqrt{(x_1 - t + R)^2 + R^2 - (x_1 + R)^2} \\
&= R - \sqrt{(R - t)^2 - 2tx_1} \\
&\geq R - (R - t) + \frac{x_1 t}{R - t} \\
&= t + \frac{x_1 t}{R - t}.
\end{aligned}
\tag{2.5.4}$$

$$\begin{aligned}
h(t) &= \sqrt{r_n^2 - ((t - t_0) - x_1)^2} \\
&= \sqrt{r_n^2 - \left(\frac{(R + t - t_0)^2 - R^2 + r_n^2}{2(R + t - t_0)} \right)^2} \\
&= r_n \sqrt{\frac{R^2}{(R + t - t_0)^2} - \left(\frac{2R(t - t_0) + (t - t_0)^2 - r_n^2}{2r_n(R + t - t_0)} \right)^2} \\
&\geq r_n \sqrt{\frac{R^2}{(R + r_n)^2} - \left(\frac{2R(t - t_0) + (t - t_0)^2 - r_n^2}{2r_n(R + t - t_0)} \right)^2} \\
&= r_n \cdot \frac{R}{R + r_n} \sqrt{1 - \left(\frac{R + r_n}{R} \frac{2R(t - t_0) + (t - t_0)^2 - r_n^2}{2r_n(R + t - t_0)} \right)^2} \\
&\geq \frac{Rr_n}{R + r_n} \sqrt{1 - \left(\frac{R + r_n}{R} \frac{2R(t - t_0) + r_n^2 - r_n^2}{2r_n(R + 0)} \right)^2} \\
&= \frac{Rr_n}{R + r_n} \sqrt{1 - \left(\frac{R + r_n}{Rr_n} (t - t_0) \right)^2}.
\end{aligned}
\tag{2.5.5}$$

So

(2.5.6)

$$\begin{aligned}
P(X \in B(x, r_n)) &\geq \int_{t_0}^{t_0+r_n} t^\alpha C_f V_{d-1} \left(\frac{Rr_n}{R+r_n} \right)^{d-1} \left(1 - \left(\frac{R+r_n}{Rr_n} (t-t_0) \right)^2 \right)^{\frac{d-1}{2}} dt \\
&\geq \int_{t_0}^{t_0+\frac{R}{R+r_n}r_n} t^\alpha C_f V_{d-1} \left(\frac{Rr_n}{R+r_n} \right)^{d-1} \left(1 - \left(\frac{R+r_n}{Rr_n} (t-t_0) \right)^2 \right)^{\frac{d-1}{2}} dt \\
&\text{Let } u = \left(\frac{R+r_n}{Rr_n} (t-t_0) \right)^2 \text{ and notice } t \geq t-t_0 \geq 0 \\
&\geq \int_{t_0}^{t_0+\frac{R}{R+r_n}r_n} (t-t_0)^\alpha C_f V_{d-1} \left(\frac{Rr_n}{R+r_n} \right)^{d-1} \left(1 - \left(\frac{R+r_n}{Rr_n} (t-t_0) \right)^2 \right)^{\frac{d-1}{2}} dt \\
&= C_f V_{d-1} \left(\frac{Rr_n}{R+r_n} \right)^{\alpha+d} \int_0^1 u^{\frac{\alpha-1}{2}} (1-u)^{\frac{d-1}{2}} du \\
&= \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)}{2} \left(\frac{R}{R+r_n} \right)^{\alpha+d} \cdot r_n^{\alpha+d}.
\end{aligned}$$

Since $\frac{R}{R+r_n} \geq 1 - \frac{r_n}{R}$, we have

$$(2.5.7) \quad P(X \in B(x, r_n)) \geq \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)}{2} r_n^{\alpha+d} \left(1 - \min(1, (\alpha+d) \frac{r_n}{R}) \right).$$

Let us cover S with $N(S, \epsilon_n)$ balls of radius ϵ_n centered at points $x_j \in S$, for $j \in \{1, \dots, N(\epsilon_n)\}$.

By assumption, $\epsilon_n < r_n$. So

$$(2.5.8) \quad P(\sup_{x \in S} \min_i \|x - X_i\| \geq r_n) \leq P(\max_{x_j} \min_i \|x_j - X_i\| \geq r_n - \epsilon_n)$$

So,

(2.5.9)

$$\begin{aligned}
P(d_H(S, \mathcal{X}_n) \geq r_n) &\leq N(S, \epsilon_n) \left(1 - \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)}{2} (r_n - \epsilon_n)^{\alpha+d} \left(1 - \min(1, (\alpha+d) \frac{r_n - \epsilon_n}{R}) \right) \right)^n \\
&\leq N(S, \epsilon_n) e^{-\frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right)}{2} (r_n - \epsilon_n)^{\alpha+d} \left(1 - \min(1, (\alpha+d) \frac{r_n - \epsilon_n}{R}) \right) n} \\
&= e^{\log N(S, \epsilon_n) - \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) r_n^{\alpha+d}}{2} \left(1 - \frac{\epsilon_n}{r_n} \right)^{\alpha+d} \left(1 - \min(1, (\alpha+d) \frac{r_n - \epsilon_n}{R}) \right) n} \\
&\leq e^{\log N(S, \epsilon_n) - \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) n r_n^{\alpha+d}}{2} \left(1 - (\alpha+d) \frac{\epsilon_n}{r_n} - \min(1, (\alpha+d) \frac{r_n - \epsilon_n}{R}) \right)}
\end{aligned}$$

Notice that in some bizarre cases, $1 - (\alpha+d) \frac{\epsilon_n}{r_n} - \min(1, (\alpha+d) \frac{r_n - \epsilon_n}{R})$ can get negative, the above bound becomes trivially true. $\square$

DEFINITION 2.5.1 (cylinder). *A cylinder in $\mathbb{R}^d$ is defined as a point set*

$$(2.5.10) \quad \mathcal{C}^u(x, r, h) = \{y \in \mathbb{R}^d : |\langle y - x, u \rangle| \leq h, \|y - x - \langle y - x, u \rangle u\| \leq r\},$$

where $x \in \mathbb{R}^d$ is the centroid, u is the unit vector in $\mathbb{R}^d$, $r \in \mathbb{R}^+$ is the radius of the basis, $h \in \mathbb{R}^+$ is half of the height.

The following lemma states that given a cylinder that is centered at x and has direction u_x , we are able to find an embedded cylinder centered at y with direction u_y when both the two points x and y and their directions u_x and u_y are close enough (smoothness condition will ensure that if x and y are close, then u_x and u_y are close).

LEMMA 2.5.2. *Given a cylinder $\mathcal{C}^u(x, r, h)$ in $\mathbb{R}^d$, we can find another cylinder $\mathcal{C}^{u'}(x', r', h') \subset \mathcal{C}^u(x, r, h)$ in $\mathbb{R}^d$, as long as the following conditions are satisfied:*

- (a) $0 < r' \leq r - e_x - 2e_u(\sqrt{r^2 + h^2} + e_x)$,
- (b) $0 < h' \leq h - e_x - e_u\sqrt{r^2 + h^2}$,

where $e_x := \|x - x'\|$ and $e_u := \|u - u'\|$.

PROOF. Let $y \in C^{u'}(x', r', h')$,

$$\begin{aligned}
(2.5.11) \quad \|y - x'\|^2 &= \|\langle y - x', u' \rangle u' + y - x' - \langle y - x', u' \rangle u'\|^2 \\
&= \|\langle y - x', u' \rangle u'\|^2 + \|y - x' - \langle y - x', u' \rangle u'\|^2 \\
&\leq h'^2 + r'^2.
\end{aligned}$$

Then,

$$\begin{aligned}
(2.5.12) \quad |\langle y - x, u \rangle| &= |\langle y - x' + x' - x, u' - u' + u \rangle| \\
&= |\langle y - x', u' \rangle + \langle y - x', u - u' \rangle + \langle x' - x, u \rangle| \\
&\leq |\langle y - x', u' \rangle| + |\langle y - x', u - u' \rangle| + |\langle x' - x, u \rangle| \\
&\leq h - e_x - \sqrt{r^2 + h^2}e_u + \|y - x'\| \|u - u'\| + \|x' - x\| \|u\| \\
&\leq h - e_x - \sqrt{r^2 + h^2}e_u + e_u \sqrt{r'^2 + h'^2} + e_x \\
&\leq h - e_x - \sqrt{r^2 + h^2}e_u + e_u \sqrt{r^2 + h^2} + e_x \\
&= h.
\end{aligned}$$

And we also have

$$\begin{aligned}
(2.5.13) \quad \|\langle y - x, u \rangle u - \langle y - x, u' \rangle u'\| &= \|\langle y - x, u \rangle u - \langle y - x, u \rangle u' + \langle y - x, u \rangle u' - \langle y - x, u' \rangle u'\| \\
&= \|\langle y - x, u \rangle (u - u') + \langle y - x, u' \rangle u'\| \\
&\leq 2 \|y - x\| \|u - u'\| \\
&= 2e_u \|y - x\| \\
&\leq 2e_u (\sqrt{r'^2 + h'^2} + e_x).
\end{aligned}$$

So that

(2.5.14)

$$\begin{aligned}
\|y - x - \langle y - x, u \rangle u\| &= \|y - x' - \langle y - x', u' \rangle u' + x' - x + \langle y - x', u' \rangle u' - \langle y - x, u \rangle u\| \\
&\leq \|y - x' - \langle y - x', u' \rangle u'\| + \|x' - x\| + \|\langle y - x', u' \rangle u' - \langle y - x, u \rangle u\| \\
&\leq r - e_x - 2e_u(\sqrt{r^2 + h^2} + e_x) + e_x + 2e_u(\sqrt{r'^2 + h'^2} + e_x) \\
&\leq r.
\end{aligned}$$

So $y \in \mathcal{C}^u(x, r, h)$.

□

The following states that the probability of having a non-intersecting (with the data cloud) tangent cylinders on the boundary of S can be bounded from above.

LEMMA 2.5.3. *Let $x \in \partial S$. Consider a tangent cylinder $\mathcal{C}^{u_x}(x, t_n, h_n)$ with u_x being the inward pointing normal vector, such that for $\mu, \xi > 0$, we have $0 \leq \xi r_n - t_n \leq p_n$, $\xi r_n < \frac{R}{2}$ and $0 \leq \frac{\mu r_n^2}{2R} - h_n \leq q_n$. Furthermore, we assume*

(2.5.15)

$$m_n := \min \left(\frac{\frac{\xi}{2(d-1)} - \frac{p_n}{r_n}}{r_n \left(1 + \sqrt{\xi^2 + \frac{\mu^2}{16}} + R \right)}, \frac{\mu - \frac{2Rq_n}{r_n}}{R \left(2 + \sqrt{\xi^2 + \frac{\mu^2}{16}} \right)}, \frac{\frac{\mu - \xi^2}{2(\alpha+1)} - \frac{16\xi^3 r_n}{9R} - \frac{2Rq_n}{r_n}}{R \left(2 + \sqrt{\xi^2 + \frac{\mu^2}{16}} \right)} \right) > 0.$$

Then for $0 < \epsilon_n < m_n$,

(2.5.16)

$$P(\exists x \in \partial S; \mathcal{C}^{u_x}(x, t_n, h_n) \cap \mathcal{X}_n = \emptyset) \leq e^{\log N(\partial S, \epsilon_n r_n^2) - \frac{C_f V_{d-1} \xi^{d-1} (\mu - \xi^2)^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)(2R)^{\alpha+1}}} (1 - M_{\alpha, d, \mu, \xi}(p_n, q_n, r_n, \epsilon_n)),$$

where

(2.5.17)

$$M_{\alpha,d,\mu,\xi}(p_n, q_n, r_n, \epsilon_n) = \frac{2(\alpha+1)R}{(\mu-\xi^2)} \frac{q_n}{r_n^2} + \frac{2(\alpha+1)R}{(\mu-\xi^2)} \left(1 + \frac{\sqrt{\xi^2 + \mu^2/16}}{2}\right) \epsilon_n + \frac{(d-1)}{\xi} \frac{p_n}{r_n} \\ + \frac{(d-1)}{\xi} \left(1 + \sqrt{\xi^2 + \frac{\mu^2}{16} + \frac{R}{2}}\right) \epsilon_n r_n + \frac{16(\alpha+1)\xi^3}{9R(\mu-\xi^2)} r_n.$$

DISCUSSION. r_n is the radius of the balls used to build the r_n -ball LoCoH estimator. t_n is the radius of the base of the tangent cylinder. h_n is the height of the cylinder. p_n is the maximum deviation of t_n from ξr_n and q_n is the maximum deviation of h_n from $\frac{\mu r_n^2}{2R}$. So the volume of the cylinder $\mathcal{C}^{u_x}(x, t_n, h_n)$ is of the order $O(r_n^{d+1}) = O(r_n V_d(B(0, r_n)))$ when p_n and q_n are chosen appropriately. Therefore $\frac{\text{Vol}_d(\mathcal{C}^{u_x}(x, t_n, h_n))}{\text{Vol}_d(B(0, r_n))} = O(r_n)$. When $r_n \rightarrow 0$, this ratio also goes to 0.

For a fixed dimension d , let $p_n = o(r_n)$ and $q_n = o(r_n^2)$. When $n \rightarrow \infty$, the right-hand side of the Inequality (2.5.15) becomes a constant

$$(2.5.18) \quad m := \min \left(\frac{\mu}{R \left(2 + \sqrt{\xi^2 + \frac{\mu^2}{16}}\right)}, \frac{\frac{\mu - \xi^2}{2(\alpha+1)}}{R \left(2 + \sqrt{\xi^2 + \frac{\mu^2}{16}}\right)} \right).$$

Therefore $\limsup_{n \rightarrow \infty} \epsilon_n < m$. In the proof, we are going to cover ∂S using balls of radius $\epsilon_n r_n^2$, which is then of the order $O(r_n^2)$. Also this choice of p_n and q_n makes sure that $\limsup_{n \rightarrow \infty} M_{\alpha,d,\mu,\xi}(p_n, q_n, r_n, \epsilon_n) < \frac{1}{2}$.

So when r_n is chosen appropriately such that

$$(2.5.19) \quad \log N(\partial S, \epsilon_n r_n^2) - \frac{C_f V_{d-1} \xi^{d-1} (\mu - \xi^2)^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)(2R)^{\alpha+1}} \frac{1}{2} \rightarrow -\infty \text{ as } n \rightarrow \infty,$$

the probability $P(\exists x \in \partial S; \mathcal{C}^{u_x}(x, t_n, h_n) \cap \mathcal{X}_n = \emptyset)$ converges to 0 as $n \rightarrow \infty$. $\square$

PROOF. Let $x \in \partial S$, $0 < r < R$. Suppose $\mathcal{C}^{u_x}(x, r, h)$ is a tangent cylinder with radius r , height $2h$ and an inward pointing normal vector u_x . Suppose $X \sim f$. We now bound the probability content of $\mathcal{C}^{u_x}(x, r, h)$.

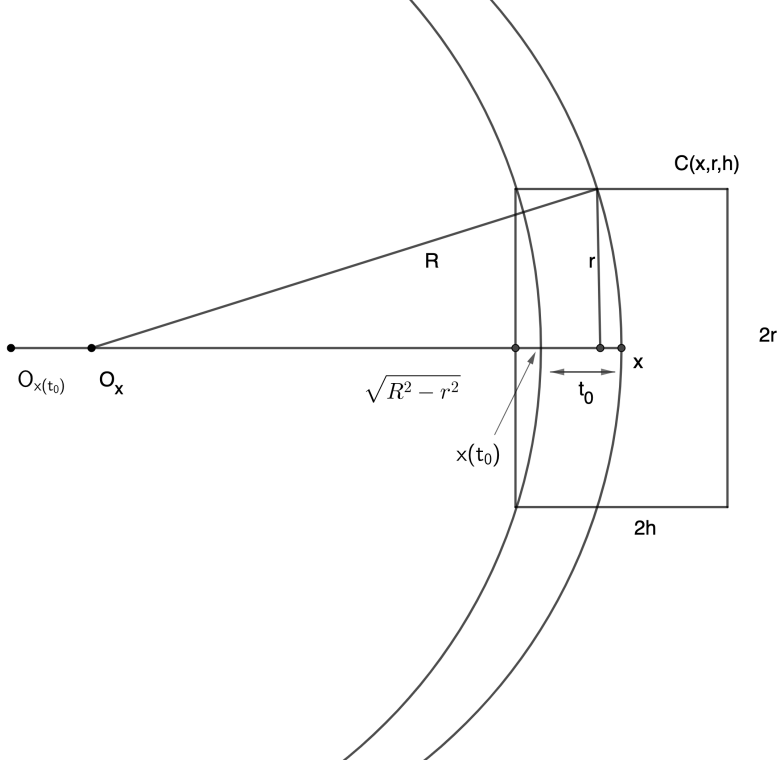


FIGURE 2.3. Lemma 2.5.3

Let $x(t) = x + u_x t$ and $\mathcal{A}(t) = \partial B(O_{x(t)}, R) \cap \mathcal{C}^{u_x}(x, r, h)$. Let $t_0 = h - R + \sqrt{R^2 - r^2}$ when $r < R$. When $0 \leq t \leq t_0$, $\text{Vol}_{d-1}(\mathcal{A}(t)) \geq V_{d-1} r^{d-1}$, see Figure 2.3.

$$\begin{aligned}
 P(X \in \mathcal{C}^{u_x}(x, r, h)) &\geq \int_0^h \text{Vol}_{d-1}(\mathcal{A}(t)) C_f t^\alpha dt \\
 (2.5.20) \qquad \qquad \qquad &\geq \int_0^{t_0} V_{d-1} r^{d-1} C_f t^\alpha dt \\
 &= \frac{C_f V_{d-1}}{\alpha + 1} r^{d-1} (h - R + \sqrt{R^2 - r^2})^{\alpha+1}
 \end{aligned}$$

With the following conditions:

- (1) $(d-1) \frac{p_n}{\xi r_n} < \frac{1}{2}$,
- (2) $(\alpha + 1) \left(\frac{2Rq_n}{(\mu - \xi^2)r_n^2} + \frac{16\xi^3 r_n}{9R(\mu - \xi^2)} \right) < \frac{1}{2}$,

we have

(2.5.21)

$$\begin{aligned}
P(X \in \mathcal{C}^{u_x}(x, t_n, h_n)) &\geq \frac{C_f V_{d-1}}{\alpha+1} t_n^{d-1} \left(h_n - R + \sqrt{R^2 - t_n^2} \right)^{\alpha+1} \\
&\geq \frac{C_f V_{d-1}}{\alpha+1} t_n^{d-1} \left(h_n - R + R - \frac{t_n^2}{2R} - \frac{8t_n^3}{9R^2} \right)^{\alpha+1} \text{ by Taylor} \\
&\geq \frac{C_f V_{d-1}}{\alpha+1} (\xi r_n - p_n)^{d-1} \left(\frac{\mu r_n^2}{2R} - q_n - R + R - \frac{\xi^2 r_n^2}{2R} - \frac{8\xi^3 r_n^3}{9R^2} \right)^{\alpha+1} \\
&\geq \frac{C_f V_{d-1}}{\alpha+1} (\xi r_n - p_n)^{d-1} \left(\frac{(\mu - \xi^2) r_n^2}{2R} - q_n - \frac{8\xi^3 r_n^3}{9R^2} \right)^{\alpha+1} \\
&\geq \frac{C_f V_{d-1} \xi^{d-1} (\mu - \xi^2)^{\alpha+1} r_n^{d+1+2\alpha}}{(\alpha+1)(2R)^{\alpha+1}} \left(1 - \frac{p_n}{\xi r_n} \right)^{d-1} \left(1 - \frac{2Rq_n}{(\mu - \xi^2) r_n^2} - \frac{16\xi^3 r_n}{9R(\mu - \xi^2)} \right)^{\alpha+1} \\
&\geq \frac{C_f V_{d-1} \xi^{d-1} (\mu - \xi^2)^{\alpha+1} r_n^{d+1+2\alpha}}{(\alpha+1)(2R)^{\alpha+1}} \left(1 - (d-1) \frac{p_n}{\xi r_n} \right) \left(1 - \frac{2(\alpha+1)Rq_n}{(\mu - \xi^2) r_n^2} - \frac{16(\alpha+1)\xi^3 r_n}{9R(\mu - \xi^2)} \right)^{\alpha+1} \\
&\geq \frac{C_f V_{d-1} \xi^{d-1} (\mu - \xi^2)^{\alpha+1} r_n^{d+1+2\alpha}}{(\alpha+1)(2R)^{\alpha+1}} \left(1 - \frac{2(\alpha+1)Rq_n}{(\mu - \xi^2) r_n^2} - (d-1) \frac{p_n}{\xi r_n} - \frac{16(\alpha+1)\xi^3 r_n}{9R(\mu - \xi^2)} \right)^{\alpha+1}.
\end{aligned}$$

Let's cover ∂S with $N(\partial S, \epsilon_n r_n^2)$ balls with radius $\epsilon_n r_n^2$. If there exists $x \in \partial S$ such that $\mathcal{C}^{u_x}(x, t_n, h_n) \cap \mathcal{X}_n = \emptyset$, then there exists a small ball $B(x_i, \epsilon_n r_n^2)$ containing x , such that $\|u_x - u_{x_i}\| \leq \epsilon_n r_n^2 / R$ by Proposition 2.2.1. Now given t_n, h_n , by Lemma 2.5.2 we have $\mathcal{C}^{u_{x_i}}(x_i, t'_n, h'_n) \subset \mathcal{C}^{u_x}(x, t_n, h_n)$ if

$$\begin{aligned}
(2.5.22) \quad t'_n &= t_n - \epsilon_x - 2(\sqrt{t_n^2 + h_n^2} + \epsilon_x) \epsilon_u > 0, \\
h'_n &= h_n - \epsilon_x - \sqrt{t_n^2 + h_n^2} \epsilon_u > 0,
\end{aligned}$$

where $\epsilon_x = \|x - x_i\| \leq \epsilon_n r_n^2$ and $\epsilon_u = \|u_x - u_{x_i}\| \leq \epsilon_n r_n^2 / R$.

