

Exact asymptotic limit for kernel estimation of regression level sets

Dau Dang, Thomas Laloë, Rémi Servien

► **To cite this version:**

Dau Dang, Thomas Laloë, Rémi Servien. Exact asymptotic limit for kernel estimation of regression level sets. *Statistics and Probability Letters*, Elsevier, 2020, 161 (108721), 10.1016/j.spl.2020.108721 . hal-02073071v2

HAL Id: hal-02073071

<https://hal.archives-ouvertes.fr/hal-02073071v2>

Submitted on 10 Feb 2020

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

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

Exact asymptotic limit for kernel estimation of regression level sets

Dau Hai Dang^a, Thomas Laloë ^{*a}, and Rémi Servien^b

^aUniversité Côte d'Azur, CNRS, LJAD, France

^bINTHERES, Université de Toulouse, INRA, ENVT, Toulouse, France.

Abstract

The asymptotic behavior of a plug-in kernel estimator of the regression level sets is studied. The exact asymptotic limit of the symmetric difference is derived for a given level and for an unknown level corresponding to a fixed probability.

1 Introduction

Level set estimation is a general statistical framework where one has to estimate the t -level set of a function g (i.e. the set $\{x \mid g(x) \geq t\}$) using observations where randomness is involved. Depending on the context, the function g can be a density ([Hartigan, 1987](#); [Tsybakov, 1997](#); [Cadre, 2006](#); [Mason and Polonik, 2009](#); [Chen et al., 2017](#)), a cumulative distribution function ([Di Bernardino et al., 2013, 2015](#)) or a regression function which is the framework of this article. Let (X, Y) be a pair of random variables taking value in $\mathbb{R}^d \times \mathbb{R}$. The function $r : x \mapsto \mathbb{E}[Y \mid X = x]$ is called the regression function of Y with respect to X . There are different applications of estimating $\mathcal{L}(t) := \mathcal{L}^r(t) = \{x \mid r(x) \geq t\}$. For example, in medical treatment, Y can be a variable characterizing the severity of a cancer. In certain cases, one compares Y to a threshold γ to choose between standard chemotherapy or aggressive chemotherapy. The problem is that measuring Y can be complicated, so one might hope to infer Y from a feature vector X of the patient where X can be obtained more easily. The problem then becomes the estimation of $\mathcal{L}(\gamma)$. Similarly, in economics, X can be demographic information concerning a person and Y can be his income. From a government point of view, it is relevant to estimate $\mathcal{L}(t)$ where t may be the poverty threshold or, on the contrary, a very high income threshold. For more details the reader is referred to [Scott and Davenport \(2007\)](#).

Despite the many potential applications, the estimation of the level sets of the regression function has not been widely studied. [Willett and Nowak \(2007\)](#) obtained minimax rates (for different smoothness classes) for

*Corresponding author : Thomas.Laloe@unice.fr

estimators based on recursive dyadic partitions. [Scott and Davenport \(2007\)](#) used a cost sensitive approach and a different measure of risk. [Cavalier \(1997\)](#) and [Polonik and Wang \(2005\)](#) used estimators based on the maximization of the excess mass. Cavalier demonstrated asymptotic minimax rate of convergence for piecewise polynomial estimators using smoothness assumptions on the boundary of the level sets. [Laloë and Servien \(2013\)](#) used a different approach and constructed a plug-in estimator $\mathcal{L}_n(t)$ defined by $\mathcal{L}_n(t) = \{x \in \mathbb{R}^d : r_n(x) \geq t\}$, where $r_n(x)$ is the kernel estimator of the regression. The main advantage of this estimator lies in the simplicity of its calculation, inherited from the plug-in approach. Moreover, this estimator does not require strong assumptions on the shape of level sets. The error is investigated in term of the volume of the symmetrical difference between the real and the estimated level sets

$$d_\lambda(\mathcal{L}_n(t), \mathcal{L}(t)) = \lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t)),$$