The above equalities together with $\xi r_n - p_n \leq t_n \leq \xi r_n$ and $\frac{\mu r_n^2}{2R} - q_n \leq h_n \leq \frac{\mu r_n^2}{2R}$ yield

(2.5.23)

$$\begin{aligned}
\xi r_n - p_n - \epsilon_x - 2(\sqrt{t_n^2 + h_n^2} + \epsilon_x) \epsilon_u &\leq t'_n \leq \xi r_n - \epsilon_x - 2(\sqrt{t_n^2 + h_n^2} + \epsilon_x) \epsilon_u \leq \xi r_n, \\
(\mu r_n^2)/(2R) - q_n - \epsilon_x - \sqrt{t_n^2 + h_n^2} \epsilon_u &\leq h'_n \leq (\mu r_n^2)/(2R) - \epsilon_x - \sqrt{t_n^2 + h_n^2} \epsilon_u \leq (\mu r_n^2)/(2R).
\end{aligned}$$

Then when $\epsilon_n < 1$

$$\begin{aligned}
(2.5.24) \quad \epsilon_x + 2(\sqrt{t_n^2 + h_n^2} + \epsilon_x)\epsilon_u &\leq \epsilon_n r_n^2 + 2 \left(\sqrt{\xi^2 r_n^2 + \frac{\mu^2 r_n^4}{4R^2}} + \epsilon_n r_n^2 \right) \frac{\epsilon_n r_n^2}{R} \\
&= \epsilon_n r_n^2 \left(1 + \frac{2r_n}{R} \sqrt{\xi^2 + \frac{\mu^2 r_n^2}{4R^2}} + \frac{2\epsilon_n r_n^2}{R} \right) \\
&\text{when } r_n \leq R/2 \\
&\leq \epsilon_n r_n^2 \left(1 + \sqrt{\xi^2 + \frac{\mu^2}{16}} + R \right)
\end{aligned}$$

And

$$\begin{aligned}
(2.5.25) \quad \epsilon_x + \sqrt{t_n^2 + h_n^2}\epsilon_u &\leq \epsilon_n r_n^2 + \sqrt{\xi^2 r_n^2 + \frac{\mu^2 r_n^4}{4R^2}} \frac{\epsilon_n r_n^2}{R} \\
&\leq \epsilon_n r_n^2 \left(1 + \frac{r_n}{R} \sqrt{\xi^2 + \frac{\mu^2 r_n^2}{4R^2}} \right) \\
&\text{when } r_n \leq R/2 \\
&\leq \epsilon_n r_n^2 \left(1 + \frac{\sqrt{\xi^2 + \mu^2/16}}{2} \right)
\end{aligned}$$

Let's denote

$$\begin{aligned}
(2.5.26) \quad p'_n &:= p_n + \epsilon_n r_n^2 \left(1 + \sqrt{\xi^2 + \frac{\mu^2}{16}} + R \right), \\
q'_n &:= q_n + \epsilon_n r_n^2 \left(1 + \frac{\sqrt{\xi^2 + \mu^2/16}}{2} \right).
\end{aligned}$$

So $\xi r_n - p'_n \leq t'_n \leq \xi r_n$ and $\frac{\mu r_n^2}{2R} - q'_n \leq h'_n \leq \frac{\mu r_n^2}{2R}$. When $\xi r_n - p'_n > 0$ and $\frac{\mu r_n^2}{2R} - q'_n > 0$,

$$(2.5.27) \quad P_n := P(\exists x \in \partial S; \mathcal{C}^{u_x}(x, t_n, h_n) \cap \mathcal{X}_n = \emptyset) \leq P(\exists x_i; \mathcal{C}^{u_{x_i}}(x_i, t'_n, h'_n) = \emptyset).$$

So with the assumptions (ab),

(2.5.28)

$$\begin{aligned}
P_n &\leq N(\partial S, \epsilon_n r_n^2) \left(1 - \frac{C_f V_{d-1} \xi^{d-1} (\mu - \xi^2)^{\alpha+1} r_n^{d+1+2\alpha}}{(\alpha+1)(2R)^{\alpha+1}} \left(1 - \frac{2(\alpha+1)Rq'_n}{(\mu - \xi^2)r_n^2} - (d-1) \frac{p'_n}{\xi r_n} - \frac{16(\alpha+1)\xi^3 r_n}{9R(\mu - \xi^2)} \right) \right)^n \\
&\leq N(\partial S, \epsilon_n r_n^2) e^{-\frac{C_f V_{d-1} \xi^{d-1} (\mu - \xi^2)^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)(2R)^{\alpha+1}} \left(1 - \frac{2(\alpha+1)Rq'_n}{(\mu - \xi^2)r_n^2} - (d-1) \frac{p'_n}{\xi r_n} - \frac{16(\alpha+1)\xi^3 r_n}{9R(\mu - \xi^2)} \right)} \\
&= e^{\log N(\partial S, \epsilon_n r_n^2) - \frac{C_f V_{d-1} \xi^{d-1} (\mu - \xi^2)^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)(2R)^{\alpha+1}} (1 - M_{\alpha, d, \mu, \xi}(p_n, q_n, r_n, \epsilon_n))},
\end{aligned}$$

where

$$\begin{aligned}
M_{\alpha, d, \mu, \xi}(p_n, q_n, r_n, \epsilon_n) &= \frac{2(\alpha+1)R}{(\mu - \xi^2)} \frac{q_n}{r_n^2} + \frac{2(\alpha+1)R}{(\mu - \xi^2)} \left(1 + \frac{\sqrt{\xi^2 + \mu^2/16}}{2} \right) \epsilon_n + \frac{(d-1)}{\xi} \frac{p_n}{r_n} + \frac{(d-1)}{\xi} \left(1 + \sqrt{\xi^2 + \frac{\mu^2}{16}} + R \right) \epsilon_n r_n + \\
&\frac{16(\alpha+1)\xi^3}{9R(\mu - \xi^2)} r_n. \quad \square
\end{aligned}$$

2.5.2. A Global Concentration Bound. We first define a function

$$\begin{aligned}
&\Psi(S, \partial S, d, R, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) \\
&:= e^{\log N(S, \epsilon_{1,n}) - \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) n \left(\frac{3\rho_n}{2}\right)^{\alpha+d}} \left(1 - (\alpha+d) \frac{2\epsilon_{1,n}}{3\rho_n} - \min(1, (\alpha+d) \frac{3\rho_n/2 - \epsilon_{1,n}}{R}) \right) \\
(2.5.29) \quad &+ e^{\log N(\partial S, \epsilon_{2,n} r_n^2) - \frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)2^{d-1}}} \left(1 - M_{\alpha, d, \frac{1}{2}, 2bR} \left(\frac{99\rho_n^2}{2r_n}, 0, r_n, \epsilon_{2,n} \right) \right) \\
&+ e^{\log N(\partial S, \epsilon_{3,n} r_n^2) - \frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)4^{d-1}}} \left(1 - M_{\alpha, d, \frac{1}{4}, 2bR - \frac{3}{16}} \left(\frac{99\rho_n^2}{4r_n} + \frac{353r_n^2}{2048R}, \frac{r_n^3}{128R^2}, r_n, \epsilon_{3,n} \right) \right)
\end{aligned}$$

where $B(\cdot, \cdot)$ is the Beta function and $M_{\alpha, d, \mu, \xi}(p_n, q_n, r_n, \epsilon_n) = \frac{2(\alpha+1)R}{(\mu - \xi^2)} \frac{q_n}{r_n^2} + \frac{2(\alpha+1)R}{(\mu - \xi^2)} \left(1 + \frac{\sqrt{\xi^2 + \mu^2/16}}{2} \right) \epsilon_n + \frac{(d-1)}{\xi} \frac{p_n}{r_n} + \frac{(d-1)}{\xi} \left(1 + \sqrt{\xi^2 + \frac{\mu^2}{16}} + R \right) \epsilon_n r_n + \frac{16(\alpha+1)\xi^3}{9R(\mu - \xi^2)} r_n$ as defined in Equation (2.5.17).

Now let's state the global bound.

THEOREM 2.5.1. *Let $\mathcal{X}_n$ be an i.i.d. sample in $\mathbb{R}^d$ from a density function f that has support S . Suppose that*

- (a) S is a compact d -manifold,
- (b) S satisfies the rolling-ball condition with $R_{in} = R_{out} = R > 0$,
- (c) f is α -quickly decreasing with constant C_f ,

then for positive numbers $b, r_n, \rho_n, \epsilon_{i,n}$ for $i \in \{1, 2, 3\}$ (for their meanings, see the discussions below), when the following conditions are satisfied,

$$\begin{aligned}
(1) \quad & r_n < \frac{R}{2}, \\
(2) \quad & br_n^2 < \rho_n < \frac{r_n}{24}, \\
(3) \quad & b > \max \left(\frac{1}{8R}, \frac{2r_n^2}{R} + \frac{3r_n^3}{8R^2} \right), \\
(4) \quad & \epsilon_{1,n} < \frac{3\rho_n}{2}, \\
(5) \quad & \epsilon_{2,n} < \min \left(\frac{\frac{1}{2(d-1)} - \frac{99\rho_n^2}{r_n^2}}{r_n(2+\sqrt{1+b^2R^2}+2R)}, \frac{4bR}{R(4+\sqrt{1+b^2R^2})}, \frac{\frac{8bR-1}{4(\alpha+1)} - \frac{4r_n}{9R}}{R(4+\sqrt{1+b^2R^2})} \right), \\
(6) \quad & \epsilon_{3,n} < \min \left(\frac{\frac{1}{2(d-1)} - \frac{99\rho_n^2}{r_n^2} - \frac{353r_n}{512R}}{r_n(4+\sqrt{1+(2bR-\frac{3}{16})^2}+4R)}, \frac{8bR-\frac{3}{4}-\frac{r_n^2}{16R}}{R(8+\sqrt{1+(2bR-\frac{3}{16})^2})}, \frac{\frac{8bR-1}{2(\alpha+1)} - \frac{r_n}{9R} - \frac{r_n^2}{16R}}{R(8+\sqrt{1+(2bR-\frac{3}{16})^2})} \right),
\end{aligned}$$

we have

$$(2.5.30) \quad P \left((\hat{H}_{r_n} \Delta S) \subset \partial S \bigoplus br_n^2 \mathcal{B} \right) \geq 1 - \Psi(S, \partial S, d, R, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}).$$

where $\Psi(\dots)$ is defined in Equation (2.5.29).

DISCUSSION. r_n is the radius of the local balls that are used to build the r -ball LoCoH estimator. We want r_n to be small relative to the rolling ball radius R so that a local ball intersect with S is almost “convex”. So we are able to estimate the support locally well using convex hulls.

$\forall x \in S$, let $d_n(x) = d(x, \partial S)$ be the distance of x to the boundary. In the proof, we discuss the cases when $d_n(x) > 3\rho_n$ and when $d_n(x) < 3\rho_n$ separately. In the cases when $d_n(x) < 3\rho_n$, we make the argument by building a tangent cylinder with base radius t_n such that $0 \leq gr_n - t_n \leq h \frac{\rho_n^2}{r_n}$ for some constant $g, h > 0$. So ρ_n controls the deviation of t_n from gr_n . As we discussed in Lemma 2.5.3, $\frac{\rho_n^2}{r_n}$ needs to be small relatively to r_n .

The symmetric difference between the r -ball LoCoH estimator $\hat{H}_{r_n}$ and S is bounded by a constant whose value is determined by the shape of ∂S times br_n . We call the constant b the error coefficient. And it is bounded from below by $\max \left(\frac{1}{8R}, \frac{2r_n^2}{R} + \frac{3r_n^3}{8R^2} \right)$. The bigger rolling ball radius R is, the smaller value b can take. So the choice of b is affected by the rolling ball parameter R , which measures the smoothness of the boundary. The value of b must be larger when the value of R is smaller, which means the boundary is less smooth, which also implies the estimation is harder.

$\epsilon_{1,n}$ is the radius of balls we use to cover S (see Lemma 2.5.1 and the discussions). $\epsilon_{2,n}r_n^2$ and $\epsilon_{3,n}r_n^2$ are the radii of balls we use to cover ∂S (see Lemma 2.5.3 and the discussions). The conditions (5) and (6) can be understood better if we look at the asymptotic forms in the Inequality (2.5.52). The values of $\epsilon_{2,n}$ and $\epsilon_{3,n}$ need be chosen small if (1) the error coefficient b is small; (2) α is large (so the density close to the boundary can be small). Note that the reason why R appears in the inequality is because that in Lemma 2.5.3, we use some rough estimates.

We can see that in order to make the right-hand side of the inequality small, $C_f^{\frac{1}{\alpha+1}}b$ must be large. When the value of C_f is small, the coefficient b needs to be large enough for the product $C_f^{\frac{1}{\alpha+1}}b$ to be large. C_f measures the speed of the density function f vanishes close to the boundary. The smaller value C_f is, the faster the density can vanish. So the fewer points close to the boundary of S in the data cloud $\mathcal{X}_n$. This makes the estimation close to the boundary harder. So the error coefficient b must be larger.

The asymptotic behavior and high dimension performance will be discussed in Corollaries 2.5.1 and 2.5.2.

□

PROOF. The structure of the proof is as follows. First we decompose the complement of the event $E = \{(\hat{H}_{r_n}\Delta S) \subset \partial S \oplus br_n^2\mathcal{B}\}$ into sub-events E_1, E_2, E_3 and E_4 . Then we bound the probability content of each sub-event from above. $P(E)$ is thus bounded from below by the Inequality (2.5.36).

(i) Decomposition of E^c :

Let $d_n(x) = d(x, \partial S)$ and $x^* \in \partial S$ such that $d(x, x^*) = d_n(x)$. E can be rewritten as $\{\forall x \in (\hat{H}_{r_n}\Delta S), d_n \leq br_n^2\}$.

$$\begin{aligned}
E^c &= \{\exists x \in ((\hat{H}_{r_n} \cap S^c) \cup (\hat{H}_{r_n}^c \cap S)) \text{ such that } d_n(x) > br_n^2\} \\
(2.5.31) \quad &= \{\exists x \in \hat{H}_{r_n} \cap S^c \text{ such that } d_n(x) > br_n^2\} \cup \{\exists x \in \hat{H}_{r_n}^c \cap S \text{ such that } d_n(x) > br_n^2\} \\
&= E_1 \cup A_1.
\end{aligned}$$

By condition (2) $br_n^2 < \rho_n < r_n$, we can write $A_1 = E_2 \cup A_2$, where

$$(2.5.32) \quad E_2 = \{\exists x \in \hat{H}_{r_n}^c \cap S, \text{ such that } d_n(x) > br_n^2 \text{ and } d_n(x) > 3\rho_n\},$$

and

$$(2.5.33) \quad A_2 = \{\exists x \in \hat{H}_{r_n}^c \cap S, \text{ such that } d_n(x) > br_n^2 \text{ and } d_n(x) \leq 3\rho_n\}.$$

We further split up A_2 :

By Proposition 2.2.1, there is a unique inward pointing vector u_{x^*} . Using this, we write $A_2 = E_3 \cup E_4$, where

$$(2.5.34) \quad E_3 = \{\exists x \in \hat{H}_{r_n}^c \cap S, \forall X \in B(x, r_n/2) \cap \mathcal{X}_n, \langle u_{x^*}, X - x \rangle \geq 0 \text{ such that } br_n^2 < d_n(x) \leq 3\rho_n\},$$

and

$$(2.5.35) \quad E_4 = \{\exists x \in \hat{H}_{r_n}^c \cap S, \exists X \in B(x, r_n/2) \cap \mathcal{X}_n, \langle u_{x^*}, X - x \rangle < 0 \text{ such that } br_n^2 < d_n(x) \leq 3\rho_n\}.$$

We now have

$$(2.5.36) \quad \begin{aligned} P\left((\hat{H}_{r_n} \cap S^c) \cup (\hat{H}_{r_n}^c \cap S)\right) &\subset \partial S \bigoplus br_n^2 \mathcal{B} = P((E_1 \cup E_2 \cup E_3 \cup E_4)^c) \\ &= 1 - P(E_1 \cup E_2 \cup E_3 \cup E_4) \\ &\geq 1 - P(E_1) - P(E_2) - P(E_3) - P(E_4) \end{aligned}$$

Since $\hat{H}_{r_n} \Delta S = ((\hat{H}_{r_n} \cap S^c) \cup (\hat{H}_{r_n}^c \cap S))$, it remains to bound the probabilities $P(E_1), \dots, P(E_4)$ from above.

(ii) Bounding $P(E_1), \dots, P(E_4)$ from above:

$(\hat{H}_{r_n} \cap S^c)$ is the over-filled region by the estimator $\hat{H}_{r_n}$ and $(\hat{H}_{r_n}^c \cap S)$ is the under-filled region. First we control the over-filled region $(\hat{H}_{r_n} \cap S^c)$ by studying the event E_1 :

Event $E_1 = \{\exists x \in \hat{H}_{r_n} \cap S^c \text{ such that } d_n(x) > br_n^2\}$

Suppose $x \in \hat{H}_{r_n} \cap S^c$. By the definition of $\hat{H}_{r_n}$, $\exists X_i \in \mathcal{X}_n$ such that $x \in \mathcal{H}(B(X_i, r_n) \cap \mathcal{X}_n) \subset \mathcal{H}(B(x, 2r_n) \cap \mathcal{X}_n)$. Then by the rolling-ball condition, there exists $O_x^- \in \mathbb{R}^d$ such that $x \in B(O_x^-, R_{out})$. Therefore, $x \in B(x, 2r_n) \cap B(O_x^-, R_{out})$.

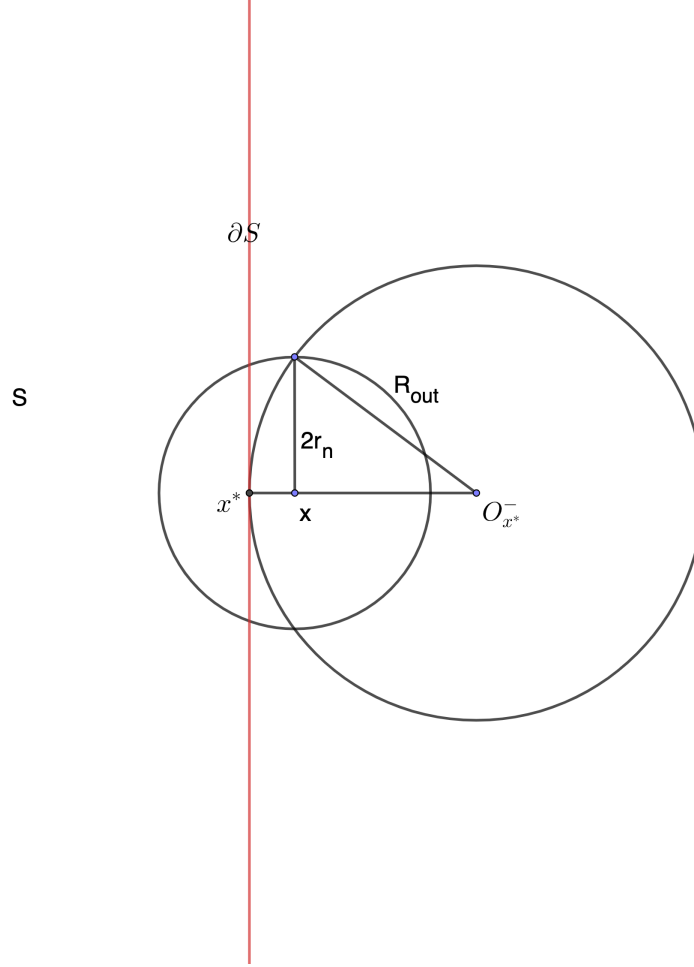


FIGURE 2.4. $d(x, \partial S) = R_{out} - \sqrt{R_{out}^2 - 4r_n^2}$.

By assumption (1) $r_n \leq R/2$, we have (see Figure 2.4) $d(x, \partial S) \leq R - \sqrt{R^2 - 4r_n^2} \leq 2r_n^2/R + 16r_n^3/9R^2$. Therefore, when $b > 2r_n^2/R + 16r_n^3/9R^2$,

$$(2.5.37) \quad P(E_1) = 0.$$

Note that this part is fully deterministic.

To ensure the under-filled region $\hat{H}_{r_n}^c \cap S$ is subset of $\partial S \oplus br_n^2 \mathcal{B}$, the three separate cases E_2, E_3 and E_4 are discussed. $\hat{H}_{r_n}^c \cap S \not\subset \partial S \oplus br_n^2 B$ implies $\exists x \in \hat{H}_{r_n}^c \cap S$ such that $d_n(x) = d(x, \partial S) > br_n^2$. E_2 is the case when $d_n > 3\rho_n$ and $E_3 \& E_4$ are the cases when $br_n^2 < d_n(x) \leq 3\rho_n$.

$$\text{Event } E_2 = \{\exists x \in \hat{H}_{r_n}^c \cap S, \text{ such that } d_n(x) > 3\rho_n \text{ and } d_n(x) \geq br_n^2\}$$

Suppose $x \in \hat{H}_{r_n}^c \cap S$ and $d_n(x) > 3\rho_n$. First if $x \in \hat{H}_{r_n}^c$, then $x \notin \mathcal{H}(B(x, r_n/2) \cap \mathcal{X}_n)$. This is because if otherwise, $\exists X_i \in B(x, r_n/2)$. Then $B(x, r_n/2) \subset B(X_i, r_n)$ implies $x \in \mathcal{H}(B(X_i, r_n) \cap \mathcal{X}_n) \subset \hat{H}_{r_n}$, contradiction. So we can find a vector u such that $\forall X \in B(x, r_n/2) \cap \mathcal{X}_n, \langle u, X - x \rangle \geq 0$ (as shown in Figure 2.5). Let $y = x - \frac{3}{2}\rho_n u$. Then $y \in S$. When n is large enough such that $\frac{3\rho_n}{2} \leq \frac{r_n}{4}$, $B(y, \frac{3\rho_n}{2}) \cap \mathcal{X}_n = \emptyset$. The reason can be seen from Figure 2.5.

From the definition of the Hausdorff distance,

$$(2.5.38) \quad \exists z \in S \text{ such that } B(z, \frac{3\rho_n}{2}) \cap \mathcal{X}_n = \emptyset \implies d_H(\mathcal{X}_n, S) \geq \frac{3\rho_n}{2}$$

So, event E_2 implies $d_H(\mathcal{X}_n, S) \geq \frac{3\rho_n}{2}$ when $\rho_n < \frac{r_n}{6}$. By Lemma 2.5.1,

$$(2.5.39) \quad \begin{aligned} P(E_2) &\leq P\left(d_H(\mathcal{X}_n, S) \geq \frac{3\rho_n}{2}\right) \\ &\leq e^{\log N(S, \epsilon_{1,n}) - \frac{C_f V_{d-1} B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) n \left(\frac{3\rho_n}{2}\right)^{\alpha+d}} \left(1 - (\alpha+d) \frac{2\epsilon_{1,n}}{3\rho_n} - \min(1, (\alpha+d) \frac{3\rho_n/2 - \epsilon_{1,n}}{R})\right) \end{aligned}$$

where $\epsilon_{1,n} < \frac{3\rho}{2} < R$.

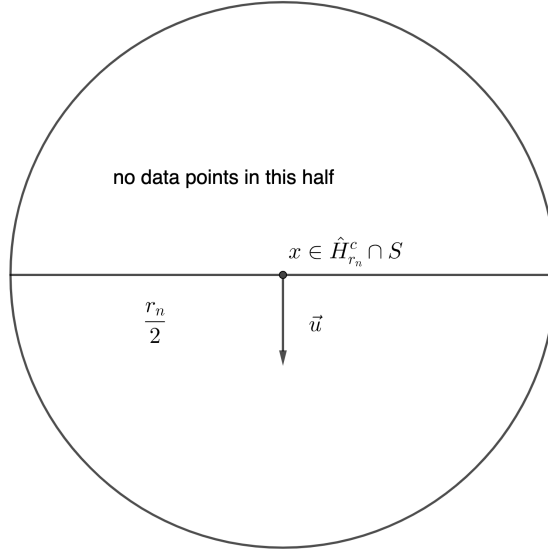


FIGURE 2.5. If $x \in \hat{H}_{r_n}^c$, then $\exists u \in \mathbb{R}^d$ such that $\forall X \in B(x, r_n/2) \cap \mathcal{X}_n$, $\langle u, X - x \rangle \geq 0$.

By the previous argument, if $x \in \hat{H}_{r_n}^c$, then we can find a vector u such that $\forall X \in B(x, r_n/2) \cap \mathcal{X}_n$, $\langle u, X - x \rangle \geq 0$ (as shown in Figure 2.5). E_3 is the case when u_{x^*} is such a vector and E_4 is the case when u_{x^*} is not. The difference is whether we can build a tangent cylinder centered at x^* .

Event $E_3 = \{\exists x \in \hat{H}_{r_n}^c \cap S, \forall X \in B(x, r_n/2) \cap \mathcal{X}_n, \langle u_{x^*}, X - x \rangle \geq 0 \text{ such that } br_n^2 < d_n(x) \leq 3\rho_n\}$

The idea is given event E_3 , to find a tangent cylinder with a centroid on ∂S , which does not intersect with $\mathcal{X}_n$ (see Figure 2.6). Probability of event E_3 is then bounded from above by the probability of such an event. Because $\forall X \in B(x, r_n/2) \cap \mathcal{X}_n, \langle u_{x^*}, X - x \rangle \geq 0$, we can construct a cylinder centered at x^* .

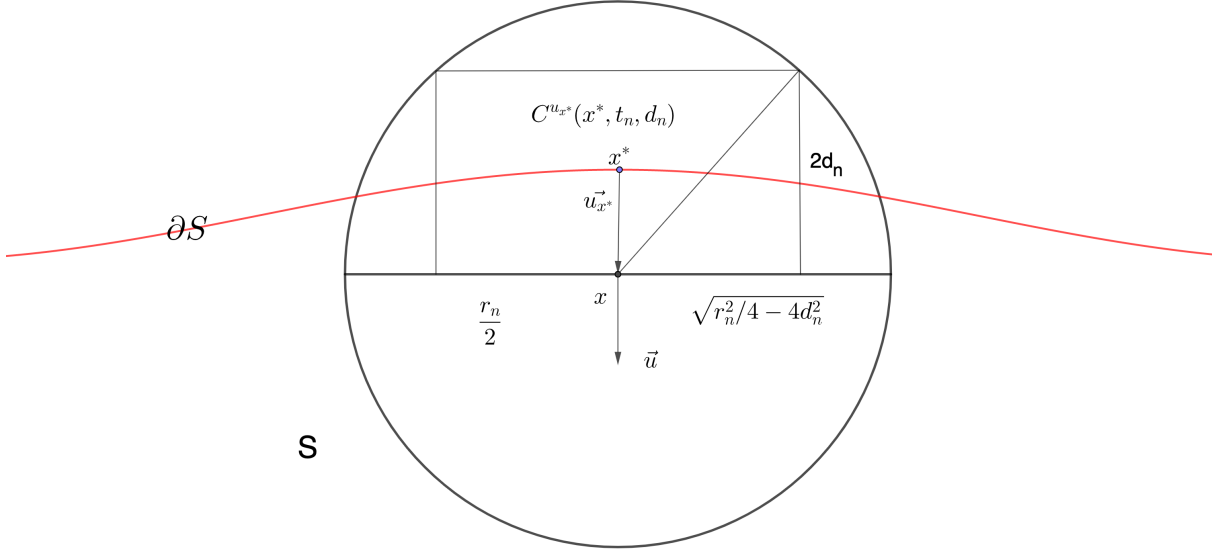


FIGURE 2.6. $\forall X \in B(x, r_n/2) \cap \mathcal{X}_n, \langle u_{x^*}, X - x \rangle \geq 0$.

Let

$$(2.5.40) \quad t_n = \sqrt{\frac{r_n^2}{4} - 4d_n^2}.$$

Then $\mathcal{C}^{u_{x^*}}(x^*, t_n, d_n)$ is the desired cylinder. Note that $\mathcal{C}^{u_{x^*}}(x^*, t_n, br_n^2) \subset \mathcal{C}^{u_{x^*}}(x^*, t_n, d_n)$. When $\rho_n \leq \frac{r_n}{24}$, this together with $br_n^2 < d_n(x) \leq 3\rho_n$ gives

$$(2.5.41) \quad 0 \leq \frac{r_n}{2} - t_n \leq \frac{99\rho_n^2}{2r_n}$$

So let $\xi = \frac{1}{2}$, $p_n = \frac{99\rho_n^2}{2r_n}$, $\mu = 2bR$ and $q_n = 0$. When $p_n < \frac{r_n}{2}$, we have $\xi r_n - p_n \leq t_n \leq \xi r_n$ and $\frac{\mu r_n^2}{2R} - q_n \leq br_n^2 \leq \frac{\mu r_n^2}{2R}$. By Lemma 2.5.3, for some $\epsilon_{2,n} > 0$ when the following condition

$$(2.5.42) \quad \epsilon_{2,n} < \min \left(\frac{\frac{1}{2(d-1)} - \frac{99\rho_n^2}{r_n^2}}{r_n(2 + \sqrt{1 + b^2 R^2} + 2R)}, \frac{4bR}{R(4 + \sqrt{1 + b^2 R^2})}, \frac{\frac{8bR-1}{4(\alpha+1)} - \frac{4r_n}{9R}}{R(4 + \sqrt{1 + b^2 R^2})} \right)$$

is satisfied, we have

$$(2.5.43) \quad \begin{aligned} & P(\exists x \in \partial S; \mathcal{C}(x, t_n, h_n) \cap \mathcal{X}_n = \emptyset) \\ & \leq e^{\log N(\partial S, \epsilon_{2,n}, r_n^2) - \frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)2^{d-1}}} \left(1 - M_{\alpha, d, \frac{1}{2}, 2bR} \left(\frac{99\rho_n^2}{2r_n}, 0, r_n, \epsilon_{2,n} \right) \right), \end{aligned}$$

where $M_{\alpha, d, \mu, \xi}(p_n, q_n, r_n, \epsilon_n)$ is given by Equation (2.5.17).

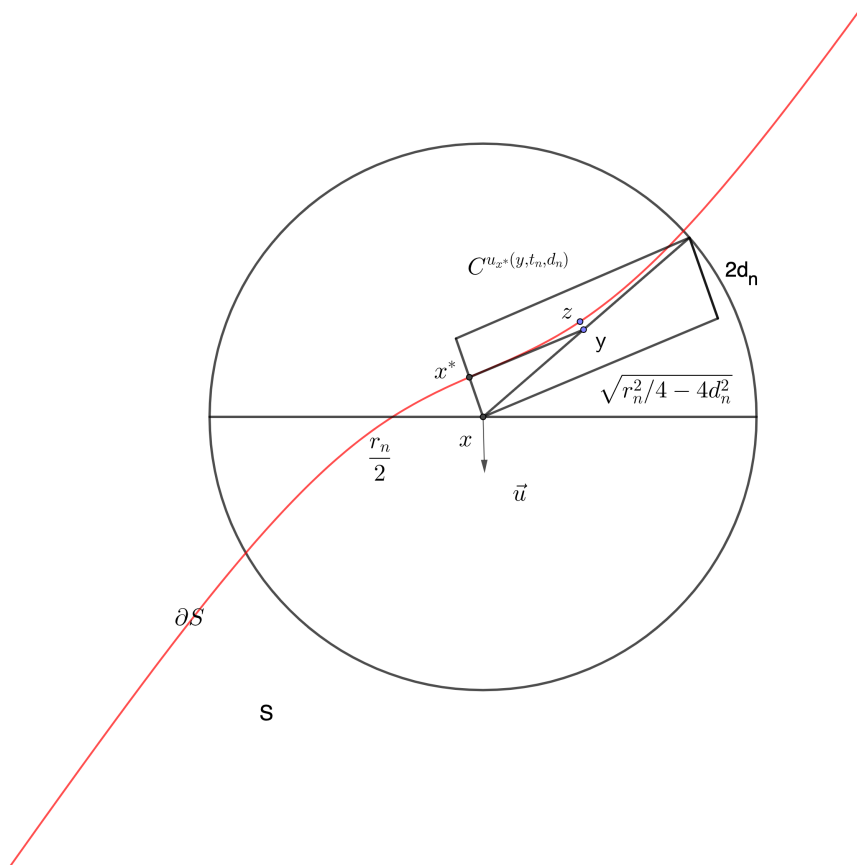


FIGURE 2.7. $w = \langle u, u_{x^*} \rangle u_{x^*} - u$, $v = w / \|w\|$. $y = x^* + 1/2\sqrt{r_n^2/4 - 4d_n^2}v$.

Event $E_4 = \{\exists x \in \hat{H}_{r_n}^c \cap S, \exists X \in B(x, r_n/2) \cap \mathcal{X}_n, \langle u_{x^*}, X - x \rangle < 0 \text{ such that } br_n^2 < d_n(x) \leq 3\rho_n\}$

The idea to control $P(E_4)$ is similar to $P(E_3)$. Given event E_4 , a tangent cylinder can be constructed with a centroid on ∂S , which does not intersect with $\mathcal{X}_n$. Because that u_{x^*} does not satisfy the condition that $\forall X \in B(x, r_n/2) \cap \mathcal{X}_n, \langle u_{x^*}, X - x \rangle \geq 0$, such a cylinder is not centered at x^* as for the event E_3 . Rather, a cylinder whose centroid y is not on ∂S (the direction vector is u_{x^*}) is first constructed such that it does not intersect with $\mathcal{X}_n$ (see Figure 2.7), then Lemma 2.5.2 is used to find another point $z \in \partial S$ (since $d(y, z) \leq \alpha_n$, which can be calculated explicitly by Figure 2.8) then to build a tangent cylinder in the direction of u_z such that it is contained in the first cylinder centered at y .

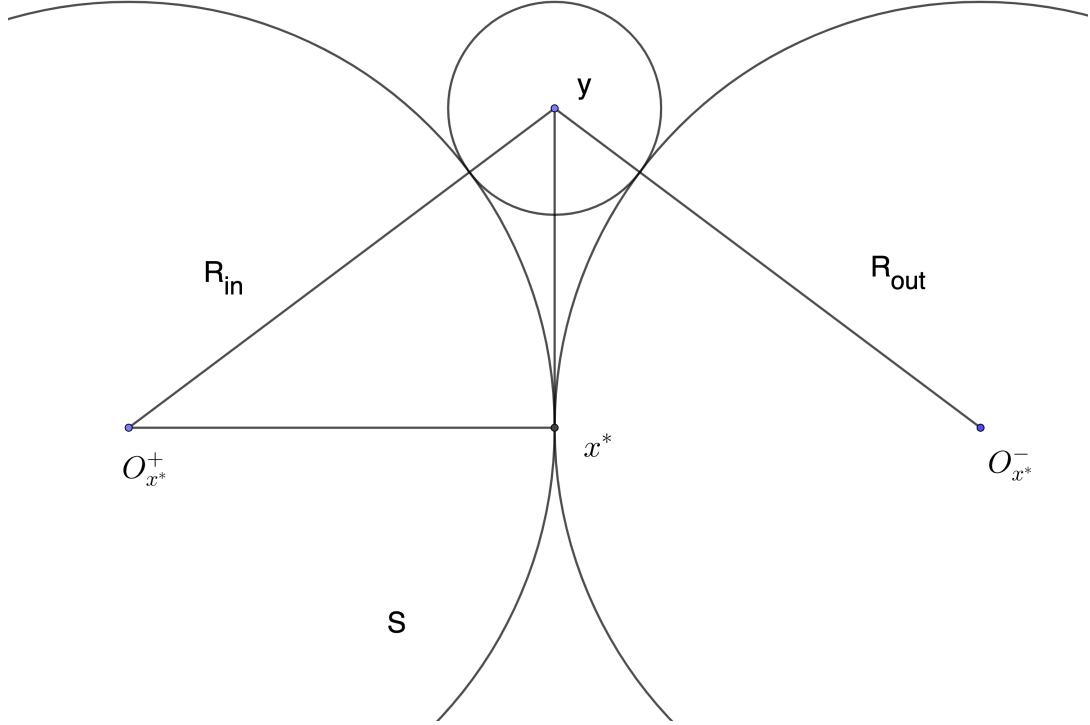


FIGURE 2.8. The distance between y and z is within α_n .

Let $t_n = \frac{1}{2}\sqrt{\frac{r_n^2}{4} - 4d_n^2}$. When $br_n^2 < d_n(x) \leq 3\rho_n$ and $\rho_n \leq \frac{r_n}{24}$,

$$(2.5.44) \quad 0 \leq \frac{r_n}{4} - t_n \leq \frac{99\rho_n^2}{4r_n}$$

The cylinder $\mathcal{C}^{u_{x^*}}(y, t_n, d_n)$ is the first cylinder we want to construct (Figure 2.7). Now to construct a smaller tangent cylinder, we consider a point $z \in \partial S$, which is the projection of y onto ∂S . Then $d(z, y) \leq \alpha_n$ and $\|u_z - u_{x^*}\| \leq \frac{1}{R}(\frac{r_n}{4} + \alpha_n)$, where $\alpha_n = -R + \sqrt{R^2 + \frac{r_n^2}{16}} \leq \frac{r_n^2}{32R}$. The reason is illustrated in Figure 2.8.

To apply Lemma 2.5.2, write

$$\begin{aligned}
e_1 &= \frac{r_n^2}{32R} + 2 \left(\sqrt{t_n^2 + b^2 r_n^4} + \frac{r_n^2}{32R} \right) \frac{1}{R} \left(\alpha_n + \frac{r_n}{4} \right) \\
&= \frac{r_n^2}{32R} + 2 \left(\sqrt{\left(\frac{1}{2} \sqrt{\frac{r_n^2}{4} - 4d_n^2} \right)^2 + b^2 r_n^4} + \frac{r_n^2}{32R} \right) \frac{1}{R} \left(\alpha_n + \frac{r_n}{4} \right) \\
&= \frac{r_n^2}{32R} + 2 \left(\sqrt{\frac{r_n^2}{16} - d_n^2 + b^2 r_n^4} + \frac{r_n^2}{32R} \right) \frac{1}{R} \left(\alpha_n + \frac{r_n}{4} \right) \\
(2.5.45) \quad &\leq \frac{r_n^2}{32R} + 2 \left(\sqrt{\frac{r_n^2}{16} - b^2 r_n^4 + b^2 r_n^4} + \frac{r_n^2}{32R} \right) \frac{1}{R} \left(\alpha_n + \frac{r_n}{4} \right) \\
&\leq \frac{r_n^2}{32R} + 2 \left(\frac{r_n}{4} + \frac{r_n^2}{32R} \right) \frac{1}{R} \left(\frac{r_n^2}{32R} + \frac{r_n}{4} \right) \\
&= \frac{5r_n^2}{32R} + \frac{r_n^3}{32R^2} + \frac{r_n^4}{512R^3} \\
&\leq \frac{353r_n^2}{2048R} \quad \text{since } r_n < R/2
\end{aligned}$$

and

$$\begin{aligned}
e_2 &= \frac{r_n^2}{32R} + \sqrt{t_n^2 + b^2 r_n^4} \frac{1}{R} \left(\alpha_n + \frac{r_n}{4} \right) \\
(2.5.46) \quad &\leq \frac{r_n^2}{32R} + \frac{r_n}{4R} \left(\alpha_n + \frac{r_n}{4} \right) \\
&= \frac{r_n^2}{32R} + \frac{r_n}{4R} \left(\frac{r_n^2}{32R} + \frac{r_n}{4} \right) \\
&= \frac{3r_n^2}{32R} + \frac{r_n^3}{128R^2}.
\end{aligned}$$

Then $\mathcal{C}^{u_z}(z, t_n - e_1, d_n - e_2) \subset \mathcal{C}^{u_{x^*}}(y, t_n, d_n)$ when $t_n - e_1 > 0$ and $d_n - e_2 > 0$. Let $\xi = \frac{1}{4}$, $p_n = \frac{99\rho_n^2}{4r_n} + \frac{353r_n^2}{2048R}$, $\mu = 2bR - \frac{3}{16} > 0$ and $q_n = \frac{r_n^3}{128R^2}$.

By Lemma 2.5.3, for some $\epsilon_{3,n} > 0$ when the following condition

$$(2.5.47) \quad \epsilon_{3,n} < \min \left(\frac{\frac{1}{2(d-1)} - \frac{99\rho_n^2}{r_n^2} - \frac{353r_n}{512R}}{r_n(4 + \sqrt{1 + (2bR - \frac{3}{16})^2} + 4R)}, \frac{8bR - \frac{3}{4} - \frac{r_n^2}{16R}}{R(8 + \sqrt{1 + (2bR - \frac{3}{16})^2})}, \frac{\frac{8bR-1}{2(\alpha+1)} - \frac{r_n}{9R} - \frac{r_n^2}{16R}}{R(8 + \sqrt{1 + (2bR - \frac{3}{16})^2})} \right)$$

is satisfied, we have

$$(2.5.48) \quad \begin{aligned} & P(\exists x \in \partial S; \mathcal{C}(x, t_n, h_n) \cap \mathcal{X}_n = \emptyset) \\ & \leq e^{\log N(\partial S, \epsilon_{3,n} r_n^2) - \frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)4^{d-1}}} \left(1 - M_{\alpha, d, \frac{1}{4}, 2bR - \frac{3}{16}} \left(\frac{99\rho_n^2}{4r_n} + \frac{353r_n^2}{2048R}, \frac{r_n^3}{128R^2}, r_n, \epsilon_{3,n} \right) \right), \end{aligned}$$

where $M_{\alpha, d, \mu, \xi}(p_n, q_n, r_n, \epsilon_n)$ is given by Equation (2.5.17).

The proof is done by plugging Equality (2.5.37), and Inequalities (2.5.39), (2.5.43), (2.5.48) in the Inequality (2.5.36). □

Now we state two corollaries that can be derived directly from the above theorem, which are used to study the consistency and high dimension behavior of the r -ball LoCoH estimator.

2.5.2.1. Consistency. The first corollary is obtained by plugging in the bounds for the covering numbers $N(S, \epsilon)$ and $N(\partial S, \epsilon)$ from Proposition 2.2.2. Let's first define the function Ψ^* . Let

$$(2.5.49) \quad \begin{aligned} & \Psi^*(S, \partial S, d, R, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) \\ & := \lambda_S e^{-d \log \epsilon_{1,n} - \frac{C_f V_{d-1} B \left(\frac{\alpha+1}{2}, \frac{d+1}{2} \right) n \left(\frac{3\rho_n}{2} \right)^{\alpha+d}} \left(1 - (\alpha+d) \frac{2\epsilon_{1,n}}{3\rho_n} - \min(1, (\alpha+d) \frac{3\rho_n/2 - \epsilon_{1,n}}{R}) \right) \\ & \quad + \lambda_{\partial S} e^{(1-d) \log(\epsilon_{2,n} r_n^2) - \frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)2^{d-1}}} \left(1 - M_{\alpha, d, \frac{1}{2}, 2bR} \left(\frac{99\rho_n^2}{2r_n}, 0, r_n, \epsilon_{2,n} \right) \right) \\ & \quad + \lambda_{\partial S} e^{(1-d) \log(\epsilon_{3,n} r_n^2) - \frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)4^{d-1}}} \left(1 - M_{\alpha, d, \frac{1}{4}, 2bR - \frac{3}{16}} \left(\frac{99\rho_n^2}{4r_n} + \frac{353r_n^2}{2048R}, \frac{r_n^3}{128R^2}, r_n, \epsilon_{3,n} \right) \right), \end{aligned}$$

where λ_S and $\lambda_{\partial S}$ are some positive numbers depending on S and ∂S , respectively. And $M_{\alpha,d,\mu,\xi}(p_n, q_n, r_n, \epsilon_n)$ is defined in Equation (2.5.17).

COROLLARY 2.5.1. *With the same assumptions as in Theorem 2.5.1, we have*

$$(2.5.50) \quad P\left((\hat{H}_{r_n}\Delta S) \subset \partial S \bigoplus br_n^2\mathcal{B}\right) \geq 1 - \Psi^*(S, \partial S, d, R, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}),$$

where $\Psi^*(\dots)$ is defined in Equation (2.5.49).

PROOF. By Proposition 2.2.2, the assumption (a) in Theorem 2.5.1 implies that there are positive numbers $\lambda_S, \lambda_{\partial S}$ such that $\forall \epsilon > 0$ the covering numbers

$$(2.5.51) \quad \begin{aligned} N(S, \epsilon) &\leq \lambda_S \epsilon^{-d}, \\ N(\partial S, \epsilon) &\leq \lambda_{\partial S} \epsilon^{1-d}. \end{aligned}$$

Plugging them in the Inequality (2.5.30) finishes the proof. □

Let $r_n, \rho_n, \epsilon_{1,n}$ be sequences of positive numbers that converges to 0 as $n \rightarrow \infty$. And $r_n^2 \ll \rho_n \ll r_n, \epsilon_{1,n} \ll \rho_n$. Let $\epsilon_{2,n}, \epsilon_{3,n} > 0$ be such that

$$(2.5.52) \quad \begin{aligned} \limsup_{n \rightarrow \infty} \epsilon_{2,n} &< \min\left(\frac{4bR}{R(4 + \sqrt{1 + b^2 R^2})}, \frac{\frac{8bR-1}{4(\alpha+1)}}{R(4 + \sqrt{1 + b^2 R^2})}\right), \\ \limsup_{n \rightarrow \infty} \epsilon_{3,n} &< \min\left(\frac{8bR - \frac{3}{4}}{R(8 + \sqrt{1 + (2bR - \frac{3}{16})^2})}, \frac{\frac{8bR-1}{2(\alpha+1)}}{R(8 + \sqrt{1 + (2bR - \frac{3}{16})^2})}\right). \end{aligned}$$

If there exists a constant $b > \frac{1}{8R}$ such that

$$\begin{aligned} (1) \quad &-d \log \epsilon_{1,n} - \frac{C_f V_{d-1} B(\frac{\alpha+1}{2}, \frac{d+1}{2}) n(\frac{3\rho_n}{2})^{\alpha+d}}{2} \rightarrow -\infty, \\ (2) \quad &(1-d) \log(\epsilon_{2,n} r_n^2) - \frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{2(\alpha+1)2^{d-1}} \rightarrow -\infty, \\ (3) \quad &(1-d) \log(\epsilon_{3,n} r_n^2) - \frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{2(\alpha+1)4^{d-1}} \rightarrow -\infty, \end{aligned}$$

then

$$(2.5.53) \quad P\left((\hat{H}_{r_n}\Delta S) \subset \partial S \bigoplus br_n^2\mathcal{B}\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

Note that $r_n \asymp \left(\frac{\log n}{n}\right)^{\frac{1}{d+1+2\alpha}}$, $\rho_n \asymp \left(\frac{\log n}{n}\right)^{\frac{1}{d+\alpha}}$, $\epsilon_{1,n} \asymp \left(\frac{1}{n}\right)^{\frac{1}{d+\alpha}}$, $\epsilon_{2,n} \asymp \epsilon_{3,n} \asymp (\log n)^{-2}$ is a valid choice as shown in Aaron and Bodart (2016).

2.5.2.2. High Dimension Performance. The second corollary is obtained by plugging in the bounds of the covering numbers $N(S, \epsilon)$ and $N(\partial S, \epsilon)$ from Proposition 2.2.3. Let's first define the function Ψ^{**} . Let

(2.5.54)

$$\begin{aligned} & \Psi^{**}(S, \partial S, d, R, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) \\ &:= \left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon_{1,n}} + \frac{1}{R} \right)^d e^{-\frac{C_f V_{d-1} B \left(\frac{\alpha+1}{2}, \frac{d+1}{2} \right) n \left(\frac{3\rho_n}{2} \right)^{\alpha+d}} \left(1 - (\alpha+d) \frac{2\epsilon_{1,n}}{3\rho_n} - \min(1, (\alpha+d) \frac{3\rho_n/2 - \epsilon_{1,n}}{R}) \right)} \\ &+ 2 \left[\left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon_{2,n} r_n^2} + \frac{1}{R} \right)^d - \frac{2^d \text{Vol}_d(S) \Gamma(d/2 + 1)}{\pi^{d/2} \epsilon_{2,n}^d r_n^{2d}} \right] \\ &\quad e^{-\frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1) 2^{d-1}}} \left(1 - M_{\alpha, d, \frac{1}{2}, 2bR} \left(\frac{99\rho_n^2}{2r_n}, 0, r_n, \epsilon_{2,n} \right) \right) \\ &+ 2 \left[\left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon_{3,n} r_n^2} + \frac{1}{R} \right)^d - \frac{2^d \text{Vol}_d(S) \Gamma(d/2 + 1)}{\pi^{d/2} \epsilon_{3,n}^d r_n^{2d}} \right] \\ &\quad e^{-\frac{C_f V_{d-1} (b-1/(8R))^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1) 4^{d-1}}} \left(1 - M_{\alpha, d, \frac{1}{4}, 2bR - \frac{3}{16}} \left(\frac{99\rho_n^2}{4r_n} + \frac{353r_n^2}{2048R}, \frac{r_n^3}{128R^2}, r_n, \epsilon_{3,n} \right) \right), \end{aligned}$$

where $M_{\alpha, d, \mu, \xi}(p_n, q_n, r_n, \epsilon_n)$ is defined in Equation (2.5.17).