where $\mathcal{L}_n(t) \Delta \mathcal{L}(t) = (\mathcal{L}_n(t) \cap \mathcal{L}^C(t)) \cup (\mathcal{L}_n^C(t) \cap \mathcal{L}(t))$ and λ is a measure on \mathbb{R}^d (for example the Lebesgue measure). The estimator is shown to be consistent with $\mathbb{E} \lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t)) = O(1/\sqrt{nh^d})$. To our knowledge, the exact asymptotic behavior of the Lebesgue measure of the volume of the symmetric difference is still unknown for the regression function whereas an exact asymptotic limit has been obtained under reasonable assumptions by [Cadre \(2006\)](#) for the density case. Thus, our objective is to give exact asymptotic rate and exact asymptotic limit in the regression case.

Section 2 is devoted to definitions and notations. In Section 3, we state the exact asymptotic rate and limit for the regression framework. Then, we extend this result to the case of an unknown level t . Proofs are sketched and gathered in Section 4.

2 Definitions, notations and assumptions

Let (X, Y) be a pair of random variables taking values in $\mathbb{R}^d \times J$ where $d \geq 2$ and J is a bounded subset of \mathbb{R} . Suppose that X has a density f and let $r(x) = \mathbb{E}[Y|X = x]$. Let K be a probability density on \mathbb{R}^d with $\tilde{K} = \int K^2 < \infty$ and denote by r_n the corresponding kernel estimation of r , that is

$$r_n(x) = \begin{cases} \frac{\phi_n(x)}{f_n(x)} & \text{if } f_n(x) > 0; \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $\phi_n(x) = \frac{1}{nh^d} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)$ and $f_n(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$ with $h = h(n) \rightarrow 0$ being the bandwidth parameter. We use λ to denote the Lebesgue measure and λ_m the measure induced by a bounded function m in the sense that $\lambda_m(S) = \int m \mathbb{1}_{x \in S} dx$ for measurable sets S . Φ denotes the cumulative distribution function of the standard Gaussian distribution $\mathcal{N}(0, 1)$ and c some strictly positive constant whose values may vary

from line to line. From now on, $\Theta \subset (\inf \mathbf{r}, \sup \mathbf{r})$ is an open interval. Let us introduce some assumptions :

- Assumptions on (X, Y) :

A0 (Definition of t^-) For all $t \in \Theta$ there exists $\inf r \leq t^- \leq t$ such that $\mathcal{L}(t^-)$ is compact;

A1 The functions r and f belong to class $C^2(\mathbb{R}^d)$ and $\forall t \in \Theta, \inf_{\mathcal{L}(t^-)} f > 0$;

A2 For all $t \in \Theta, \inf_{r^{-1}(t)} \|\nabla r\| > 0$;

A3 The function $q(x) = \text{Var}(Y|X = x)$ belongs to class C^2 and satisfies $\inf_{r^{-1}(\Theta)} q > 0$.

- Assumptions on the kernel and bandwidth :

B1 The kernel K belongs to the class C^2 , has compact support which we suppose to be contained in $B(0, 1)$ and satisfies $K(x) = K(-x)$;

B2 The bandwidth satisfies $\frac{nh^d}{\log^8 n} \rightarrow \infty$ and $nh^{\min(d+4, 2d)} \log^8 n \rightarrow 0$.

Note that **A0** is a standard assumption in the theory of set estimation. Indeed, it is very difficult to measure errors if we try to estimate an unbounded set. Assumption **A2** means that r has no plateau in levels around t . If r had a plateau at level t , then even a very good estimator of r (in the sense that $|r_n - r|$ is small) might still produce a bad estimator of $\mathcal{L}(t)$. Note that under **B2**, the bias is dominated by the variance. This is a not ideal but acceptable setting in statistics in general and in kernel estimation in particular, as mentioned in [Cadre \(2006\)](#).

3 Theoretical results

3.1 Known t

We first consider the case where the level t is known and obtain a rate of consistency for the volume of the symmetric difference $\lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t))$:

Theorem 1. *Under assumptions A0 to B2, we have for $d \geq 2$*

$$\sqrt{nh^d} \lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t)) \xrightarrow{\mathbb{P}} \sqrt{\frac{2\tilde{K}}{\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}, \quad (2)$$

and

$$\sqrt{nh^d} \mathbb{E} [\lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t))] \rightarrow \sqrt{\frac{2\tilde{K}}{\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H} \quad (3)$$

where \mathcal{H} denotes the $(d-1)$ -dimensional Hausdorff measure on \mathbb{R}^d ([Evans and Gariepy, 2015](#)).

This theorem gives the exact asymptotic limit for the Lebesgue measure of the symmetric difference between the true level set $\mathcal{L}(t)$ and the estimated set $\mathcal{L}_n(t)$. Note that (2) is a natural extension of Theorem 2.1 in [Cadre \(2006\)](#) to the regression case and that (3) is an improvement of Theorem 2.1 in [Laloë and Servien \(2013\)](#) where only a convergence rate of $O(\sqrt{nh^d})$ is proved.

3.2 Unknown t

We now consider the problem of estimating $\mathcal{L}(t)$ when t is unknown because it is defined via a probability p such that $\mathbb{P}(r(X) \geq t) = p$. We define an estimator t_n for t and the estimator for $\mathcal{L}(t)$ becomes $\mathcal{L}_n(t_n)$. This definition of t can be of interest for practical applications, for example, when an insurance company wishes to focus on cost levels for unlikely risk factor values. In this case, t defines a level such that the set $\mathcal{L}(t)$ of risk values for which the cost is greater than this level is of probability $\mathbb{P}(\mathcal{L}(t)) = p$. In this spirit, we can mention the hydrology application presented in [Di Bernardino et al. \(2015\)](#). For this problem, we need a more restrictive bandwidth assumption (which implies assumption **B2**):

B2' $d \geq 3$ and the bandwidth satisfies $\frac{nh^{d+2}}{\log n} \rightarrow \infty$ and $nh^{\min(d+4, 2d)} \log^8 n \rightarrow 0$.

Note that this assumption excludes the case $d = 2$, contrary to **B2**, and that a problem of similar nature occurs in [Cadre \(2006\)](#) when an unknown level is considered.

Theorem 2. *Define the function P by $P(u) = \mathbb{P}(r(X) \geq u)$ and let $\Theta_s = [s_1, s_2]$ be a strict subinterval of Θ . Let p be such that $p \in P(\Theta_s)$ and let t be such that $P(t) = p$. Then, under assumptions A0-B2':*

1. *Almost surely for $n > n_0 = n_0(\omega)$, there exists a unique t_n such that $\int_{\mathcal{L}(t^-)} f_n \mathbb{1}_{r_n \geq t_n} = p$. In particular, $\mathbb{P}\left(\exists! t_n \text{ s.t. } \int_{\mathcal{L}(t^-)} f_n \mathbb{1}_{r_n \geq t_n} = p\right) \rightarrow 1$;*
2. *We have $t_n \xrightarrow{\text{a.s.}} t$;*
3. *We have $\sqrt{nh^d} \lambda[\mathcal{L}_n(t_n) \Delta \mathcal{L}(t)] \xrightarrow{\mathbb{P}} \sqrt{\frac{2\tilde{K}}{\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}$.*

Remark that to compute this estimation in practice, we propose to replace $\mathcal{L}(t^-)$ in the integral by the smallest compact d dimensional cube containing all the data set. Furthermore, we can note that even if the problem of estimating $\mathcal{L}(t)$ where t is unknown seems to be more difficult, the convergence rate stays the same.

3.3 Discussion of the results

First of all, note that the convergence speed and the exact form of the limit are similar in the two theorems. Moreover if we consider the restrictions on the bandwidth we see that the best convergence rate we can get is $O\left(n^{\frac{2}{d+4}}\right)$ up to some $\log n$ factor. We are in a situation similar to [Cadre \(2006\)](#) with a clear curse of dimensionality.