COROLLARY 2.5.2. *With the same assumptions as in Theorem 2.5.1, we have*

$$(2.5.55) \quad P\left((\hat{H}_{r_n} \Delta S) \subset \partial S \bigoplus b r_n^2 \mathcal{B}\right) \geq 1 - \Psi^{**}(S, \partial S, d, R, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}),$$

where $\Psi^{**}(\dots)$ is defined in Equation (2.5.54).

PROOF. By Proposition 2.2.3, the assumption (b) in Theorem 2.5.1 implies that $\forall \epsilon > 0$ the covering numbers

$$(2.5.56) \quad N(S, \epsilon) \leq \left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon} + \frac{1}{R} \right)^d,$$

and

$$(2.5.57) \quad N(\partial S, \epsilon) \leq 2 \left[\left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon} + \frac{1}{R} \right)^d - \frac{2^d \text{Vol}_d(S) \Gamma(d/2 + 1)}{\pi^{d/2} \epsilon^d} \right].$$

Plugging them in the Inequality (2.5.30) finishes the proof. $\square$

Now let's analyze the performance of the r -ball LoCoH estimator $\hat{H}_{r_n}$ in high dimensions. In order to quantify it precisely, we consider the special case when S is a d -dimensional ball of volume 1. Then the boundary of S has $d - 1$ dimension volume $dV_d^{1/d}$. The rolling ball radius $R = V_d^{-1/d}$. The covering number bounds for $N(S, \epsilon)$ and $N(\partial S, \epsilon)$ from Proposition 2.2.3 can be further simplified as

$$(2.5.58) \quad \begin{aligned} N(S, \epsilon) &\leq \left(\frac{2R}{\epsilon} + 1 \right)^d, \\ N(\partial S, \epsilon) &\leq 2d \left(\frac{2R}{\epsilon} + 1 \right)^{d-1}. \end{aligned}$$

Let's define

$$(2.5.59) \quad \begin{aligned} &\psi(d, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) \\ &:= e^{d \log \left(\frac{2V_d^{-1/d}}{\epsilon_{1,n}} + 1 \right) - \frac{C_f V_{d-1} B \left(\frac{\alpha+1}{2}, \frac{d+1}{2} \right) n \left(\frac{3\rho_n}{2} \right)^{\alpha+d}}{2} \left(1 - (\alpha+d) \frac{2\epsilon_{1,n}}{3\rho_n} - \min(1, (\alpha+d) \frac{3\rho_n/2 - \epsilon_{1,n}}{V_d^{-1/d}}) \right)} \\ &+ e^{(d-1) \log 2d + (d-1) \log \left(\frac{2V_d^{-1/d}}{\epsilon_{2,n} r_n^2} + 1 \right) - \frac{C_f V_{d-1} (b - V_d^{1/d}/8)^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)2^{d-1}} \left(1 - M_{\alpha, d, \frac{1}{2}, 2bV_d^{-1/d}} \left(\frac{99\rho_n^2}{2r_n}, 0, r_n, \epsilon_{2,n} \right) \right)} \\ &+ e^{(d-1) \log 2d + (d-1) \log \left(\frac{2V_d^{-1/d}}{\epsilon_{3,n} r_n^2} + 1 \right) - \frac{C_f V_{d-1} (b - V_d^{1/d}/8)^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1)4^{d-1}} \left(1 - M_{\alpha, d, \frac{1}{4}, 2bV_d^{-1/d} - \frac{3}{16}} \left(\frac{99\rho_n^2}{4r_n} + \frac{353r_n^2}{2048} V_d^{1/d}, \frac{r_n^3}{128} V_d^{2/d}, r_n, \epsilon_{3,n} \right) \right)}, \end{aligned}$$

where $M_{\alpha, d, \mu, \xi}(p_n, q_n, r_n, \epsilon_n)$ is defined in Equation (2.5.17).

Then

$$(2.5.60) \quad P \left((\hat{H}_{r_n} \Delta S) \subset \partial S \bigoplus br_n^2 \mathcal{B} \right) \geq 1 - \psi(d, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}).$$

We want see how fast the dimension d can blow up while $\psi(d, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n})$ still converges to 0. And we are going to use the following Stirling's approximation,

$$(2.5.61) \quad \begin{aligned} V_d &= \frac{\pi^{d/2}}{\Gamma(\frac{d}{2} + 1)} \sim \frac{1}{\sqrt{d\pi}} \left(\frac{2\pi e}{d} \right)^{\frac{d}{2}}, \\ B\left(\frac{\alpha+1}{2}, \frac{d+1}{2}\right) &\sim \Gamma\left(\frac{\alpha+1}{2}\right) \left(\frac{d+1}{2}\right)^{-\frac{\alpha+1}{2}}. \end{aligned}$$

First we let $r_n = \left(\frac{1}{x_n}\right)^{\frac{1}{d+1+2\alpha}}$ for some x_n that converges to ∞ as $n \rightarrow \infty$. For simplicity, we do not change the notation of d . But keep in mind that d can change with n .

Let's look at the exponent of the second term of the function $\psi(d, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n})$.

$$(2.5.62) \quad \begin{aligned} &(d-1) \log 2d + (d-1) \log \left(\frac{2V_d^{-\frac{1}{d}}}{\epsilon_{2,n} r_n^2} + 1 \right) - \frac{C_f V_{d-1} (b - V_d^{\frac{1}{d}}/8)^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1) 2^{d-1}} \left(1 - M_{\alpha, d, \frac{1}{2}, 2bV_d^{-\frac{1}{d}}} \left(\frac{99\rho_n^2}{2r_n}, 0, r_n, \epsilon_{2,n} \right) \right) \\ &\leq d \log 2d + d \log \frac{3}{\epsilon_{2,n}} + d \log V_d^{-\frac{1}{d}} - d \log r_n^2 - \frac{C_f V_{d-1} (b - V_d^{\frac{1}{d}}/8)^{\alpha+1} n r_n^{d+1+2\alpha}}{(\alpha+1) 2^d} \\ &= d \log d + d \log \frac{6}{\epsilon_{2,n}} - \log V_d + \frac{2d}{d+1+2\alpha} \log x_n - \frac{C_f V_{d-1} (b - V_d^{\frac{1}{d}}/8)^{\alpha+1}}{(\alpha+1) 2^d} \frac{n}{x_n} \\ &\sim \frac{3}{2} d \log d + d \log \frac{6}{\epsilon_{2,n} \sqrt{2\pi e}} + \frac{1}{2} \log(\pi d) + \frac{2d}{d+1+2\alpha} \log x_n - \frac{C_f (b - V_d^{\frac{1}{d}}/8)^{\alpha+1}}{2(\alpha+1) \sqrt{\pi}} (d-1)^{-\frac{d}{2}} \left(\frac{\pi e}{2} \right)^{\frac{d-1}{2}} \frac{n}{x_n}. \end{aligned}$$

First we observe that

- (1) If d is fixed, then in order for the above quantity to converge to $-\infty$, the fourth term needs to grow slower than the absolute value of the fifth term. This means that we want x_n to grow slowly. So r_n converges slowly.
- (2) Now let d grow with n . x_n cannot grow too slowly. For example, if $x_n = d$, then $r_n \rightarrow 1$. Then the estimator will never converge with respect to the symmetric difference distance.
- (3) For the sum of the first four terms, only the first one and the fourth one can be the dominating term.

If we let $x_n = d^d$ then

(2.5.63)

$$\begin{aligned} & \frac{3}{2}d \log d + d \log \frac{6}{\epsilon_{2,n} \sqrt{2\pi e}} + \frac{1}{2} \log(\pi d) + \frac{2d}{d+1+2\alpha} \log x_n - \frac{C_f(b - V_d^{\frac{1}{d}}/8)^{\alpha+1}}{2(\alpha+1)\sqrt{\pi}} (d-1)^{-\frac{d}{2}} \left(\frac{\pi e}{2}\right)^{\frac{d-1}{2}} \frac{n}{x_n} \\ &= \frac{3}{2}d \log d + d \log \frac{6}{\epsilon_{2,n} \sqrt{2\pi e}} + \frac{1}{2} \log(\pi d) + \frac{2d}{d+1+2\alpha} d \log d - \frac{C_f(b - V_d^{\frac{1}{d}}/8)^{\alpha+1}}{2(\alpha+1)\sqrt{\pi}} (d-1)^{-\frac{d}{2}} d^{-d} \left(\frac{\pi e}{2}\right)^{\frac{d-1}{2}} n. \end{aligned}$$

If we let $d \log d \ll (d-1)^{-\frac{d}{2}} d^{-d} \left(\frac{\pi e}{2}\right)^{\frac{d-1}{2}} n$, the above quantity converges to $-\infty$ as we want. So we have $d = O\left(\frac{\log n}{\log \log n}\right)$. The analysis for the first and second terms of the function $\psi(d, \alpha, C_f, b, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n})$ is similar.

The maximum speed of the dimension d goes to ∞ depends on how fast the radius r_n approaches zero. The faster the small ball radius r_n can go to zero, the slower d_n can blow up. This can be explained by that if r_n is fixed for all n , as the dimension goes higher, the volume of balls of radius r_n gets smaller. If r_n go to zero faster, the volume of balls of radius r_n gets even smaller (with the same dimension d) so fewer points can be used to build local convex hulls.

Notice that the error coefficient b can also be arbitrarily small when the dimension d gets large.

2.5.3. A Local Concentration Bound. The global result is bottlenecked by the worse case scenario. The rolling ball radius R is upper bounded by the radius of the smallest balls that can roll freely inside and outside S . However, there are regions where a larger ball can roll freely (i.e., whose boundary is more smooth). In the following, we state a local bound, which may have better constants than the global case.

Let's first define a local neighborhood. Suppose $x \in \partial S$, let $\text{CN}(x, r_x) \subset \partial S$ be a connected neighborhood of x such that $\forall y \in \text{CN}(x, r_x)$, the distance between x and y along ∂S is at most r_x . Suppose that $\forall y \in \text{CN}(x, r_x)$, there exists $O_y^- \in \mathbb{R}^d$ such that $y \in B(O_y^-, R_x) \subset \overline{S}^c$ and there exists $O_y^+ \in \mathbb{R}^d$ such that $y \in B(O_y^+, R_x) \subset \overline{S}$. Now let's grow this neighborhood of to be a compact set by union the balls of radius R_x that can roll on $\text{CN}(x, r_x)$ inside S .

$$(2.5.64) \quad \text{LN}(x, r_x) = \cup_{\forall c \in S \text{ s.t. } c^* \in \text{CN}(x, r_x) \text{ and } d(c, c^*) = R_x} B(c, R_x)$$

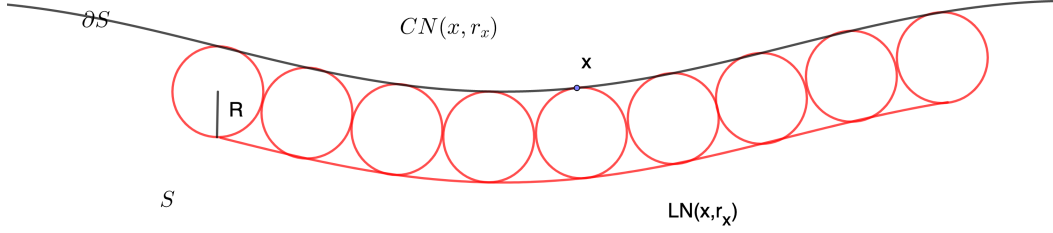


FIGURE 2.9. Growing $CN(x, r_x)$ to get $LN(x, r_x)$.

THEOREM 2.5.2. *Let $\mathcal{X}_n$ be an i.i.d. sample in $\mathbb{R}^d$ from an α -decreasing density function f that has a compact d -manifold support S . Let $x \in \partial S$. Let $CN(x, r_x) \subset \partial S$ and $LN(x, r_x) \subset S$ be neighborhoods of x as defined above. Let $\hat{H}_{r_n}^{local}$ be the local convex hull estimator built from $\mathcal{X}_n \cap LN(x, r_x)$. Suppose conditions (1 ~ 6) in Theorem 2.5.1 hold with $R = R_x$ and $b = b_x$. Then*

$$(2.5.65) \quad P \left((\hat{H}_{r_n}^{local} \Delta CN(x, r_x)) \subset \partial CN(x, r_x) \bigoplus b_x r_n^2 \mathcal{B} \right) \geq 1 - \Psi(LN(x, r_x), \partial LN(x, r_x), d, R_x, \alpha, C_f, b_x, r_n, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}).$$

When the local rolling ball radius R_x is large, the value of b_x can be chosen small. This means the error term with respect to the symmetric difference is smaller when the boundary of the region is more smooth. The proof is the same as the global bound. One just needs to check all the assumptions in Theorem 2.5.1 are satisfied for the local neighborhoods.

2.6. Concentration Bounds for the k NN LoCoH Estimator $\hat{H}_{k_n}$

Similar to the r -ball LoCoH estimator $\hat{H}_{r_n}$, we now derive a global finite sample concentration bound for the k NN LoCoH estimator $\hat{H}_{k_n}$.

2.6.1. A Global Concentration Bound.

THEOREM 2.6.1. *Let $\mathcal{X}_n$ be an i.i.d. sample in $\mathbb{R}^d$ from a density function f . Let $S \subset \mathbb{R}^d$ be the support of f . Suppose that*

- (a) S is compact d -manifold,
- (b) S satisfies the rolling ball condition with $R_{out} = R_{in} = R > 0$,

$$(c) \ f_\infty = \sup_{x \in S} f(x) < \infty,$$

$$(d) \ f_0 = \inf_{x \in S} f(x) > 0.$$

Let $\eta_1 = \left(\frac{1}{2f_\infty V_d}\right)^{\frac{1}{d}}$, $\eta_2 = \left(\frac{3}{2\gamma f_0 V_d}\right)^{\frac{1}{d}}$, $\gamma = \frac{1}{2}I_{\frac{3}{4}}\left(\frac{d+1}{2}, \frac{1}{2}\right)$ and $\eta_1 \leq \eta_2$. For positive integer k_n and positive numbers $b, \rho_n, \epsilon_{i,n}$ for $i \in \{1, 2, 3\}$ (for their meanings, see the discussions below), when the following conditions are satisfied,

$$(1) \ k_n < n \left(\frac{R}{2\eta_2}\right)^d,$$

$$(2) \ b \left(\frac{k_n}{n}\right)^{\frac{2}{d}} < \rho_n < \frac{1}{24}\eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}},$$

$$(3) \ b > \max \left(\frac{\eta_1^2}{8R}, \frac{2\left(\frac{k_n}{n}\right)^{\frac{2}{d}}}{R} + \frac{3\eta_2\left(\frac{k_n}{n}\right)^{\frac{3}{d}}}{8R^2} \right),$$

$$(4) \ \epsilon_{1,n} < \frac{3\rho_n}{2},$$

$$(5) \ \epsilon_{2,n} < \min \left(\frac{\frac{1}{2(d-1)} - \frac{99\rho_n^2}{\eta_1^2\left(\frac{k_n}{n}\right)^{\frac{2}{d}}}}{\eta_1\left(\frac{k_n}{n}\right)^{\frac{1}{d}}(2 + \sqrt{1 + \frac{b^2 R^2}{\eta_1^4} + 2R})}, \frac{4bR}{R(4\eta_1^2 + \sqrt{\eta_1^4 + b^2 R^2})}, \frac{\frac{8bR - \eta_1^2}{4(\alpha+1)} - \frac{4\eta_1^3}{9R}\left(\frac{k_n}{n}\right)^{\frac{1}{d}}}{R(4\eta_1^2 + \sqrt{\eta_1^4 + b^2 R^2})} \right),$$

$$(6) \ \epsilon_{3,n} < \min \left(\frac{\frac{1}{2(d-1)} - \frac{99\rho_n^2}{\eta_1^2\left(\frac{k_n}{n}\right)^{\frac{2}{d}}} - \frac{353\eta_1}{512R}\left(\frac{k_n}{n}\right)^{\frac{1}{d}}}{\eta_1\left(\frac{k_n}{n}\right)^{\frac{1}{d}}(4 + \sqrt{1 + (\frac{2bR}{\eta_1^2} - \frac{3}{16})^2 + 4R})}, \frac{8bR - \frac{3}{4}\eta_1^2 - \frac{\eta_1^4}{16R}\left(\frac{k_n}{n}\right)^{\frac{2}{d}}}{R(8\eta_1^2 + \sqrt{\eta_1^4 + (2bR - \frac{3}{16}\eta_1^2)^2})}, \frac{\frac{8bR - \eta_1^2}{2(\alpha+1)} - \frac{\eta_1^3}{9R}\left(\frac{k_n}{n}\right)^{\frac{1}{d}} - \frac{\eta_1^4}{16R}\left(\frac{k_n}{n}\right)^{\frac{2}{d}}}{R(8\eta_1^2 + \sqrt{\eta_1^4 + (2bR - \frac{3}{16}\eta_1^2)^2})} \right),$$

we have

(2.6.1)

$$P \left((\hat{H}_{k_n} \Delta S) \subset \partial S \oplus b \left(\frac{k_n}{n}\right)^{\frac{2}{d}} \mathcal{B} \right) \geq 1 - \Psi(S, \partial S, d, R, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) - 2\delta_n,$$

where $\delta_n = 2e^{-\left(\frac{k_n}{1024d \log n}\right)^{1/2}}$ and $\Psi(\dots)$ is defined in Equation (2.5.29).

DISCUSSION. k_n is the number of nearest neighbors that are used to build the local convex hull for the k NN LoCoH estimator. k_n needs to be small because the radius of the k -nearest neighbors needs to be small enough relatively to the rolling ball radius R such that the region where these data points sampled is locally “convex”. However, it cannot be too small otherwise the right-hand side of the Inequality (2.6.1) is small.

The meanings of $\rho_n, \epsilon_{i,n}$ for $i \in \{1, 2, 3\}$ and the error coefficient b are the same as in Theorem reffound:hrglobal. We omit the discussions here.

□

PROOF. We first find the lower bound $r_{1,n}$ and the upper bound $r_{2,n}$ for the k_n -nearest neighbor distance $R_{k_n}(x)$ for $x \in S$, i.e., with probability at least $1 - \delta_n$, $\inf_{x \in S} R_{k_n}(x) \geq r_{1,n}$ and $\sup_{x \in S} R_{k_n}(x) \leq r_{2,n}$. When this inequality is true, $\hat{H}_{r_{1,n}} \subset \hat{H}_{k_n} \subset \hat{H}_{r_{2,n}}$. So $(\hat{H}_{k_n} \cap S^c) \subset (\hat{H}_{r_{2,n}} \cap S^c)$ and $(\hat{H}_{k_n}^c \cap S) \subset (\hat{H}_{r_{1,n}}^c \cap S)$. So the symmetric difference $\hat{H}_{k_n} \Delta S = (\hat{H}_{k_n} \cap S^c) \cup (\hat{H}_{k_n}^c \cap S)$ is contained in the union of $(\hat{H}_{r_{2,n}} \cap S^c)$ and $(\hat{H}_{r_{1,n}}^c \cap S)$, whose behavior is well studied in the proof of Theorem 2.5.1. The probability that $\hat{H}_{k_n} \Delta S$ falls outside of the union is small and controlled by δ_n .

(i) The bounds for the k_n -nearest neighbor distance.

First

$$(2.6.2) \quad \frac{2^8}{(3/2 - 1)^2} \log^2\left(\frac{2}{\delta_n}\right) d \log n = \frac{2^8}{(1 - 1/2)^2} \log^2\left(\frac{2}{\delta_n}\right) d \log n = k_n.$$

By Lemma 2.2.1, the assumption (b) implies $\text{Leb}(S \cap B(x, r)) \geq \gamma \text{Leb}(B(x, r))$ for $0 < r \leq R$. Let $r_{1,n} = \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}$ and $r_{2,n} = \eta_2 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}$.

By Corollary 2.3.1,

$$(2.6.3) \quad P\left(\sup_{x \in S} R_k(x) \leq \left(\frac{3}{2\gamma f_0 V_d} \frac{k}{n}\right)^{\frac{1}{d}} = r_{2,n}\right) \geq 1 - \delta_n.$$

By Corollary 2.3.3,

$$(2.6.4) \quad P\left(\inf_{x \in S} R_k(x) \geq \left(\frac{b}{f_\infty V_d} \frac{k}{n}\right)^{\frac{1}{d}} = r_{1,n}\right) \geq 1 - \delta_n.$$

Write events $U = \{\sup_{x \in S} R_k(x) \leq r_{2,n}\}$, $L = \{\inf_{x \in S} R_k(x) \geq r_{1,n}\}$. Then $P(U) \geq 1 - \delta_n$ and $P(L) \geq 1 - \delta_n$.

(ii) Event $E_1 = \{(\hat{H}_{k_n} \cap S^c) \subset \partial S \oplus b \left(\frac{k_n}{n}\right)^{\frac{2}{d}}\}^c$

Given the event U , we have $\hat{H}_{k_n} \subset \hat{H}_{r_{2,n}}$. So $(\hat{H}_{r_{2,n}} \cap S^c) \subset \partial S \oplus \frac{b}{\eta_2^2} r_{2,n}^2 \implies E_1^c$. Let event $A_1 = \{(\hat{H}_{r_{2,n}} \cap S^c) \subset \partial S \oplus \frac{b}{\eta_2^2} r_{2,n}^2\}$. So $P(E_1|U) = 1 - P(E_1^c|U) \leq 1 - P(A_1|U)$.

So

$$\begin{aligned}
P(E_1) &= P(E_1 \cap U) + P(E_1 \cap U^c) \\
&= P(E_1|U)P(U) + P(E_1|U^c)P(U^c) \\
(2.6.5) \quad &\leq P(E_1|U) + P(U^c) \\
&\leq (1 - P(A_1|U)) + P(U^c) \\
&\leq \delta_n.
\end{aligned}$$

The last inequality is because when $r_{2,n} < \frac{R}{2}$ and $\frac{b}{\eta_2^2} > \frac{2r_{2,n}^2}{R} + \frac{3r_{2,n}^3}{8R^2}$, $P((\hat{H}_{r_{2,n}} \cap S^c) \subset \partial S \oplus \frac{b}{\eta_2^2} r_{2,n}^2) = 1$ by Equality (2.5.37) in the proof of Theorem 2.5.1.

$$(iii) \text{ Event } E_2 = \{(\hat{H}_{k_n}^c \cap S) \subset \partial S \oplus b \left(\frac{k_n}{n}\right)^{\frac{2}{d}}\}^c$$

Given the event L , we have $\hat{H}_{r_{1,n}} \subset \hat{H}_{k_n}$, so $\hat{H}_{r_{1,n}}^c \cap S \subset \partial S \oplus \frac{b}{\eta_1^2} r_{1,n}^2 \implies E_2^c$. Let event $A_2 = \{\hat{H}_{r_{1,n}}^c \cap S \subset \partial S \oplus \frac{b}{\eta_1^2} r_{1,n}^2\}$. So $P(E_2|L) \leq 1 - P(A_2|L)$.