Let us now discuss the form of the exact limit :

- The term $\sqrt{q/f}$ in the asymptotic limit is natural: estimating $\mathcal{L}(t)$ is easier when the variability is low and the density is high;
- Note that in the trivial case $q = 0$ everywhere, Y would be a deterministic function of X . Estimating $\mathcal{L}(t)$ would be a lot easier and thus we would have a faster convergence speed than $\sqrt{nh^d}$;
- The exact limit depends on an integral over $r^{-1}\{t\}$. This is not surprising because $r^{-1}\{t\}$ corresponds to the boundary of $\mathcal{L}(t)$ which is the place where the estimation is more difficult.

The results presented in Theorems 1 and 2 provide a natural and elegant extension of the results presented in (Cadre, 2006; Laloë and Servien, 2013). However a nice perspective would be to get a asymptotic normality as the one obtained in Polonik and Wang (2005) in the density function case. In a more practical perspective, addressing the problems of the choice of the bandwidth and of the estimation of the integral $\int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}$ using the recent works of Doss and Weng (2018); Qiao (2018, 2019) would be of great interest.

4 Proofs

4.1 Proof of Theorem 1

In all this section, c will denote generic constants whose values can change from line to line.

Sketch of the proof : By definition of the symmetric difference, one can write $\lambda(\mathcal{L}_n(t)\Delta\mathcal{L}(t)) = \int_{\{r < t\} \cap \mathcal{L}(t^-)} \mathbb{1}_{r_n(x) \geq t} dx + \int_{\{r > t\} \cap \mathcal{L}(t^-)} \mathbb{1}_{r_n(x) < t} dx$. We focus on the first term as the second one can be treated similarly. One can approximate the integral of $\mathbb{1}_{r_n(x) \geq t}$ over $\{r < t\} \cap \mathcal{L}(t^-)$ by the integral of $\mathbb{1}_{r_n(x) \geq t}$ over $V_n^t = \{r \in [t - \Gamma_n, t]\}$, where Γ_n is some small number using Lemma 3 (i.e. $\Gamma_n := \Gamma \sqrt{\frac{\log n}{nh^d}}$).

So we only need to show that $\sqrt{nh^d} \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx$ tends in probability to some explicit number c . This can be derived by showing that its expectation tends to c and its variance tends to 0. By Fubini's theorem, one can write $\mathbb{E} \left(\sqrt{nh^d} \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx \right)$ as $\sqrt{nh^d} \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx$. Then, by using Berry-Esseen inequality and properties of kernel estimators, one can give an explicit formula to approximate $\mathbb{P}(r_n(x) \geq t)$ with Lemmas 4 and 5. This explicit formula eases the calculation of the limit of $\sqrt{nh^d} \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx$ which is provided in Proposition 7. Then, the variance term is treated in Proposition 9. Again by Fubini's theorem, one can show that

$$\text{Var} \left(\sqrt{nh^d} \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx \right) = nh^d \iint_{(V_n^t)^2} \mathbb{P}(r_n(x) \geq t, r_n(y) \geq t) - \mathbb{P}(r_n(x) \geq t) \mathbb{P}(r_n(y) \geq t) dx dy.$$

Recall that $r_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}$ and that K has support in $B(0,1)$. This means that $r_n(x)$ only depends on X_i if $\|X_i - x\| \leq h$ and so $r_n(x)$ and $r_n(y)$ are "almost independent" if $\|x - y\| \geq 2h$. One might therefore hope that $\mathbb{P}(r_n(x) \geq t, r_n(y) \geq t) - \mathbb{P}(r_n(x) \geq t)\mathbb{P}(r_n(y) \geq t)$ is small. In the proof of Proposition 3.2 in [Cadre \(2006\)](#), Cadre solved this problem by approximating $\mathbb{P}(f_n(x) \geq t, f_n(y) \geq t) - \mathbb{P}(f_n(x) \geq t)\mathbb{P}(f_n(y) \geq t)$ in terms of $C_{x,y}(u, v) - C_x(u)C_y(v)$ (where $C_{x,y}$, C_x and C_y are respectively characteristic functions of $(f_n(x), f_n(y)), f_n(x)$ and $f_n(y)$). Then he proved that $C_{x,y}(u, v) - C_x(u)C_y(v)$ is small. Unfortunately, this approach can not be generalized easily to the regression case.