So

$$\begin{aligned}
P(E_2) &= P(E_2 \cap L) + P(E_2 \cap L^c) \\
&= P(E_2|L)P(L) + P(E_2|L^c)P(L^c) \\
(2.6.6) \quad &\leq P(E_2|L) + P(L^c) \\
&\leq (1 - P(A_2|L)) + P(L^c) \\
&\leq \Psi(S, \partial S, d, R, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) + \delta_n.
\end{aligned}$$

The last inequality is because

$$P((\hat{H}_{r_{1,n}} \cap S^c) \subset \partial S \oplus \frac{b}{\eta_1^2} r_{1,n}^2) \geq 1 - \Psi(S, \partial S, d, R, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n})$$

when conditions (1 ~ 6) are satisfied by Inequalities (2.5.39, 2.5.43&2.5.48) in the proof of Theorem 2.5.1.

(iv) *Combine the results*

(2.6.7)

$$\begin{aligned}
P((\hat{H}_{k_n} \Delta S) \subset \partial S \bigoplus br_n^2) &= P(E_1^c \cap E_2^c) \\
&= 1 - P(E_1 \cup E_2) \\
&\geq 1 - P(E_1) - P(E_2) \\
&\geq 1 - \delta_n - \Psi(S, \partial S, d, R, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) - \delta_n \\
&= 1 - 2\delta_n - \Psi(S, \partial S, d, R, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n})
\end{aligned}$$

□

Now we state two corollaries that can be derived directly from the above theorem, which are used to study the consistency and high dimension behavior of the r -ball LoCoH estimator.

2.6.1.1. *Consistency.* The first corollary is obtained by plugging in the bounds for the covering numbers $N(S, \epsilon)$ and $N(\partial S, \epsilon)$ from Proposition 2.2.2.

COROLLARY 2.6.1. *With the same notations and assumptions as in Theorem 2.6.1, we have*

(2.6.8)

$$P\left((\hat{H}_{k_n} \Delta S) \subset \partial S \bigoplus b \left(\frac{k_n}{n}\right)^{\frac{2}{d}} \mathcal{B}\right) \geq 1 - \Psi^*(S, \partial S, d, R, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) - 2\delta_n,$$

where $\Psi^*(\dots)$ is defined in Equation (2.5.49).

PROOF. By Proposition 2.2.2, the assumption (a) in Theorem 2.6.1 implies that there are positive numbers $\lambda_S, \lambda_{\partial S}$ such that $\forall \epsilon > 0$ the covering numbers

$$\begin{aligned}
N(S, \epsilon) &\leq \lambda_S \epsilon^{-d}, \\
N(\partial S, \epsilon) &\leq \lambda_{\partial S} \epsilon^{1-d}.
\end{aligned}$$

Plugging them in the Inequality (2.6.1) finishes the proof.

□

Let $k_n > 0$ be a sequence of positive integers such that $k_n \rightarrow \infty$ and $\frac{k_n}{n} \rightarrow 0$ as $n \rightarrow \infty$. Let $\rho_n, \epsilon_{1,n}$ be sequences of positive numbers that converges to 0 as $n \rightarrow \infty$. And $(\frac{k_n}{n})^2 \ll \rho_n \ll r_n, \epsilon_{1,n} \ll \rho_n$. Let $\epsilon_{2,n}, \epsilon_{3,n} > 0$ be such that

(2.6.10)

$$\begin{aligned} \limsup_{n \rightarrow \infty} \epsilon_{2,n} &< \min \left(\frac{4bR}{R(4\eta_1^2 + \sqrt{\eta_1^4 + b^2 R^2})}, \frac{\frac{8bR - \eta_1^2}{4(\alpha+1)}}{R(4\eta_1^2 + \sqrt{\eta_1^4 + b^2 R^2})} \right), \\ \limsup_{n \rightarrow \infty} \epsilon_{3,n} &< \min \left(\frac{8bR - \frac{3}{4}\eta_1^2}{R(8\eta_1^2 + \sqrt{\eta_1^4 + (2bR - \frac{3}{16}\eta_1^2)^2})}, \frac{\frac{8bR - \eta_1^2}{2(\alpha+1)}}{R(8\eta_1^2 + \sqrt{\eta_1^4 + (2bR - \frac{3}{16}\eta_1^2)^2})} \right). \end{aligned}$$

If there exists a constant $b > \frac{\eta_1^2}{8R}$ such that

$$\begin{aligned} (1) \quad &-d \log \epsilon_{1,n} - \frac{C_f V_{d-1} B(\frac{\alpha+1}{2}, \frac{d+1}{2}) n(\frac{3\rho_n}{2})^{\alpha+d}}{2} \rightarrow -\infty, \\ (2) \quad &(1-d) \log (\epsilon_{2,n} \eta_1^2 (\frac{k_n}{n})^{\frac{2}{d}}) - \frac{C_f V_{d-1} (\frac{b}{\eta_1^2} - \frac{1}{8R})^{\alpha+1} n \eta_1^{d+1+2\alpha} (\frac{k_n}{n})^{\frac{d+1+2\alpha}{d}}}{2(\alpha+1)2^{d-1}} \rightarrow -\infty, \\ (3) \quad &(1-d) \log (\epsilon_{3,n} \eta_1^2 (\frac{k_n}{n})^{\frac{2}{d}}) - \frac{C_f V_{d-1} (\frac{b}{\eta_1^2} - \frac{1}{8R})^{\alpha+1} n \eta_1^{d+1+2\alpha} (\frac{k_n}{n})^{\frac{d+1+2\alpha}{d}}}{2(\alpha+1)4^{d-1}} \rightarrow -\infty, \end{aligned}$$

as $n \rightarrow \infty$, then

$$(2.6.11) \quad P \left((\hat{H}_{k_n} \Delta S) \subset \partial S \bigoplus b \left(\frac{k_n}{n} \right)^{\frac{2}{d}} \mathcal{B} \right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

Note that $k_n \asymp n^{\frac{1}{d+1}} (\log n)^{\frac{d}{d+1}}$, $\rho_n \asymp \left(\frac{\log n}{n} \right)^{\frac{1}{d}}$, $\epsilon_{1,n} \asymp \left(\frac{1}{n} \right)^{\frac{1}{d}}$, $\epsilon_{2,n} \asymp \epsilon_{3,n} \asymp (\log n)^{-2}$ is a valid choice to ensure the consistency of the k NN LoCoH estimator $\hat{H}_{k_n}$ with respect to the symmetric difference distance.

2.6.1.2. High Dimension Performance. The second corollary is obtained by plugging in the bounds of the covering numbers $N(S, \epsilon)$ and $N(\partial S, \epsilon)$ from Proposition 2.2.3.

COROLLARY 2.6.2. *With the same notations and assumptions as in Theorem 2.6.1, we have*

(2.6.12)

$$P\left((\hat{H}_{k_n} \Delta S) \subset \partial S \bigoplus b\left(\frac{k_n}{n}\right)^{\frac{2}{d}} \mathcal{B}\right) \geq 1 - \Psi^{**}(S, \partial S, d, R, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) - 2\delta_n,$$

where $\Psi^{**}(\dots)$ is defined in Equation (2.5.54).

PROOF. By Proposition 2.2.3, the assumption (b) in Theorem 2.6.1 implies that $\forall \epsilon > 0$ the covering numbers

$$(2.6.13) \quad N(S, \epsilon) \leq \left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon} + \frac{1}{R} \right)^d,$$

and

$$(2.6.14) \quad N(\partial S, \epsilon) \leq 2 \left[\left(\frac{\text{Vol}_{d-1}(\partial S) \Gamma(d/2 + 1)}{\pi^{d/2}} \right)^{\frac{d}{d-1}} \left(\frac{2}{\epsilon} + \frac{1}{R} \right)^d - \frac{2^d \text{Vol}_d(S) \Gamma(d/2 + 1)}{\pi^{d/2} \epsilon^d} \right].$$

Plugging them in the Inequality (2.6.1) finishes the proof.

□

First of all, in order to ensure that $b\left(\frac{k_n}{n}\right)^{\frac{2}{d}} \rightarrow 0$ to control the error in symmetric difference distance, we need $\frac{\eta_1^2}{8R} \left(\frac{k_n}{n}\right)^{\frac{2}{d}} \rightarrow \infty$ as $n \rightarrow \infty$. So $d = O\left(\frac{\log \frac{n}{k_n}}{\log \log \frac{n}{k_n}}\right)$, and further $d = O\left(\frac{\log n}{\log \log n}\right)$. Then we can do a similar analysis on a special case when S is a d -dimensional ball of volume 1. Then the boundary of S has $d - 1$ dimension volume $dV_d^{1/d}$. The rolling ball radius $R = V_d^{-1/d}$.

Then

(2.6.15)

$$P\left((\hat{H}_{k_n} \Delta S) \subset \partial S \bigoplus b\left(\frac{k_n}{n}\right)^{\frac{2}{d}} \mathcal{B}\right) \geq 1 - \psi\left(d, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}\right),$$

where $\psi(\dots)$ is defined in Equation (2.5.59).

We can do the same analysis on the function $\psi\left(d, 0, f_0, \frac{b}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}\right)$ as in the Section 2.5.2.2 to reach the following observation.

- (1) If d is fixed, we want k_n to blow up fast so that ψ vanishes.
- (2) Now let d grow with n . k_n cannot grow too slowly.

- (3) The maximum speed of the dimension d goes to ∞ depends on how fast the radius k_n approaches ∞ . The slower the number of neighbors k_n can go to ∞ , the slower d_n can blow up.

The dimension $d = O\left(\frac{\log n}{\log \log n}\right)$ also ensures the function $\psi \rightarrow 0$ as $n \rightarrow \infty$. Notice that the error coefficient $b \rightarrow \infty$ as the dimension $d \rightarrow \infty$. This is because the radius of a local k NN set is lower bounded by the value $\eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}$ and upper bounded by the value $\eta_2 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}$. So the absolute value of the difference $|\eta_2 - \eta_1|$ can go to ∞ as $d \rightarrow \infty$. And the error coefficient b needs to accommodate this large fluctuation.

2.6.2. A Local Concentration Bound. Similar to the r -ball LoCoH estimator $\hat{H}_{r_n}$ case, suppose $x \in \partial S$. Let $CN(x, r_x) \subset \partial S$ be a connected neighborhood of x such that $\forall y \in CN(x, r_x)$, the distance between x and y along ∂S is at most r_x . Let $LN(x, r_x) \subset S$ be the expanded neighborhood of x defined in Equation (2.5.64). The following local concentration bound holds.

THEOREM 2.6.2. *Let $\mathcal{X}_n$ be an i.i.d. sample in $\mathbb{R}^d$ from an density function f that has a compact d -manifold support S . Let $x \in \partial S$. Let $CN(x, r_x) \subset \partial S$ and $LN(x, r_x) \subset S$ be neighborhoods of x as defined above. Let $\hat{H}_{k_n}^{local}$ be a local k NN LoCoH estimate built using $\mathcal{X}_n \cap LN(x, r_x)$. Suppose conditions (1 $\sim$ 6) in Theorem 2.5.1 hold with $R = R_x$ and $b = b_x$. Then*

$$(2.6.16) \quad \begin{aligned} & P\left(\left(\hat{H}_{r_n}^{local} \Delta LN(x, r_x)\right) \subset \partial LN(x, r_x) \oplus b_x \left(\frac{k_n}{n}\right)^{\frac{2}{d}} \mathcal{B}\right) \\ & \geq 1 - \Psi(S, \partial S, d, R_x, 0, f_0, \frac{b_x}{\eta_1^2}, \eta_1 \left(\frac{k_n}{n}\right)^{\frac{1}{d}}, \rho_n, \epsilon_{1,n}, \epsilon_{2,n}, \epsilon_{3,n}) - 2\delta_n, \end{aligned}$$

where $\delta_n = 2e^{-\left(\frac{k_n}{1024d \log n}\right)^{1/2}}$ and $\Psi(\dots)$ is defined in Equation (2.5.29).

Similar to the case of r -ball LoCoH estimator, when the local rolling ball radius R_x is large, the value of b_x can be chosen small. This means the error term with respect to the symmetric difference is smaller when the boundary of the region is more smooth. The proof is the same as the global bound. One just needs to check all the assumptions in Theorem 2.6.1 are satisfied for the local neighborhoods.

2.7. Appendix: Cited Theorems

Suppose F_n is the empirical measure corresponding to the sample $\mathcal{X}_n$.

LEMMA 2.7.1 (Chaudhuri and Dasgupta (2010)). *For $0 < \delta < 1$ and $k \geq d \log n$. With probability at least $1 - \delta$, for every ball $B \subset \mathbb{R}^d$,*

$$(2.7.1) \quad \begin{aligned} F(B) &\geq C_{\delta,n} \frac{\sqrt{d \log n}}{n} \implies F_n(B) > 0, \\ F(B) &\geq \frac{k}{n} + C_{\delta,n} \frac{\sqrt{k}}{n} \implies F_n(B) \geq \frac{k}{n}, \text{ and} \\ F(B) &\leq \frac{k}{n} - C_{\delta,n} \frac{\sqrt{k}}{n} \implies F_n(B) < \frac{k}{n}, \end{aligned}$$

where $C_{\delta,n} = 2^4 \log(2/\delta) \sqrt{d \log n}$.

THEOREM 2.7.1 (Aaron and Bodart (2016)). *Suppose F is absolutely continuous with respect to the Lebesgue measure. Assume that the support of F is a compact d -manifold $S \subset \mathbb{R}^d$. Assume that the probability density f is α -quickly decreasing and that there exists $R > 0$ such that balls of radius R roll freely inside and outside S . Let*

$$(2.7.2) \quad r_n = \lambda (\log n / n)^{1/(d+1+2\alpha)}, \quad \lambda > 0;$$

then

$$(2.7.3) \quad \text{Leb}(\hat{H}_{r_n} \Delta S) (n / \log n)^{2/(d+1+2\alpha)} \text{ is e.a.s. bounded.}$$

CHAPTER 3

Generalized Switching Relation and Its Applications

3.1. Introduction

The idea of statistical estimations under shape constraints is to assume that the object f to estimate lies in a space $\mathcal{F}$, whose elements satisfy certain shape constraints, for example, monotonicity or convexity, alternative to smoothness constraints. Two popular approaches to nonparametric estimation under such constraints are the least squares estimator (LSE) and the maximum likelihood estimator (MLE), where these criteria functions are optimized over the constraint class $\mathcal{F}$. Advantages of this approach over many popular parametric method include that it usually imposes fewer requirements on $\mathcal{F}$ and that the estimating procedure is automatic in the sense that no additional tuning is needed.

Grenander (1956) found that the MLE of monotonically decreasing density functions defined on $[0, \infty)$ is the left-hand derivative of the least concave majorant of the empirical cumulative distribution function. The asymptotic properties such as convergence rate and limiting distribution of the Grenander estimator were then studied by Rao (1969) and later by Groeneboom (1984). However, the estimation of monotonic functions in higher dimension is much less explored. There are many possible orders in multivariate cases, all leading to different types of monotonicity. One example are block-decreasing functions, which refers to functions that are decreasing with respect to every variable. Polonik (1998) has shown that such MLEs can be obtained in a similar fashion as the classical Grenander estimator. Polonik (1998), Biau and Devroye (2003), Gao and Wellner (2007) and Pavlides (2012) studied the block-decreasing density functions and bounds for convergence rate of its nonparametric MLE with respect to certain norms when the dimension is 2 and above. However, no limiting distribution result is available for monotonic density estimators in high dimensions.

In this chapter, we are attempting to address this problem. One difficulty is that such monotonic density estimators, and the nonparametric maximum likelihood estimators (NPMLEs) under shape constraints in general, are defined indirectly, and an explicit representation is not available. The key in Groeneboom's proof for the limiting distribution of Grenander estimator is the so-called switching relation, which is used to quantify the behavior of density estimators through the behavior of the empirical cdf. Its validity is closely related to a majorization process. Polonik (1998) has shown that there exists a similar graphical representation of the nonparametric MLEs under order constraints with a similar majorization process. We use this representation to propose a generalized switching relation and outline the idea to obtain asymptotic results of monotonic density estimators in high dimensions.

Monotonicity is not the only shape constraint that has been considered for function estimation. Other shape constraints include two estimation problems related to monotonicity initiated by Groeneboom *et al.* (2001a) and Groeneboom *et al.* (2001b): (1) the MLE of the concave decreasing density functions on $[0, \infty)$ and (2) the least square estimator of concave regression functions. Groeneboom *et al.* (2001b) also derived an asymptotic distributional result for the ML density estimator. Less was known about the limiting distribution of these estimators in dimension other than one until recently Han and Zhang (2020) showed some distributional result in the regression setting. Another related problem is the log-concave density estimation. See Samworth (2018) for a review of recent progress with an emphasis on the ML approach.

3.2. Generalized Switching Relation

3.2.1. Classical Switching Relation. The classical switching relation was introduced when studying the Grenander estimator for monotonically decreasing functions defined on $[0, \infty)$. For a sample of n iid observations $X_1, \dots, X_n \sim f$, where f is defined on $[0, \infty)$, the Grenander estimator $\hat{f}_n(x)$ is the left-hand derivative of the least concave majorant of the empirical distribution function $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$. Define, for $\lambda > 0$,

$$(3.2.1) \quad \hat{s}_n(\lambda) = \sup_{u \geq 0} \{s \in \operatorname{argmax}\{F_n(u) - \lambda u\}\}.$$

The switching relation then states that

$$(3.2.2) \quad \hat{f}_n(x) < \lambda \iff \hat{s}_n(\lambda) < x.$$

This is equivalent to

$$(3.2.3) \quad \hat{f}_n(x) \geq \lambda \iff \hat{s}_n(\lambda) \geq x.$$

The graphical interpretation is shown in Figure 3.1. A formal proof of (3.2.3) will follow as a byproduct of the proof of a multivariate version of (3.2.3) that will be derived below (see Theorem 3.2.1).

Note that both sides of Equivalence 3.2.2 are strict inequalities. However, $\hat{f}_n(x) \leq \lambda \iff \hat{s}_n(\lambda) \leq x$ only fails to hold when $\lambda = \hat{f}_n(x_\lambda)$ for some x_λ . And $\hat{f}_n(x)$ can only take finitely many different values. The significance of this relation is that it allows us to study the empirical distribution function rather than to study the density estimator itself. Much more is known about the former function.

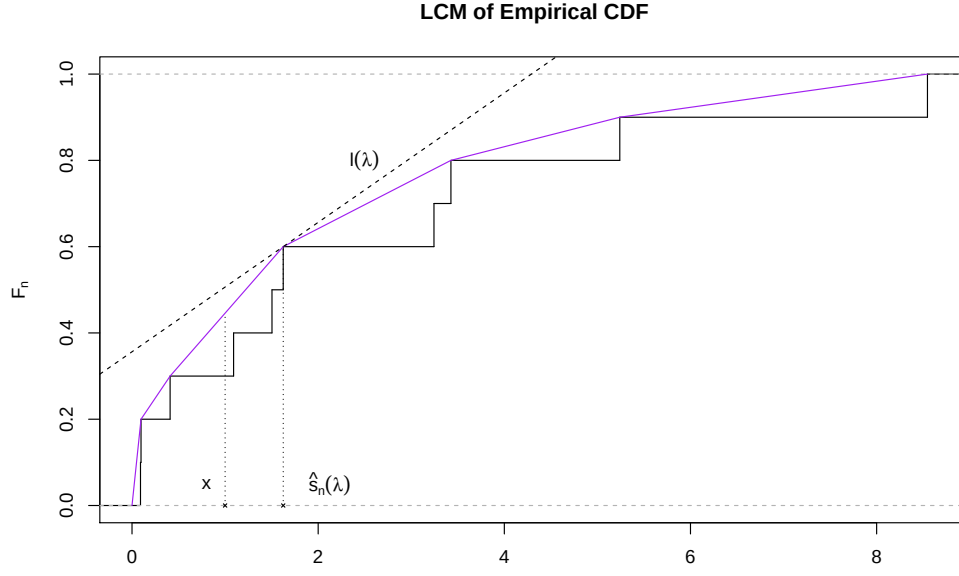


FIGURE 3.1. Illustration of classical switching relation (3.2.3). $\hat{f}_n(x) \geq \lambda$. The dotted line $l(\lambda)$ has slope λ . $\hat{s}_n(\lambda)$ is the x -coordinate where $l(\lambda)$ is tangent to the least concave majorant of the empirical cdf.

3.2.2. A Crucial Observation. First we state a simple but crucial observation, which will be used to formulate a generalized switching relation. Consider a metric space S , with Borel σ -algebra $\mathcal{S}$, and a measure μ . Let g be a measurable function defined on S . Let $\Gamma_g(\lambda) := \{t : g(t) \geq \lambda\}$ be the upper level sets of g at level λ .

PROPOSITION 3.2.1. *For $x \in S$,*

$$(3.2.4) \quad g(x) \geq \lambda \iff \Gamma_g(g(x)) \subseteq \Gamma_g(\lambda) \iff \mu(\Gamma_g(\lambda)) \geq \mu(\Gamma_g(g(x)))$$

Moreover, suppose that $\mu(\Gamma_g(g(x)) \setminus \overline{\{t \in S : g(t) > g(x)\}}) > 0$, we have

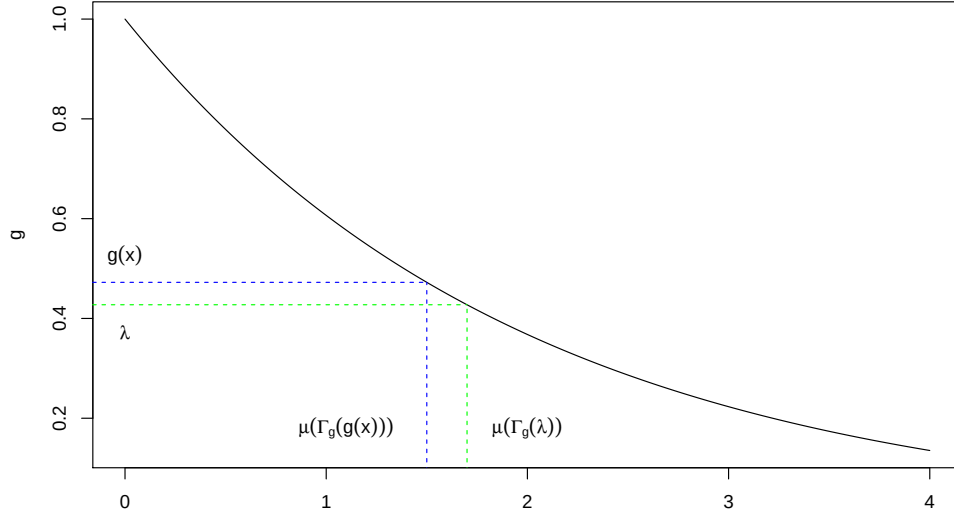
$$(3.2.5) \quad g(x) \geq \lambda \iff \mu(\Gamma_g(\lambda)) \geq \mu(D) \text{ for any set } D \text{ with } \overline{\{t \in S : g(t) > g(x)\}} \subseteq D \subseteq \{t \in S : g(t) \geq g(x)\}.$$

PROOF. If $g(x) \geq \lambda$, then $\forall t$ such that $g(t) \geq g(x)$, we have $g(t) \geq \lambda$, which means $\Gamma_g(g(x)) \subseteq \Gamma_g(\lambda)$. If $g(x) < \lambda$, $x \in \Gamma_g(g(x))$, but $x \notin \Gamma_g(\lambda)$, which means $\Gamma_g(g(x)) \not\subseteq \Gamma_g(\lambda)$. This proves that $g(x) \geq \lambda \iff \Gamma_g(g(x)) \subseteq \Gamma_g(\lambda)$.