In fact, even when $\|x - y\| \geq 2h$, $r_n(x)$ and $r_n(y)$ are still dependent because the random variables N_x and N_y are not independent (where N_x and N_y are respectively the number of points belonging to $B(x, h)$ and $B(y, h)$ among X_1, \dots, X_n). Therefore, we study the dependence between $r_n(x)$ and $r_n(y)$ in Lemma 8. Using this Lemma, Proposition 9 proves that $\text{Var}\left(\sqrt{nh^d} \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx\right)$ tends to 0, which concludes the proof of Theorem 1. \square

Proof of Theorem 1. By Lemma 3, one has $\sqrt{nh^d} \left[\lambda(\mathcal{L}_n(t)\Delta\mathcal{L}(t)) - \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx - \int_{\bar{V}_n^t} \mathbb{1}_{r_n(x) < t} dx \right] \xrightarrow{\mathbb{P}} 0$. Proposition 7 allows us to conclude that $\mathbb{E} \left[\sqrt{nh^d} \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx \right] \rightarrow \sqrt{\frac{\bar{K}}{2\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}$, while Proposition 9 gives $\text{Var}\left(\sqrt{nh^d} \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx\right) \rightarrow 0$. This ensures that

$$\sqrt{nh^d} \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx \xrightarrow{\mathbb{P}} \sqrt{\frac{\bar{K}}{2\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}.$$

The same arguments applied to $\sqrt{nh^d} \int_{\bar{V}_n^t} \mathbb{1}_{r_n(x) < t} dx$ complete the proof. \square

Lemma 3. Let $\Gamma > 0$ be a constant such that $\sqrt{nh^d} \mathbb{P}\left(\sup_{\mathcal{L}(t-)} |r_n - r| > \Gamma \sqrt{\frac{\log n}{nh^d}}\right) \rightarrow 0$ ([Einmahl and Mason, 2005](#); [Laloë and Servien, 2013](#)). Moreover define $V_n^t = r^{-1}\left[t - \Gamma \sqrt{\frac{\log n}{nh^d}}, t\right]$ and $\bar{V}_n^t = r^{-1}\left[t, t + \Gamma \sqrt{\frac{\log n}{nh^d}}\right]$.

Then, under the assumptions of Theorem 1, we have :

$$(i) \sqrt{nh^d} \left(\lambda(\mathcal{L}_n(t)\Delta\mathcal{L}(t)) - \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx - \int_{\bar{V}_n^t} \mathbb{1}_{r_n(x) < t} dx \right) \xrightarrow{\mathbb{P}} 0;$$

$$(ii) \sqrt{nh^d} \left[\mathbb{E} \lambda(\mathcal{L}_n(t)\Delta\mathcal{L}(t)) - \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx - \int_{\bar{V}_n^t} \mathbb{P}(r_n(x) < t) dx \right] \rightarrow 0.$$

Remark. Using Prop A.2 of [Cadre \(2006\)](#), we have $\max\left(\lambda(V_n^t), \lambda(\bar{V}_n^t)\right) \leq c\sqrt{\frac{\log n}{nh^d}}$.

Proof. Since the proofs are quite similar, we only deal with the point (ii) of Lemma 3. Note that

$$\begin{aligned} & \sqrt{nh^d} \int_{\{r \leq t - \Gamma \sqrt{\frac{\log n}{nh^d}}\} \cap \mathcal{L}(t^-)} \mathbb{P}(r_n(x) \geq t) dx \leq \sqrt{nh^d} \int_{\{r \leq t - \Gamma \sqrt{\frac{\log n}{nh^d}}\} \cap \mathcal{L}(t^-)} \mathbb{P}\left(\sup_{\mathcal{L}(t^-)} |r_n - r| \geq \Gamma \sqrt{\frac{\log n}{nh^d}}\right) dx \\ & \leq |\mathcal{L}(t^-)| \sqrt{nh^d} \mathbb{P}\left(\sup_{\mathcal{L}(t^-)} |r_n - r| \geq \Gamma \sqrt{\frac{\log n}{nh^d}}\right) \rightarrow 0 \text{ by Lemma 4.1. in Laloë and Servien (2013)} \end{aligned}$$