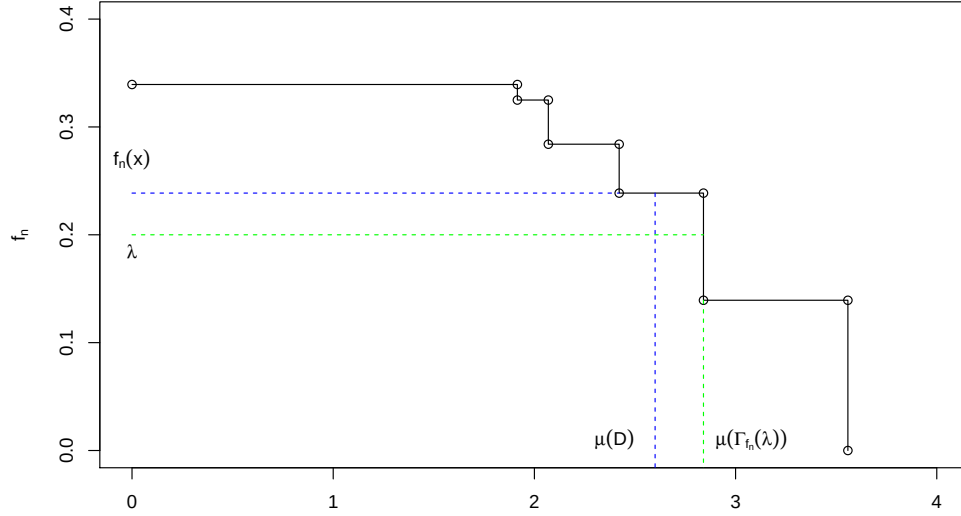
$\Gamma_g(g(x)) \subseteq \Gamma_g(\lambda) \implies \mu(\Gamma_g(\lambda)) \geq \mu(\Gamma_g(g(x)))$ is obviously true. The other direction is because of the fact that $\{\Gamma_g(\lambda) : \lambda \in \mathbb{R}\}$ is a class of nested sets by the above argument. This proves $\Gamma_g(g(x)) \subseteq \Gamma_g(\lambda) \iff \mu(\Gamma_g(\lambda)) \geq \mu(\Gamma_g(g(x)))$.

For (3.2.5), if $g(x) \geq \lambda$, then $\forall D$ such that $\overline{\{t \in S : g(t) > g(x)\}} \subseteq D \subseteq \{t \in S : g(t) \geq g(x)\}$, $\mu(\Gamma_g(\lambda)) \geq \mu(\Gamma_g(g(x))) \geq \mu(D)$. If $g(x) < \lambda$, then $\Gamma_g(\lambda) \subseteq \overline{\{t \in S : g(t) > g(x)\}} \subset \Gamma_g(g(x))$. So $\mu(\Gamma_g(\lambda)) < \mu(\Gamma_g(g(x)))$ by the assumption that $\mu(\Gamma_g(g(x)) \setminus \overline{\{t \in S : g(t) > g(x)\}}) > 0$. So let $D = \Gamma_g(g(x))$ finishes the proof. □

In (3.2.4), the measure μ can be replaced by any function G that is monotonic on the level sets of g . It turns out that the classical switching relation (3.2.3) is a special case of (3.2.5), by choosing μ equal the Lebesgue measure on the real line, $g(x) = \hat{f}_n(x)$, the Grenander estimator and $\mu(D) = x$. Note that x is a valid choice because the value x lies in between $\mu(\overline{\{t \in S : g(t) > g(x)\}})$ and $\mu(\Gamma_g(g(x)))$. Actually, any value that lies in this range could be chosen to make the argument work. The assumption $\mu(\Gamma_g(g(x)) \setminus \overline{\{t \in S : g(t) > g(x)\}}) > 0$ indicates that the “flat part” of the function g at level $g(x)$ has positive measure. The choice of D is crucial in deriving the pointwise limit distribution. In the Grenander estimator the value $\mu(D) = x$ is known and non-random. In general, $\mu(D)$ has to be chosen random. In such cases the behavior of $\mu(D)$ will influence the kind of asymptotic results that can be obtained by using the switching relation.



(a)



(b)

FIGURE 3.2. Illustration of generalized switching relation (a): (3.2.4), (b): (3.2.5). x -axis is the volume of upper level set with point x on its boundary. y -axis is the function value at the point x .

3.2.3. Order Restrictions and σ -lattice. We now talk about how to define a monotone function in dimension higher than 1. A quasi-order is a binary relation that is reflexive and

transitive. Given a quasi-order ' $\preceq$ ', the class $\mathcal{F}_{\preceq}$ of functions that are monotonically decreasing with respect to this order is defined as

$$(3.2.6) \quad \mathcal{F}_{\preceq} = \{f : \mathbb{R}^d \rightarrow \mathbb{R}, x \preceq y \iff f(x) \geq f(y)\}.$$

Given a class $\mathcal{C}$ of subsets of $\mathbb{R}^d$, the class $\mathcal{F}_{\mathcal{C}}$ of positive functions with shape constraints $\mathcal{C}$ is defined as follows,

$$(3.2.7) \quad \mathcal{F}_{\mathcal{C}} = \{f : \mathbb{R}^d \rightarrow \mathbb{R}, \Gamma_f(\lambda) \in \mathcal{C} \quad \forall \lambda \geq 0\}.$$

A classical result (see Robertson *et al.* (1988)) is that for any quasi-order ' $\preceq$ ', there exists a σ -lattice $\mathcal{C}_{\preceq}$, a class of subsets of $\mathbb{R}^d$ that is closed under countable intersections and countable unions, such that

$$(3.2.8) \quad \mathcal{F}_{\preceq} = \mathcal{F}_{\mathcal{C}_{\preceq}},$$

meaning that f is monotonically decreasing w.r.t. ' $\preceq$ ' if and only if its upper level sets fall into the class $\mathcal{C}_{\preceq}$.

For instance, let $S = [0, \infty)$, ' $\preceq$ ' = ' $\leq$ ' be the usual order in R . The corresponding σ -lattice $\mathcal{C}_{\leq}$ is the class of intervals starting at 0, i.e., $\mathcal{C}_{\leq} = \{[0, a], a \geq 0\} \cup \{[0, a), a > 0\}$. Throughout the text, $\mathcal{C}$ is assumed to be a σ -lattice (in most cases, $\mathcal{C} = \mathcal{C}_{\preceq}$) unless it is stated otherwise.

Below we discuss how to formulate the switching relation in the case of NPMLEs under order restrictions on $\mathbb{R}^d$. The underlying space setting is $(S, \mathcal{S}, \mu) = (\mathbb{R}^d, \mathcal{B}^{\mathbb{R}^d}, \text{Leb})$.

3.2.4. NPMLEs for Density Functions.

3.2.4.1. *Definition of NPMLEs.* Given a sample of n observations $X_1, \dots, X_n \sim f$ with $f \in \mathcal{F}_{\preceq}$, let $\hat{f}_{n, \preceq}(x)$ denote the NPMLE of f under the order restriction with respect to ' $\preceq$ ', i.e., $\hat{f}_{n, \preceq}(x)$ satisfies

$$(3.2.9) \quad \prod_{i=1}^n \hat{f}_{n,\preceq}(X_i) = \max \left\{ \prod_{i=1}^n f(X_i), f \in \mathcal{F}_{\preceq} \right\}.$$

3.2.4.2. *Graphical Representations.* It can be shown that the $\hat{f}_{n,\preceq}(x)$ is of a piecewise constant, assuming only finitely many values $t_1 < \dots < t_{k_n}$ on the sets $C_j \setminus C_{j+1}, j = 1, \dots, k_n$, respectively, where $C_{k_n+1} = \emptyset$, and the sets $C_{k_n} \subset C_{k_n-1} \subset \dots \subset C_1$ denote the different level sets of $\hat{f}_{n,\preceq}(x)$.

A Lévy-type concentration function $\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell)$ is defined as

$$(3.2.10) \quad \tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell) = \sup_{C \in \mathcal{C}_{\preceq}} \{F_n(C) : \text{Leb}(C) \leq \ell\}, \quad \ell \geq 0,$$

where F_n is the empirical measure, i.e., $F_n(C) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(C) = \frac{1}{n} \#\{i : X_i \in C\}$.

Let $\tilde{F}_{n,\mathcal{C}_{\preceq}}^*(\ell) = \text{LCM}(\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell))$, the least concave majorant of $\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell)$. Then $\tilde{F}_{n,\mathcal{C}_{\preceq}}^*(\ell)$ (see Figure 3.3) is a piecewise linear, increasing functions with at most n changes of slope (the number of slope changes is equal to k_n , the number of different level sets), which happen at levels ℓ for which $\tilde{F}_{n,\mathcal{C}_{\preceq}}^*(\ell) = \tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell)$. Write $\{\ell_1 > \dots > \ell_{k_n}\} = \{\ell : \tilde{F}_{n,\mathcal{C}_{\preceq}}^*(\ell) = \tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell)\}$. Then t_i is the left-hand derivatives of $\tilde{F}_{n,\mathcal{C}_{\preceq}}^*$ at ℓ_i for all $i = 1, \dots, k_n$.

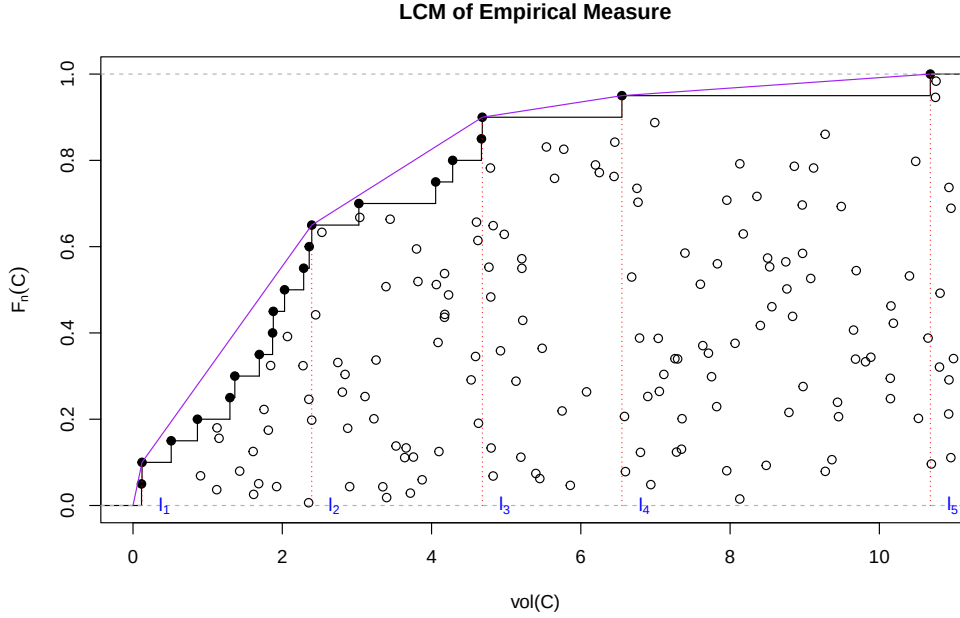


FIGURE 3.3. Graphical representation of the NPMLE.

3.2.5. General Switching Relation for Order Restricted NPMLEs. The usefulness of the classical switching relation lies in the fact that in this case, the right-hand side of (3.2.3) can be written as a functional of the empirical distribution. The question now is whether this can be achieved in multivariate situations as well. The reason why the classical switching relation can be generalized is that there is a graphical representation of the MLEs based on least concave majorants of certain Lévy-type concentration functions, which can be seen as a generalization of the well-known representation of the Grenander estimator to higher dimensions. To formulate the switching relation for NPMLEs, we first define a few quantities. Given a sample of n iid observations $X_1, \dots, X_n \sim F$, the Lévy-type concentration function $\tilde{F}_{n, \mathcal{C}_{\preceq}}(\ell)$ mentioned above is defined as

$$(3.2.11) \quad \tilde{F}_{n, \mathcal{C}_{\preceq}}(\ell) = \sup\{F_n(C) : C \in \mathcal{C}_{\preceq}, \text{Leb}(C) \leq \ell\}.$$

Define

$$(3.2.12) \quad \hat{s}_{n, \mathcal{C}_{\preceq}}^*(\lambda) = \sup\{s \in \text{argmax}\{\ell \geq 0 : \tilde{F}_{n, \mathcal{C}_{\preceq}}(\ell) - \lambda\ell\}\}.$$

It is shown in Polonik (1998) that $\hat{s}_{n, \mathcal{C}_{\preceq}}^*(\lambda)$ is equal to $\text{Leb}(\Gamma_g(\lambda))$ for $g = \hat{f}_{n, \preceq}$, and we will plug this into (3.2.5). As we have discussed before, the choice of $\mu(D)$ in (3.2.5) can be flexible, and it might depend on the structure of $\mathcal{C}_{\preceq}$. The following is when $\mu(D)$ is picked to be $\text{Leb}(\Gamma_{\hat{f}_{n, \preceq}}(\hat{f}_{n, \preceq}(x)))$. In the total order example below, when $\mathcal{C}_{\preceq}$ is relatively simple, other choices of $\mu(D)$ are being used.

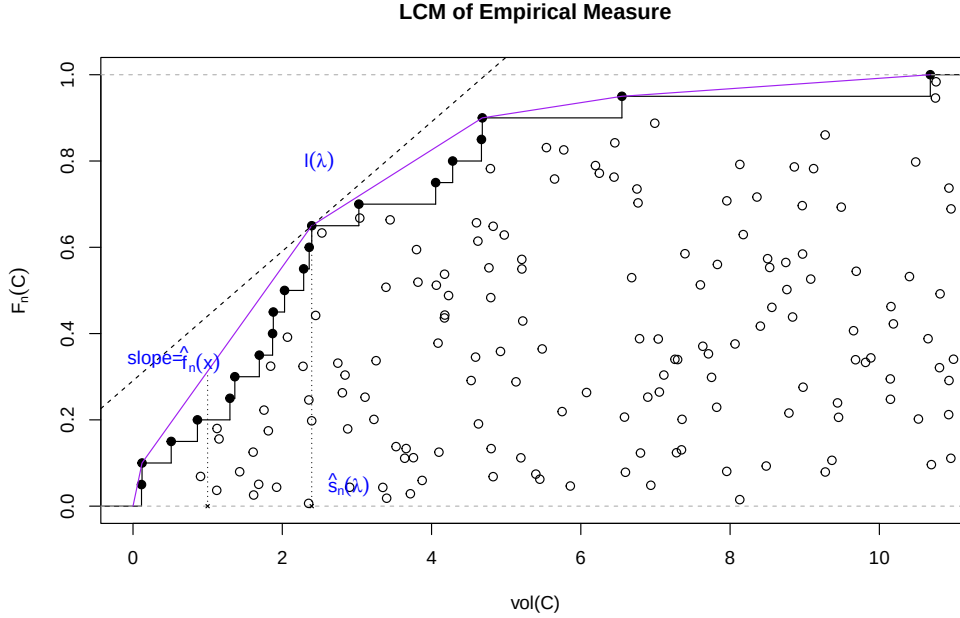


FIGURE 3.4. Illustration of generalized switching relation (3.2.13). $\hat{f}_{n, \preceq}(x) \geq \lambda$. The dotted line $l(\lambda)$ has slope λ . $\hat{s}_{n, \mathcal{C}_{\preceq}}^*(\lambda)$ is the x -coordinate where the straight line $l(\lambda)$ with slope λ is tangent to the least concave majorant of the empirical cdf.

Let $\ell_{n, x} := \text{Leb}(\Gamma_{\hat{f}_{n, \preceq}}(\hat{f}_{n, \preceq}(x)))$, we can use Proposition 3.2.1 to obtain:

THEOREM 3.2.1 (Generalized Switching Relation for Order Restricted NPMLEs). *For any quasi-order ' $\preceq$ ' on $\mathbb{R}^d$ and the corresponding σ -lattice $\mathcal{C}_{\preceq}$ the order restricted NPMLE $\hat{f}_{n, \preceq}(x)$ satisfies the switching relation*

$$(3.2.13) \quad [\hat{f}_{n, \preceq}(x) \geq \lambda] \iff [\hat{s}_{n, \mathcal{C}_{\preceq}}^*(\lambda) \geq \ell_{n, x}].$$

Notice that the reason why the generalized switching relation is helpful is that the r.h.s. only depends on $\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell)$. Also in case of the Grenander estimator, where $\mathcal{C}_{\preceq} = \mathcal{C}_{\leq}$, we have $\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell) = F_n(\ell)$ and we obtain the classical switching relation.

PROOF. The result follows directly from Proposition 3.2.1 if we have shown that the following equality is true:

$$(3.2.14) \quad \hat{s}_{n,\mathcal{C}_{\preceq}}^*(\lambda) = \text{Leb}(\Gamma_{\hat{f}_{n,\preceq}}(\lambda)).$$

From the definition in (3.2.12), $\hat{s}_{n,\mathcal{C}_{\preceq}}^*(\lambda)$ is the largest ℓ such that there exists a line $(\lambda\ell + b$ for some $b)$ with slope λ that is tangent to the concave majorant $\tilde{F}_{n,\mathcal{C}_{\preceq}}^*(\ell)$ (see Figure 3.2.12). From the graphical representation described in Section 3.2.4.2, $\hat{s}_{n,\mathcal{C}_{\preceq}}^*(\lambda)$ can only take values when the slope of the least concave majorant $\tilde{F}_{n,\mathcal{C}_{\preceq}}^*(\ell)$ changes, which is exactly $\text{Leb}(\Gamma_{\hat{f}_{n,\preceq}}(\lambda))$. □

We first state a lemma for the concentration function defined in Equation (3.2.11).

LEMMA 3.2.1 (Consistency of Concentration Function). *For any class $\mathcal{C}$, we have*

$$(3.2.15) \quad \sup_{\ell \geq 0} |\tilde{F}_{n,\mathcal{C}}(\ell) - \tilde{F}_{\mathcal{C}}(\ell)| \leq \|F_n - F\|_{\mathcal{C}}.$$

Hence, if $\mathcal{C}$ is a GC-class for F , then, with probability 1,

$$(3.2.16) \quad \sup_{\ell \geq 0} |\tilde{F}_{n,\mathcal{C}}(\ell) - \tilde{F}_{\mathcal{C}}(\ell)| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

PROOF.

$$\begin{aligned} |\tilde{F}_{n,\mathcal{C}}(\ell) - \tilde{F}_{\mathcal{C}}(\ell)| &= \left| \sup_{\substack{\text{Leb}(C) \leq \ell \\ C \in \mathcal{C}}} F_n(C) - \sup_{\substack{\text{Leb}(C) \leq \ell \\ C \in \mathcal{C}}} F(C) \right| \\ &\leq \sup_{\substack{\text{Leb}(C) \leq \ell \\ C \in \mathcal{C}}} |F_n(C) - F(C)| \\ &\leq \sup_{C \in \mathcal{C}} |F_n(C) - F(C)| \\ &= \|F_n - F\|_{\mathcal{C}}. \end{aligned}$$

□

In the next theorem, we discuss how to control the quantity $\ell_{n,x}$.

THEOREM 3.2.2 (Property of $\ell_{n,x}$). *Let $\Lambda \subset [0, \infty)$ be a closed neighborhood of $f(x)$. Suppose the following assumptions hold:*

- i. $\mathcal{C}_{\preceq}$ is an n -deviation measurable (v, m) constructible VC-class,
- ii. there exist constants $\gamma > 0$, $c \geq 0$ such that, for all η small enough, $\sup_{\lambda \in \Lambda} \text{Leb}(\{x : |f(x) - \lambda| < \eta\}) \leq c\eta^\gamma$,
- iii. $|\hat{f}_{n,\preceq}(x) - f(x)| = o_p(1)$.

Let $\ell_x = \text{Leb}(\Gamma_f(f(x)))$, then

$$(3.2.17) \quad \ell_{n,x} \xrightarrow{P} \ell_x \text{ as } n \rightarrow \infty.$$

PROOF. For convenience, write $\Gamma_n(\cdot) = \Gamma_{\hat{f}_{n,\preceq}}(\cdot)$, $\Gamma(\cdot) = \Gamma_f(\cdot)$, $\lambda_n = \hat{f}_{n,\preceq}(x)$, and $\lambda = f(x)$.

$$\begin{aligned} |\ell_{n,x} - \ell_x| &= |\text{Leb}(\Gamma_n(\lambda_n)) - \text{Leb}(\Gamma(\lambda))| \\ &\leq |\text{Leb}(\Gamma_n(\lambda_n)) - \text{Leb}(\Gamma(\lambda_n))| + |\text{Leb}(\Gamma(\lambda_n)) - \text{Leb}(\Gamma(\lambda))| \\ &= \left| \int_{\Gamma_n(\lambda_n) \setminus \Gamma(\lambda_n)} ds \right| + |\text{Leb}(\Gamma(\lambda_n)) - \text{Leb}(\Gamma(\lambda))| \\ &\leq \int_{\Gamma_n(\lambda_n) \Delta \Gamma(\lambda_n)} ds + |\text{Leb}(\Gamma(\lambda_n)) - \text{Leb}(\Gamma(\lambda))| \\ &= \text{Leb}(\Gamma_n(\lambda_n) \Delta \Gamma(\lambda_n)) + |\text{Leb}(\Gamma(\lambda_n)) - \text{Leb}(\Gamma(\lambda))| \end{aligned}$$

Now we write

$$\begin{aligned} I &= \text{Leb}(\Gamma_n(\lambda_n) \Delta \Gamma(\lambda_n)) \\ II &= |\text{Leb}(\Gamma(\lambda_n)) - \text{Leb}(\Gamma(\lambda))|. \end{aligned}$$

For I , by Theorem 3.6 of Polonik (1995b), assumptions (i, ii) imply there exists a constant $K = K(m, c, \gamma, \mathcal{C}_{\preceq})$ such that

$$(3.2.18) \quad P^* \left(\sup_{\lambda \in \Lambda} \text{Leb}(\Gamma(\lambda) \Delta \Gamma_n(\lambda)) > K \left(\frac{n}{\log n} \right)^{-\gamma/(2+\gamma)} \right) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

For II , by assumption (ii) , $|\text{Leb}(\Gamma(\lambda_n) \Delta \Gamma(\lambda))| = O(|\hat{f}_{n,\preceq}(x) - f(x)|^\gamma)$.

□

3.3. Asymptotic Results for Nonparametric Maximum-Likelihood Estimators (NPMLEs) Under Order Restrictions

In this section, we illustrate how to study the behavior of NPMLEs under order restrictions using the generalized switching relation by some examples. The results we are able to get usually depend on how well we are able to control the chosen $\ell_{n,x}$. In Section 3.1 we talk about a simple case in which ‘ $\preceq$ ’ is a total order. In this scenario, the asymptotic distributional results can be derived. Note that this essentially is just a revisit of the proof of the Grenander estimator (see Groeneboom and Jongbloed (2014)) under our framework.

3.3.1. Pointwise Limit Distribution for NPMLEs Under a Total Order Restriction.

Below, we use an example when $\mathcal{C}_{\preceq}$ is a class of nested ellipsoids of a given shape to illustrate how generalized switching relation is used to derive the pointwise limit distribution of $\hat{f}_{n,\preceq}(x)$. We need some more definitions. Let A be a positive-definite $d \times d$ matrix. Let

$$(3.3.1) \quad Q_{x_0,A,t} = \{u : (u - x_0)' A (u - x_0) \leq t\},$$

$$(3.3.2) \quad Q_{x_0,A,t}^o = \{u : (u - x_0)' A (u - x_0) < t\},$$

and

$$(3.3.3) \quad \mathcal{Q}_{x_0,A} = \{Q_{x_0,A,t}, Q_{x_0,A,t}^o : 0 \leq t \leq T\}$$

Let ‘ $\preceq$ ’ be the total order on $\mathcal{Q}_{x_0,A,T}$ defined by

$$(3.3.4) \quad Q_1 \preceq Q_2 \iff Q_1 \subseteq Q_2.$$

Without loss of generality, let $x_0 = 0$. With a fixed positive-definite $d \times d$ matrix A , write $\mathcal{Q} = \mathcal{Q}_{0,A}$, with elements $Q_t = \{u : u' A u \leq t\}$, $0 \leq t \leq T$. Let $X_1, \dots, X_n$ be iid from a population with density function f and distribution function F . Then f being monotonically decreasing with respect to ' $\preceq$ ' is equivalent to that f has a shape constraint with respect to $\mathcal{C}_{\preceq} = \mathcal{Q}$. Such distributions are called elliptical distributions in the literature. Due to the fact that ' $\preceq$ ' is a total order, $\mu(D)$ in (3.2.5) can be defined as

- (1) $\ell_{n,x}^{(1)} = \ell_x = \text{Leb}(Q_{x'Ax})$,
- (2) $\ell_{n,x}^{(2)} = \text{Leb}(Q_{\hat{t}_x^{max}})$, where $\hat{t}_x^{max} = \max\{X_i' A X_i \leq x' A x, i = 1, \dots, n\}$,
- (3) $\ell_{n,x}^{(3)} = \text{Leb}(Q_{\hat{t}_x^{min}})$, where $\hat{t}_x^{min} = \min\{X_i' A X_i \geq x' A x, i = 1, \dots, n\}$.

For almost all λ ,

$$(3.3.5) \quad \hat{f}_{n,\preceq}(x) \leq \lambda \iff \hat{s}_{n,\mathcal{C}_{\preceq}}^*(\lambda) \leq \ell_{n,x}^{(i)}, \quad \forall i = 1, 2, 3.$$

It's easy to calculate that all the above choices of $\ell_{n,x}$ converge to ℓ_x in probability as $n \rightarrow \infty$ with some regularity conditions.

3.3.1.1. Uniform Case.

THEOREM 3.3.1 (Pointwise Limit Distribution of NPMLEs Under a Total Order Restriction for Uniform Distributions). *Suppose $F = \mathcal{U}(Q_T)$ is the uniform distribution defined on Q_T where $\text{Leb}(Q_T) = 1$. $\hat{f}_{n,\preceq}$ is the NPMLE of the density f under the total order restriction defined as previously. Then*

$$(3.3.6) \quad P(\sqrt{n}(\hat{f}_{n,\preceq}(x) - f(x)) \leq t) \rightarrow P\left(\arg\max_{\ell \geq 0} \{B(\ell) - t\ell\} \leq \ell_x\right),$$

as $n \rightarrow \infty$, where B is a standard Brownian bridge on the unit interval.

PROOF. Suppose $F = \mathcal{U}(Q_T)$, where $\text{Leb}(Q_T) = 1$. Then $f(x) = 1$ for all $x \in Q_T$ and $l = \tilde{F}_{\mathcal{C}_{\preceq}}(l)$.

Then

$$\begin{aligned}
& \{\alpha_n^{-1}(\hat{f}_{n,\preceq}(x) - f(x)) \leq t\} \\
&= \{\hat{f}_{n,\preceq}(x) \leq f(x) + \alpha_n t\} \\
&= \{\hat{s}_{n,\mathcal{C}_{\preceq}}^*(f(x) + \alpha_n t) \leq \ell_x\} \\
&= \left\{ \operatorname{argmax}_{\ell > 0} \{\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell) - (f(x) + \alpha_n t)\ell\} \leq \ell_x \right\} \\
&= \left\{ \operatorname{argmax}_{\ell > 0} \{\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell_x + \ell) - (f(x) + \alpha_n t)(\ell_x + \ell)\} \leq 0 \right\} \\
&= \left\{ \operatorname{argmax}_{\ell > 0} \{\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell_x + \ell) - \tilde{F}_{\mathcal{C}_{\preceq}}(\ell_x + \ell) - \alpha_n t\ell - \alpha_n t\ell_x\} \leq 0 \right\} \\
&= \left\{ \operatorname{argmax}_{\ell > 0} \{\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell_x + \ell) - \tilde{F}_{\mathcal{C}_{\preceq}}(\ell_x + \ell) - \alpha_n t\ell\} \leq 0 \right\} \\
&\quad \text{Let } \alpha_n = n^{-1/2}, \text{ the above formula becomes} \\
&= \left\{ \operatorname{argmax}_{\ell > 0} \{\sqrt{n}(\tilde{F}_{n,\mathcal{C}_{\preceq}}(\ell_x + \ell) - \tilde{F}_{\mathcal{C}_{\preceq}}(\ell_x + \ell)) - t\ell\} \leq 0 \right\}
\end{aligned}$$

Notice that $\alpha_n t \ell_x$ does not depend on ℓ . We can drop this term inside the ‘argmax’ operation without changing the result.

Let

$$(3.3.7) \quad \hat{C}(\ell) \in \operatorname{argmax}_{C \in \mathcal{C}_{\preceq}} \{F_n(C), \operatorname{Leb}(C) \leq \ell\},$$

$$(3.3.8) \quad C(\ell) \in \operatorname{argmax}_{C \in \mathcal{C}_{\preceq}} \{F(C), \operatorname{Leb}(C) \leq \ell\}.$$

$\mathcal{C}_{\preceq} = \mathcal{Q}$ is Donsker, so the process $\{\sqrt{n}(F_n - F)(C), C \in \mathcal{C}_{\preceq}\}$ converges to $B(F(C))$ where B is a standard Brownian bridge on the unit interval. Again, due to the fact that we have a total ordering, $\hat{C}(\ell)$ and $C(\ell)$ can be chosen to be $Q_{V^{-1}(\ell)}$. The process $\{\sqrt{n}(\tilde{F}_n - \tilde{F})(\ell) = \sqrt{n}(F_n - F)(Q_{V^{-1}(\ell)}), 0 \leq t \leq T\}$, converges to $B(F(Q_{V^{-1}(\ell)})) = B(\ell)$.

Now we can apply argmax continuous theorem, i.e., Theorem 3.5.1 to show

$$(3.3.9) \quad \operatorname{argmax}_{\ell > 0} \left\{ \sqrt{n}(\tilde{F}_{n, \mathcal{C}_{\leq}}(\ell_x + \ell) - \tilde{F}_{\mathcal{C}_{\leq}}(\ell_x + \ell)) - t\ell \right\} \xrightarrow{D} \operatorname{argmax}_{\ell > 0} \left\{ B(\ell_x + \ell) - t\ell \right\}.$$

The conditions for the theorem hold obviously. Let $\hat{\ell}_n$ be the maximizer of $\ell \rightarrow \sqrt{n}(\tilde{F}_{n, \mathcal{C}_{\leq}}(\ell_x + \ell) - \tilde{F}_{\mathcal{C}_{\leq}}(\ell_x + \ell)) - t\ell$ in $[0, K]$ for some large enough K . Then $\hat{\ell}_n = O_P(1)$ is uniformly tight. $\square$

3.3.1.2. Non-uniform Case.

THEOREM 3.3.2 (Pointwise Limit Distribution of NPMLEs Under a Total Order Restriction for Non-uniform Distributions). *Suppose there is no flat part in a neighborhood of the level curve $\partial\Gamma_f(f(x))$ for the density function f . Let $\delta_n = \sqrt[3]{\frac{4f(x)|h(x)|}{n}}$, where $h(x) = \frac{d^2\tilde{F}(\ell_x)}{d\ell^2}$. Let $\hat{f}_{n, \leq}$ be the NPMLE of the density f under the total order restriction defined as above. Suppose that $h(x) \neq 0$. Then*

$$(3.3.10) \quad P\left(\delta_n^{-1}(\hat{f}_{n, \leq}(x) - f(x)) \leq t\right) \rightarrow P\left(\operatorname{argmax}_u \left\{ W(u) - (u + t)^2 \right\} \leq 0\right),$$

as $n \rightarrow \infty$, where W is the standard two-sided Brownian motion.

PROOF. Let $\ell_{n,x}$ be any one of the first three options in the above.

We use the fact that $\tilde{F}_n(\ell) = F_n(C(\ell))$. With probability 1, we have the relation

$$(3.3.11) \quad \hat{f}_{n, \mathcal{C}_{\leq}}(x) \leq \lambda \iff \hat{s}_{n, \mathcal{C}_{\leq}}(\lambda) \leq \ell_x.$$

Let α_n be a sequence of positive numbers that converges to 0.

$$\begin{aligned}
& P\left(\alpha_n^{-1}(\hat{f}_{n,\preceq}(x) - f(x)) \leq t\right) \\
&= P\left(\arg\max_{\ell \geq 0} \{\tilde{F}_n(\ell) - (f(x) + \alpha_n t)\ell\} \leq \ell_x\right) \\
&= P\left(\arg\max_{\ell \geq 0} \{\tilde{F}_n(\ell_x + \ell) - (f(x) + \alpha_n t)(\ell_x + \ell)\} \leq 0\right) \\
&= P\left(\delta_n^{-1} \arg\max_{\ell \geq 0} \{\tilde{F}_n(\ell_x + \ell) - (f(x) + \alpha_n t)(\ell_x + \ell)\} \leq 0\right) \\
&= P\left(\arg\max_{\ell \geq 0} \{\tilde{F}_n(\ell_x + \delta_n \ell) - (f(x) + \alpha_n t)(\ell_x + \delta_n \ell)\} \leq 0\right) \\
&= P\left(\arg\max_{\ell \geq 0} \{(\tilde{F}_n(\ell_x + \delta_n \ell) - \tilde{F}_n(\ell_x)) - (\tilde{F}(\ell_x + \delta_n \ell) - \tilde{F}(\ell_x)) + (\tilde{F}(\ell_x + \delta_n \ell) - \tilde{F}(\ell_x)) - f(x)(\ell_x + \delta_n \ell) - \alpha_n t(\ell_x + \delta_n \ell)\} \leq 0\right)
\end{aligned}$$

Now, let $A_n(\ell) = (\tilde{F}_n(\ell_x + \delta_n \ell) - \tilde{F}_n(\ell_x)) - (\tilde{F}(\ell_x + \delta_n \ell) - \tilde{F}(\ell_x))$, $b_n(\ell) = (\tilde{F}(\ell_x + \delta_n \ell) - \tilde{F}(\ell_x)) - f(x)\delta_n \ell$.

To study $b_n(\ell)$ first.

$$\frac{d\tilde{F}(\ell)}{d\ell} = \lim_{h \rightarrow 0} \frac{\tilde{F}(\ell + h) - \tilde{F}(\ell)}{h} = \lim_{h \rightarrow 0} \frac{\int_{C(\ell+h)\Delta C(\ell)} f(x)dx}{\int_{C(\ell+h)\Delta C(\ell)} dx} = f(x_\ell) = \lambda_\ell,$$

where $x_\ell \in \partial C(\ell)$, and λ_ℓ is such that $C(\ell) = \Gamma_f(\lambda_\ell)$.

Let $\psi(\lambda) = \text{Leb}(\Gamma_f(\lambda))$. By Reynold's transport theorem,

$$\frac{d\psi(\lambda)}{d\lambda} = \frac{d}{d\lambda} \int_{\Gamma_f(\lambda)} dA = - \int_{f=\lambda} \frac{1}{|\nabla f|} dS.$$

So

$$\frac{d^2 \tilde{F}(\ell)}{d\ell^2} = \frac{1}{\frac{d\psi(\lambda)}{d\lambda}} = - \frac{1}{\int_{f=\lambda} \frac{1}{|\nabla f|} dS}.$$

So $b_n(\ell) \sim \frac{1}{2} \frac{d^2 \tilde{F}(\ell_x)}{d\ell^2} \ell^2 \delta_n^2 (1 + o(1))$ if $n \rightarrow \infty$.

For the random part, we use the fact that $\hat{C}(\ell) = C(\ell)$.

$$\begin{aligned}
A_n(\ell) &= (\tilde{F}_n(\ell_x + \delta_n \ell) - \tilde{F}_n(\ell_x)) - (\tilde{F}(\ell_x + \delta_n \ell) - \tilde{F}(\ell_x)) \\
&= F_n(C(\ell_x + \delta_n \ell)) - F_n(C(\ell_x)) - (F(C(\ell_x + \delta_n \ell)) - F(C(\ell_x))) \\
&= \int (\mathbb{1}_{C(\ell_x + \delta_n \ell)} - \mathbb{1}_{C(\ell_x)}) d(F_n - F)
\end{aligned}$$

By Theorem 3.5.2, with appropriate scaling factors, the above quantity converges to a Gaussian process if the variance covariance function is convergent pointwise to a uniformly bounded function, details will be provided later. We have

$$\mathbb{E} A_n(\ell) = 0,$$

and

$$\begin{aligned}
\text{var}(A_n(\ell)) &= \mathbb{E} \left(F_n(C(\ell_x + \ell\delta_n)) - F_n(C(\ell_x)) - (F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x))) \right)^2 \\
&= \mathbb{E} \left(F_n(C(\ell_x + \ell\delta_n)) - F_n(C(\ell_x)) \right)^2 \\
&\quad - 2\mathbb{E} \left(F_n(C(\ell_x + \ell\delta_n)) - F_n(C(\ell_x)) \right) \left(F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x)) \right) \\
&\quad + \left(F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x)) \right)^2 \\
&= \mathbb{E} F_n(C(\ell_x + \ell\delta_n))^2 - 2\mathbb{E} F_n(C(\ell_x + \ell\delta_n)) F_n(C(\ell_x)) + \mathbb{E} F_n(C(\ell_x))^2 \\
&\quad - \left(F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x)) \right)^2 \\
&= \frac{1}{n} F(C(\ell_x + \ell\delta_n)) + \frac{n-1}{n} F(C(\ell_x + \ell\delta_n))^2 - 2\frac{1}{n} F(C(\ell_x)) \\
&\quad - \frac{2(n-1)}{n} F(C(\ell_x + \ell\delta_n)) F(C(\ell_x)) + \frac{1}{n} F(C(\ell_x)) + \frac{n-1}{n} F(C(\ell_x))^2 \\
&\quad - \left(F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x)) \right)^2 \\
&= \frac{1}{n} \left(F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x)) \right) - \frac{1}{n} \left(F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x)) \right)^2 \\
&= \frac{1}{n} \left(F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x)) \right) \left(1 - (F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x))) \right),
\end{aligned}$$

which gives us $\frac{1}{n} \left(F(C(\ell_x + \ell\delta_n)) - F(C(\ell_x)) \right) \sim \frac{1}{n} f(x) \ell \delta_n$ as $n \rightarrow \infty$. Recall from above that $b_n(\ell)$ is of the order δ_n^2 . $\delta_n^{-2} b_n(\ell)$ stabilizes. Moreover, $\text{var}(\delta_n^{-2} A_n(\ell)) \sim \frac{f(x)\ell}{n\delta_n^3}$ and we see that δ_n should be of order $n^{-1/3}$.

$$\begin{aligned}
&P \left(\alpha_n^{-1} (\hat{f}_{n,\leq}(x) - f(x)) \leq t \right) \\
&= P \left(\operatorname{argmax}_{\ell \geq 0} \left\{ (\tilde{F}_n(\ell_x + \delta_n \ell) - \tilde{F}_n(\ell_x)) - (\tilde{F}(\ell_x + \delta_n \ell) - \tilde{F}(\ell_x)) + (\tilde{F}(\ell_x + \delta_n \ell) - \tilde{F}(\ell_x)) \right. \right. \\
&\quad \left. \left. - f(x)(\ell_x + \delta_n \ell) - \alpha_n t(\ell_x + \delta_n \ell) \right\} \leq 0 \right) \\
&= P \left(\operatorname{argmax}_{\ell \geq 0} \left\{ A_n(\ell) + b_n(\ell) - \alpha_n t \delta_n \ell \right\} \leq 0 \right) \\
&= P \left(\operatorname{argmax}_{\ell \geq 0} \left\{ \delta_n^{-2} A_n(\ell) + \delta_n^{-2} b_n(\ell) - \alpha_n \delta_n^{-1} t \ell \right\} \leq 0 \right)
\end{aligned}$$

With this choice, we obtain (see below) that the process $\{\delta_n^{-2}A_n(\ell) + \delta_n^{-2}b_n(\ell) - t\ell\}$ converges to $\sqrt{\frac{f(x)}{s^3}}W(\ell) + \frac{1}{2}\frac{d^2\tilde{F}(\ell_x)}{d\ell^2}\ell^2 - t\ell$ weakly, where $W(\ell)$ is a two-sided Brownian motion. We can further simplify the formula using Brownian scaling, namely,

$$(3.3.12) \quad t \rightarrow \alpha(W(\beta t) + (\beta t)^2) =_D t \rightarrow \alpha\sqrt{\beta}W(t) + \alpha\beta^2t^2.$$

To decide the choice of s . Write $h(x) = \frac{d^2\tilde{F}(\ell_x)}{d\ell^2}$. Let $u = |D|\ell$.

$$\begin{aligned} |D| \operatorname{argmax}_{\ell \geq 0} \left\{ \sqrt{\frac{f(x)}{s^3}}W(\ell) + \frac{1}{2}h(x)\ell^2 - t\ell \right\} &= \operatorname{argmax}_{u \geq 0} \left\{ \sqrt{\frac{f(x)}{s^3}}W\left(\frac{u}{|D|}\right) + \frac{1}{2}h(x)\frac{u^2}{D^2} - \frac{tu}{|D|} \right\} \\ &= \operatorname{argmax}_{u \geq 0} \left\{ \sqrt{\frac{f(x)}{s^3|D|}}W(u) + \frac{1}{2}h(x)\frac{u^2}{D^2} - \frac{tu}{|D|} \right\} \end{aligned}$$

Let

$$\sqrt{\frac{f(x)}{s^3|D|}} = \frac{1}{|D|},$$

$$\frac{h(x)}{2D^2} = \frac{1}{|D|}.$$

So we can choose $D = |h(x)|$, $s = \sqrt[3]{4f(x)|h(x)|}$. Putting everything together, and recalling that

$$\begin{aligned} &P\left(\alpha_n^{-1}(\hat{f}_{n,\leq}(x) - f(x)) \leq t\right) \\ &= P\left(\operatorname{argmax}_{\ell \geq 0} \{A_n(\ell) + b_n(\ell) - \alpha_n t \delta_n \ell\} \leq 0\right) \\ &= P\left(\operatorname{argmax}_{\ell \geq 0} \{\delta_n^{-2}A_n(\ell) + \delta_n^{-2}b_n(\ell) - \alpha_n \delta_n^{-1}t\ell\} \leq 0\right), \end{aligned}$$

We obtain with $\delta_n = n^{-\frac{1}{3}}$ that

$$\begin{aligned}
& P\left(\sqrt[3]{\frac{n}{4f(x)|h(x)|}}(\hat{f}_{n,\preceq}(x) - f(x)) \leq t\right) \\
& \rightarrow P\left(|D| \operatorname{argmax}_{u \geq 0} \left\{W(u) - u^2 - t\right\} \leq 0\right) \\
& = P\left(\operatorname{argmax}_{u \geq 0} \left\{W(u) - (u+t)^2\right\} \leq 0\right).
\end{aligned}$$

Now, let's establish the convergence of $\delta_n^{-2}A_n(\ell)$ in $\ell^\infty([-K, K])$ for $K > 0$ with $\delta_n = n^{-1/3}$ rigorously. To this end, we use Theorem 3.5.2. We need to check validity of all the conditions (3.5.1)~(3.5.4) in this theorem. Let $\mathcal{G}_n = \{g_{n,\ell} = n^{1/6}\mathbb{1}_{C(\ell_x + \ell n^{-1/3})} - n^{1/6}\mathbb{1}_{C(\ell_x)}, \ell \in [-K, K]\}$. The envelope function $G_{n,en} = \sup_{|\ell| \leq K} |g_{n,\ell}| = n^{1/6}\mathbb{1}_{C(\ell_x + Kn^{-1/3}) \setminus C(\ell_x)}$.

Suppose $f(x) > 0$, we can find $L, R > 0$ such that $f(z) \in [L, R]$ for $z \in C(\ell_x + \ell_b) \setminus C(\ell_x - \ell_b)$ with a small enough ℓ_b . Then with n large,

$$\begin{aligned}
\|G_{n,en}\|_{2,f}^2 &= \int n^{1/3}\mathbb{1}_{C(\ell_x + Kn^{-1/3}) \setminus C(\ell_x)}(z)f(z)dz \\
&\leq \int n^{1/3}\mathbb{1}_{C(\ell_x + Kn^{-1/3}) \setminus C(\ell_x)}(z)Rdz \\
&= n^{1/3}Kn^{-1/3}R \\
&= KR.
\end{aligned}$$

This is (3.5.1). Next, we show (3.5.2). For every $\eta > 0$,

$$\begin{aligned}
\int_{\{z: G_{n,en}(z) > \eta\sqrt{n}\}} G_{n,en}(z)^2 f(z) dz &= \int_{\{z: G_{n,en}(z) > \eta\sqrt{n}\}} n^{1/3}\mathbb{1}_{C(\ell_x + Kn^{-1/3}) \setminus C(\ell_x)}(z)f(z)dz \\
&\rightarrow 0 \text{ as } n \rightarrow \infty,
\end{aligned}$$

since when n is large enough, $\sqrt{n}\eta > n^{1/6}$. When $\xi_n \downarrow 0$,

$$\begin{aligned}
& \sup_{\{k,j \in [-K,K]: |k-j| < \xi_n\}} \int (g_{n,k}(z) - g_{n,j}(z))^2 f(z) dz \\
&= \sup_{\{k,j \in [-K,K]: |k-j| < \xi_n\}} \int n^{1/3} (\mathbb{1}_{C(\ell_x + kn^{-1/3})}(z) - \mathbb{1}_{C(\ell_x + jn^{-1/3})}(z))^2 f(z) dz \\
&\leq n^{1/3} \xi_n n^{-1/3} R \quad \text{for } n \text{ large enough} \\
&\rightarrow 0 \text{ as } n \rightarrow \infty.
\end{aligned}$$

To verify (3.5.3), we first show that the bracketing number,

$$(3.3.13) \quad N_{[]}(\epsilon \|G_{n,en}\|_{2,f}, \mathcal{G}_n, \|\cdot\|_{2,f}) \lesssim \frac{1}{\epsilon^2},$$

moreover, there exists a constant C that is independent of n such that

$$(3.3.14) \quad N_{[]}(\epsilon \|G_{n,en}\|_{2,f}, \mathcal{G}_n, \|\cdot\|_{2,f}) \leq \frac{C}{\epsilon^2}.$$

Suppose $f(x) > 0$, we can find $L, R > 0$ such that $f(z) \in [L, R]$ for $z \in C(\ell_x + \ell_b) \setminus C(\ell_x - \ell_b)$ with a small enough ℓ_b . Then with n large,

$$\sqrt{KL} \leq \|G_{n,en}\|_{2,f} = \left(\int n^{1/3} \mathbb{1}_{C(\ell_x + Kn^{-1/3}) \setminus C(\ell_x)} f(z) dz \right)^{1/2} \leq \sqrt{KR}.$$

So

$$(3.3.15) \quad N_{[]}(\epsilon \|G_{n,en}\|_{2,f}, \mathcal{G}_n, \|\cdot\|_{2,f}) \leq N_{[]}(\epsilon \sqrt{KL}, \mathcal{G}_n, \|\cdot\|_{2,f}).$$

Consider function brackets of the form $[g_i, g_j] = [n^{1/6} \mathbb{1}_{C(\ell_x + \ell_i)} - n^{1/6} \mathbb{1}_{C(\ell_x)}, n^{1/6} \mathbb{1}_{C(\ell_x + \ell_j)} - n^{1/6} \mathbb{1}_{C(\ell_x)}]$ such that

$$\begin{aligned}
\|g_j - g_i\|_{2,f} &= \left(\int (g_j - g_i)(z)^2 f(z) dz \right)^{1/2} \\
&= \left(\int n^{1/3} \mathbb{1}_{C(\ell_x + \ell_j) \setminus C(\ell_x + \ell_i)}(z) f(z) dz \right)^{1/2} \\
&\leq \sqrt{n^{1/3}(\ell_j - \ell_i)R} \\
&< \epsilon \sqrt{KL}.
\end{aligned}$$

This implies (3.5.3). Finally, we show (3.5.4). Then $\ell_j - \ell_i < \epsilon^2 n^{-1/3} KL/R$. In order to cover $\mathcal{G}_n$, we can to construct a sequence $-K = \ell_0 \leq \ell_1 \leq \dots \leq \ell_N = K$ such that $\ell_i - \ell_{i-1} < \epsilon^2 n^{-1/3} KL/R$ for all $i \in \{1, \dots, N\}$. It can be easily constructed such that $N \leq 2 \times 2Kn^{-1/3}/(\epsilon^2 n^{-1/3} KL/R) = 4R/(\epsilon^2 L)$ by letting $\ell_i - \ell_{i-1} = 1/2 \times \epsilon^2 n^{-1/3} KL/R$ for $i \in \{1, \dots, N-1\}$ and $\ell_N - \ell_{N-1} \leq 1/2 \times \epsilon^2 n^{-1/3} KL/R$. The desired inequality just follows with C being $4R/L$.

$$\begin{aligned}
&\int_0^{\xi_n} \sqrt{\log N_{[]}(\epsilon \|G_{n,en}\|_{2,f}, \mathcal{G}_n, \|\cdot\|_{2,f})} d\epsilon \\
&\leq \int_0^{\xi_n} \sqrt{\log(C/\epsilon^2)} d\epsilon \\
&\rightarrow 0 \text{ as } n \rightarrow \infty
\end{aligned}$$

So by Theorem 3.5.2, $\delta_n^{-2} A_n(\ell)$ converges in $\ell^\infty([-K, K])$ to a Gaussian process, since the choice of scaling factors have guaranteed that the sequence of covariance functions converges pointwise on $[-K, K]^2$.

Now, we need to apply the argmax continuous theorem (Theorem 3.5.1) to show that

$$(3.3.16) \quad \operatorname{argmax}_{\ell \geq 0} \{ \delta_n^{-2} A_n(\ell) + \delta_n^{-2} b_n(\ell) - \alpha_n \delta_n^{-1} t \ell \} \xrightarrow{D} |D| \operatorname{argmax}_{u \geq 0} \{ W(u) - u^2 - t \}.$$

Let

$$(3.3.17) \quad IM_n(\ell) = F_n(C(\ell_x + \ell)) - F_n(C(\ell_x)) - (f(x) + \alpha_n t)\ell,$$

$$(3.3.18) \quad IM(\ell) = F(C(\ell_x + \ell)) - F(C(\ell_x)) - f(x)\ell.$$

Suppose that $\hat{\ell}_n$ is the maximizer of $\ell \mapsto IM_n(\ell)$, then we want $\delta_n^{-1}\hat{\ell}_n$ to be bounded in probability, and the sequence $\{\delta_n^{-1}\hat{\ell}_n\}$ uniformly tight in order to apply the theorem. $IM_n(\hat{\ell}_n) \geq \sup_n IM_n(\ell) - o_P(1)$ is guaranteed in definition of $\hat{\ell}_n$. To check these conditions, we can apply Theorem 3.5.3. First by Taylor, we have

$$(3.3.19) \quad IM(\ell) = \frac{1}{2}h(x)\ell^2 + o(\ell^2), \text{ as } \ell \rightarrow 0.$$

So

$$(3.3.20) \quad IM(\ell) - IM(0) = \frac{1}{2}h(x)\ell^2 + o(\ell^2) \lesssim |\ell - 0|^2.$$

For δ sufficiently small, $f(z) \in [L, R]$ for all $z \in C(\ell_x + \delta) \setminus C(\ell_x - \delta)$. Consider the class of functions $\mathcal{G}' = \{g'_\ell = \mathbb{1}_{C(\ell_x + \ell)} - \mathbb{1}_{C(\ell_x)}, |\ell| < \delta\}$. $G'_{en}(z) = \sup_{|\ell| < \delta} \{\mathbb{1}_{C(\ell_x + \ell)} - \mathbb{1}_{C(\ell_x)}\} = \mathbb{1}_{C(\ell_x + \delta) \setminus C(\ell_x)}$ is its measurable natural envelope and $\|G'_{en}\|_{2,f} \leq \sqrt{\delta R}$. By an almost the same argument as above, $N_{[]}(\epsilon \|G'_{en}\|_{2,f}, \mathcal{G}', \|\cdot\|_{2,f}) \leq C/\epsilon^2$.

$$\begin{aligned} J_{[]} (1, \mathcal{G}', \|\cdot\|_{2,f}) &= \int_0^1 \sqrt{1 + \log N_{[]}(\epsilon \|G'_{en}\|_{2,f}, \mathcal{G}', \|\cdot\|_{2,f})} d\epsilon \\ &\leq \int_0^1 \sqrt{1 + \log(C/\epsilon^2)} d\epsilon \\ &= C'. \end{aligned}$$

By Theorem 3.5.4,

$$(3.3.21) \quad \mathbb{E}^* \sup_{g \in \mathcal{G}'} \left| \sqrt{n} \int g(z) d(F_n - F)(z) \right| \lesssim J_{[]} (1, \mathcal{G}', \|\cdot\|_{2,f}) \|G'_{en}\|_{2,f} \lesssim \sqrt{\delta}.$$

Therefore,

$$\begin{aligned} & \mathbb{E}^* \sup_{|\ell| < \delta} |(IM_n - IM)(\ell) - (IM_n - IM)(0)| \\ &= \mathbb{E}^* \sup_{|\ell| < \delta} \left| \int (\mathbb{1}_{C(\ell_x + \ell)} - \mathbb{1}_{C(\ell_x)}) d(F_n - F) + \alpha_n t \ell \right| \\ &\leq \mathbb{E}^* \sup_{g \in \mathcal{G}'} \left| \int g(z) d(F_n - F) \right| + \sup_{|\ell| < \delta} |\alpha_n t \ell| \\ &\lesssim \sqrt{\delta/n} + n^{-1/3} \delta \\ &= \frac{\sqrt{\delta} + \delta n^{1/6}}{\sqrt{n}}. \end{aligned}$$

Take $\phi_n(\delta) = \sqrt{\delta} + \delta n^{1/6}$, $r_n = n^{1/3}$, we have

$$(3.3.22) \quad r_n^2 \phi_n\left(\frac{1}{r_n}\right) = n^{2/3}(n^{-1/6} + n^{-1/3} n^{1/6}) = \sqrt{n}.$$

Moreover, we have $IM_n(\hat{\ell}_n) \geq 0 = IM_n(0)$. Lastly, we want to show $\hat{\ell}_n \xrightarrow{P} 0$.

For x fixed, suppose $f(x) > 0$. $f(x) + \alpha_n t > 0$ for large enough n . Then $IM_n(\ell)$ goes to $-\infty$ as $\ell \rightarrow \infty$. So $\hat{\ell}_n$ is bounded for each x and for large enough n .

$$\begin{aligned} P(|\hat{\ell}_n| > \epsilon) &\leq P\left(\sup_{|\ell| > \epsilon} IM_n(\ell) \geq 0 = IM_n(0)\right) \\ &= P\left(\sup_{|\ell| > \epsilon} (F_n(C(\ell_x + \ell)) - F_n(C(\ell_x)) - (f(x) + \alpha_n t)\ell) \geq 0\right) \end{aligned}$$

But $\|F_n - F\|_{\mathcal{Q}} = O_P(n^{-1/2})$, we have for $|\ell| > \epsilon > 0$, when ϵ is sufficiently small and $n \rightarrow \infty$

$$\begin{aligned}
F_n(C(\ell_x + \ell)) - F_n(C(\ell_x)) - (f(x) + \alpha_n t)\ell &= F(C(\ell_x + \ell)) - F(C(\ell_x)) - (f(x) + \alpha_n t)\ell + O_P(n^{-1/2}) \\
&= \frac{1}{2}h(x)\ell^2(1 + o(1)) - \alpha_n t + O_P(n^{-1/2}) \\
&\leq \frac{1}{2}h(x)\epsilon^2(1 + o(1)) - \alpha_n t + O_P(n^{-1/2}) \\
&\leq \frac{1}{4}h(x)\epsilon^2 - \alpha_n t + O_P(n^{-1/2})
\end{aligned}$$

Notice $h(x) < 0$. Therefore $\sup_{|\ell| > \epsilon} (F_n(C(\ell_x + \ell)) - F_n(C(\ell_x)) - (f(x) + \alpha_n t)\ell) < 0$ when $n \rightarrow \infty$, hence, $P(|\hat{\ell}_n| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$.

By Theorem 3.5.3, we have $\hat{\ell}_n = O_P(n^{-1/3})$. So the uniform tightness of $n^{1/3}\hat{\ell}_n$ is established. By Theorem 3.5.1, $\hat{\ell}_n \xrightarrow{D} \hat{\ell}$, where $|D|\hat{\ell}$ is the maximizer of the process $u \rightarrow W(u) - u^2 - t$.

□

3.4. Pointwise Convergence Rate of NPMLE Under Block-decreasing Order Restriction

In this section, we present a pointwise consistency result of the NPMLE under block-decreasing order restriction. Unfortunately, an attempt to show this consistency via the switching relation was not successful. However, we found the following simple proof using the known L1-consistency.

Let ‘ $\preceq$ ’ be such that $\forall x, y \in [0, \infty)^d$

$$(3.4.1) \quad x \preceq y \iff x_i \leq y_i \text{ for all } i \in \{1, \dots, d\}.$$

Functions that are monotonic with respect to the above ‘ $\preceq$ ’ are called block-decreasing functions. The special case when $d = 1$ is the usual monotonic decreasing function. Polonik (1998) calls it doubly monotone when $d = 2$.

PROPOSITION 3.4.1. *Let $X_1, \dots, X_n$ be iid with df F and density function f with support S . Let $\hat{f}_n$ be the NPMLE under the block-decreasing order restriction. Suppose that $\|\hat{f}_n - f\|_1 = O_p(\alpha_n)$. For any $x \in S$, let $A_x = \{y : x \preceq y\} \cap S$ and $B_x = \{y : y \preceq x\} \cap S$. If*

- (1) $\text{Vol}(A_x) > 0$ and $\text{Vol}(B_x) > 0$,

(2) $f(y) = f(x)$ for all $y \in A_x \cup B_x$,

then

$$|\hat{f}_n(x) - f(x)| = O_p(\alpha_n).$$

PROOF. For any $\epsilon > 0$, there exists $C > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$,

$$(3.4.2) \quad P\left(\alpha_n^{-1} \left\| \hat{f}_n - f \right\|_1 > 2C \min(\text{Vol}(A_x), \text{Vol}(B_x))\right) < \frac{\epsilon}{2},$$

since $\left\| \hat{f}_n - f \right\|_1 = O_p(\alpha_n)$.

For all $n \geq N$,

(3.4.3)

$$\begin{aligned} & P\left(\alpha_n^{-1}(\hat{f}_n(x) - f(x)) < -C\right) \\ & \leq P\left(\alpha_n^{-1} \int_{A_x} (\hat{f}_n(y) - f(y)) dy < -C \text{Vol}(A_x)\right) \text{ since } \forall y \in A_x, \hat{f}_n(x) \geq \hat{f}_n(y) \text{ and } f(y) = f(x) \text{ for all } y \in A_x \\ & \leq P\left(\alpha_n^{-1} \frac{1}{2} \left\| \hat{f}_n - f \right\|_1 > C \text{Vol}(A_x)\right) \text{ since } TV(\hat{f}_n - f) = \frac{1}{2} \left\| \hat{f}_n - f \right\|_1, \text{ where } TV(g) = \sup_{A \subseteq S} \left| \int_A g dy \right|. \\ & \leq P\left(\alpha_n^{-1} \left\| \hat{f}_n - f \right\|_1 > 2C \min(\text{Vol}(A_x), \text{Vol}(B_x))\right) \\ & < \frac{\epsilon}{2}. \end{aligned}$$

The proof of $P\left(\alpha_n^{-1}(\hat{f}_n(x) - f(x)) > C\right) < \frac{\epsilon}{2}$ is essentially the same as (3.4.3) but with A_x being replaced by B_x . So

$$\begin{aligned} & P\left(\alpha_n^{-1} |\hat{f}_n(x) - f(x)| > C\right) \\ (3.4.4) \quad & \leq P\left(\alpha_n^{-1}(\hat{f}_n(x) - f(x)) < -C\right) + P\left(\alpha_n^{-1}(\hat{f}_n(x) - f(x)) > C\right) \\ & < \epsilon. \end{aligned}$$

This proves $|\hat{f}_n(x) - f(x)| = O_p(\alpha_n)$.

□

REMARK 3.4.1. *By Polonik (1995a) and Gao and Wellner (2007),*

$$(3.4.5) \quad \left\| \hat{f}_n - f \right\|_1 = \begin{cases} O_p(n^{-1/4} \log n) & \text{if } d = 2 \\ O_p(n^{-1/2d}) & \text{if } d > 2 \end{cases}.$$

Therefore by the above proposition, for all x that satisfies conditions (1) and (2),

$$(3.4.6) \quad |\hat{f}_n(x) - f(x)| = \begin{cases} O_p(n^{-1/4} \log n) & \text{if } d = 2 \\ O_p(n^{-1/2d}) & \text{if } d > 2 \end{cases}.$$

3.5. Appendix: Cited Theorems

THEOREM 3.5.1 (Argmax Continuous Mapping, Theorem 3.1 in Groeneboom and Jongbloed (2014), Theorem 3.2.2 in van der Vaart and Wellner (1996)). *Let Z_n and Z be stochastic processes indexed by $\mathbb{R}$ such that $Z_n \rightsquigarrow Z$ in $\ell^\infty([-K, K])$ for each $K > 0$. Moreover, suppose that almost all sample paths of Z are continuous and have a unique maximizer $\hat{h}$, assumed to be bounded in probability. If the sequence $\hat{h}_n$ is uniformly tight, and satisfies $Z_n(\hat{h}_n) \geq \sup_h Z_n(h) - o_P(1)$, then $\hat{h}_n \xrightarrow{D} \hat{h}$ in $\mathbb{R}$.*

THEOREM 3.5.2 (Classes of Functions Changing with n , Theorem 3.2 in Groeneboom and Jongbloed (2014), Theorem 2.11.23 in van der Vaart and Wellner (1996)). *Assume F, f are the distribution function and corresponding density function respectively. F_n is the empirical distribution function. For each n , let $\mathcal{G}_n = \{g_{n,k} : h \in [-K, K]\}$ be a class of measurable functions on $\mathbb{R}$ with measurable envelope $G_{n,en}(z) = \sup_{|k| \leq K} |g_{n,k}(z)|$ satisfying the following two conditions:*

$$(3.5.1) \quad \|G_{n,en}\|_{2,f}^2 = \int G_{n,en}(z)^2 f(z) dz = O(1), \text{ as } n \rightarrow \infty,$$

and for every $\eta > 0$,

$$(3.5.2) \quad \int_{\{z: G_{n,en}(z) > \eta\sqrt{n}\}} G_{n,en}(z)^2 f(z) dz \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Moreover, let the classes $\mathcal{G}_n$ be such that for all $\xi_n \downarrow 0$,

$$(3.5.3) \quad \sup_{\{k,j \in [-K,K]: |k-j| < \xi_n\}} \int (g_{n,k}(z) - g_{n,j}(z))^2 f(z) dz \rightarrow 0 \text{ as } n \rightarrow \infty,$$

and

$$(3.5.4) \quad \int_0^{\xi_n} \sqrt{\log N_{[]}(\epsilon \|G_{n,en}\|_{2,f}, \mathcal{G}_n, \|\cdot\|_{2,f})} d\epsilon \rightarrow 0.$$

Then the sequence of stochastic processes

$$(3.5.5) \quad k \rightarrow \sqrt{n} \int g_{n,k}(z) d(F_n - F)(z)$$

converges in distribution to a Gaussian process, provided the sequence of covariance functions converges pointwise on $[-K, K]^2$.

THEOREM 3.5.3 (Establishing Rate of Convergence, Theorem 3.3 in Groeneboom and Jongbloed (2014), Theorem 3.2.5 in van der Vaart and Wellner (1996)). *Let IM_n be stochastic processes on an interval $I \subset \mathbb{R}$ and IM a deterministic function on I such that for all θ in a neighborhood of θ_0*

$$(3.5.6) \quad IM(\theta) - IM(\theta_0) \gtrsim |\theta - \theta_0|^2.$$

Moreover, suppose that, for every n and sufficiently small δ , the centered process $IM_n - IM$ satisfies

$$(3.5.7) \quad \mathbb{E}^* \sup_{|\theta - \theta_0| < \delta} |(IM_n - IM)(\theta) - (IM_n - IM)(\theta_0)| \lesssim \frac{\phi_n(\delta)}{\sqrt{n}}$$

for functions such that $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ (not depending on n). Let r_n be a sequence of positive numbers such that

$$(3.5.8) \quad r_n^2 \phi_n\left(\frac{1}{r_n}\right) \lesssim \sqrt{n} \text{ for all } n.$$

If the sequence $\hat{\theta}_n$ satisfies $IM_n(\hat{\theta}_n) \geq IM_n(\theta_0) - O_P(r_n^{-2})$ and converges to θ_0 in probability, then $\hat{\theta}_n = \theta_0 + O_P(r_n^{-1})$.

THEOREM 3.5.4 (An Integral Number Bound, Theorem 3.4 in Groeneboom and Jongbloed (2014), Theorem 2.14.2 in van der Vaart and Wellner (1996)).

$$(3.5.9) \quad E_f^* \sup_{g \in \mathcal{G}} |\sqrt{n} \int g(z) d(F_n - F)(z)| \lesssim J_{[]} (1, \mathcal{G}, \|\cdot\|_{2,f}) \|G_{n,en}\|_{2,f}.$$

Bibliography

- Aaron, C. and Bodart, O. (2016) Local convex hull support and boundary estimation. *Journal of Multivariate Analysis*, **147**, 82–101.
- Arias-Castro, E., Rodríguez-Casal, A. *et al.* (2017) On estimating the perimeter using the alpha-shape. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, vol. 53, 1051–1068. Institut Henri Poincaré.
- Biau, G., Cadre, B. and Pelletier, B. (2008) Exact rates in density support estimation. *Journal of Multivariate Analysis*, **99**, 2185–2207.
- Biau, G. and Devroye, L. (2003) On the risk of estimates for block decreasing densities. *Journal of multivariate analysis*, **86**, 143–165.
- Brunel, V.-E. (2018) Methods for estimation of convex sets. *Statistical Science*, **33**, 615–632.
- Chacón, J. E. and Duong, T. (2018) *Multivariate kernel smoothing and its applications*. CRC Press.
- Chaudhuri, K. and Dasgupta, S. (2010) Rates of convergence for the cluster tree. In *NIPS*, 343–351. Citeseer.
- Colesanti, A. and Manselli, P. (2010) Geometric and isoperimetric properties of sets of positive reach in $\mathbb{E}^d$. *Preprint*.
- Cuevas, A., Fraiman, R. and Pateiro-López, B. (2012) On statistical properties of sets fulfilling rolling-type conditions. *Advances in Applied Probability*, **44**, 311–329.
- Cuevas, A., Fraiman, R. *et al.* (1997) A plug-in approach to support estimation. *The Annals of Statistics*, **25**, 2300–2312.
- Devroye, L. and Wise, G. L. (1980) Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal on Applied Mathematics*, **38**, 480–488.
- Downs, J. A. (2010) Time-geographic density estimation for moving point objects. In *International Conference on Geographic Information Science*, 16–26. Springer.

- Dümbgen, L. and Walther, G. (1996) Rates of convergence for random approximations of convex sets. *Advances in applied probability*, 384–393.
- Federer, H. (1969) *Geometric measure theory*. Springer.
- Gao, F. and Wellner, J. A. (2007) Entropy estimate for high-dimensional monotonic functions. *Journal of Multivariate Analysis*, **98**, 1751–1764.
- Getz, W. M., Fortmann-Roe, S., Cross, P. C., Lyons, A. J., Ryan, S. J. and Wilmers, C. C. (2007) LoCoH: nonparameteric kernel methods for constructing home ranges and utilization distributions. *PloS One*, **2**, e207.
- Getz, W. M. and Wilmers, C. C. (2004) A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, **27**, 489–505.
- Grenander, U. (1956) On the theory of mortality measurement: Part II. *Scandinavian Actuarial Journal*, **1956**, 125–153.
- Groeneboom, P. (1984) Estimating a monotone density. *Department of Mathematical Statistics*.
- Groeneboom, P. and Jongbloed, G. (2014) *Nonparametric Estimation Under Shape Constraints*, vol. 38. Cambridge University Press.
- Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001a) A canonical process for estimation of convex functions: the “invelope” of integrated brownian motion+ t^4 . *Annals of Statistics*, 1620–1652.
- Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001b) Estimation of a convex function: characterizations and asymptotic theory. *The Annals of Statistics*, **29**, 1653–1698.
- Han, Q. and Zhang, C.-H. (2020) Limit distribution theory for block estimators in multiple isotonic regression. *The Annals of Statistics*, **48**, 3251–3282.
- Horne, J. S., Garton, E. O., Krone, S. M. and Lewis, J. S. (2007) Analyzing animal movements using Brownian bridges. *Ecology*, **88**, 2354–2363.
- Jiang, H. (2019) Non-asymptotic uniform rates of consistency for k-NN regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 3999–4006.
- Kranstauber, B., Kays, R., LaPoint, S. D., Wikelski, M. and Safi, K. (2012) A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, **81**, 738–746.

- Kühnel, W. (2015) *Differential Geometry*, vol. 77. American Mathematical Soc.
- Pavlidis, M. G. (2012) Local asymptotic minimax theory for block-decreasing densities. *Journal of Statistical Planning and Inference*, **142**, 2322–2329.
- Polonik, W. (1995a) Density estimation under qualitative assumptions in higher dimensions. *Journal of Multivariate Analysis*, **55**, 61–81.
- Polonik, W. (1995b) Measuring mass concentrations and estimating density contour clusters-An excess mass approach. *The annals of Statistics*, **23**, 855–881.
- Polonik, W. (1997) Minimum volume sets and generalized quantile processes. *Stochastic processes and their applications*, **69**, 1–24.
- Polonik, W. (1998) The silhouette, concentration functions and ML-density estimation under order restrictions. *Annals of statistics*, 1857–1877.
- Rao, B. P. (1969) Estimation of a unimodal density. *Sankhyā: The Indian Journal of Statistics, Series A*, 23–36.
- Robertson, T., Wright, F. and Dykstra, R. (1988) *Order Restricted Statistical Inference*. Chichester: John Wiley & Sons.
- Rodríguez Casal, A. (2007) Set estimation under convexity type assumptions. In *Annales de l’IHP Probabilités et statistiques*, vol. 43, 763–774.
- Rousseeuw, P. J. (1985) Multivariate estimation with high breakdown point. *Mathematical statistics and applications*, **8**, 283–297.
- Samworth, R. J. (2018) Recent progress in log-concave density estimation. *Statistical Science*, **33**, 493–509.
- Scott, D. W. (2015) *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.
- Thäle, C. (2008) 50 years sets with positive reach-a survey-. *Surveys in Mathematics & its Applications*, **3**, 123–165.
- van der Vaart, A. W. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes*. Springer.
- Vershynin, R. (2018) *High-dimensional Probability: An Introduction with Applications in Data Science*, vol. 47. Cambridge university press.

Walther, G. (1999) On a generalization of Blaschke's rolling theorem and the smoothing of surfaces.
Mathematical methods in the applied sciences, **22**, 301–316.