and, similarly $\sqrt{nh^d} \int_{\{r > t + \Gamma \sqrt{\frac{\log n}{nh^d}}\}} \mathbb{P}(r_n(x) < t) dx \rightarrow 0$. As we have

$$\begin{aligned} \sqrt{nh^d} \mathbb{E} \lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t)) &= \sqrt{nh^d} \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx + \sqrt{nh^d} \int_{\bar{V}_n^t} \mathbb{P}(r_n(x) < t) dx \\ &+ \sqrt{nh^d} \int_{\{r \leq t - \Gamma \sqrt{\frac{\log n}{nh^d}}\} \cap \mathcal{L}(t^-)} \mathbb{P}(r_n(x) \geq t) dx + \sqrt{nh^d} \int_{\{r > t + \Gamma \sqrt{\frac{\log n}{nh^d}}\}} \mathbb{P}(r_n(x) < t) dx, \end{aligned}$$

we have (ii). \square

Lemma 4. Let $V_n(x, t) = \text{Var}((Y - t)K(\frac{x-X}{h}))$ and set $\tilde{\tau}_n(x) = (tf(x) - \phi(x)) \sqrt{\frac{nh^d}{\tilde{K}f(x)[q+(r(x)-t)^2]}}$. Then, under the assumptions of Theorem 1, $\sqrt{nh^d} \left(\int_{V_n^t} \mathbb{P}(r_n(x) \geq t) - \int_{V_n^t} \bar{\Phi}(\tilde{\tau}_n(x)) \right) dx \rightarrow 0$.

Proof. Set $\tau_n(x) = (t\mathbb{E}f_n(x) - \mathbb{E}\phi_n(x)) \sqrt{\frac{nh^d}{V_n(x,t)h^{-d}}}$. Using Lemma 4.1. in Laloë and Servien (2013) and Prop A.2. in Cadre (2006), we have

$$\begin{aligned} & \left| \sqrt{nh^d} \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx - \int_{V_n^t} \bar{\Phi}(\tau_n(x)) dx \right| \leq \sqrt{nh^d} |V_n^t| \sup_{V_n^t} |\mathbb{P}(r_n(x) \geq t) - \bar{\Phi}(\tau_n(x))| \\ & \leq \sqrt{nh^d} \left(c \sqrt{\frac{\log n}{nh^d}} \right) \frac{c}{\sqrt{nh^d}} = c \sqrt{\frac{\log n}{nh^d}} \rightarrow 0 \text{ under assumptions on } h. \end{aligned} \quad (4)$$

Straightforward calculations show that

$$\left| V_n(x, t) h^{-d} - \tilde{K}f(q + (r - t)^2)(x) \right| \leq ch^2, \forall n, \forall x \in V_n^t \quad \text{and} \quad (5)$$

$$|\mathbb{E}f_n(x) - f(x)| \leq ch^2, |\mathbb{E}\phi_n(x) - \phi(x)| \leq ch^2, \forall x \in r^{-1}(\Theta). \quad (6)$$

Therefore, for all $x \in V_n^t$,

$$\begin{aligned} & \frac{1}{\sqrt{nh^d}} |\tau_n(x) - \tilde{\tau}_n(x)| \leq |t\mathbb{E}f_n - \mathbb{E}\phi_n| \left| \frac{1}{\sqrt{V_n(x,t)h^{-d}}} - \frac{1}{\sqrt{\tilde{K}w(x)}} \right| + |t(\mathbb{E}f_n - f) - (\mathbb{E}\phi_n - \phi)| \frac{1}{\sqrt{\tilde{K}w(x)}} \\ & \leq c \frac{\left| \tilde{K}w(x) - V_n(x,t)h^{-d} \right|}{\sqrt{V_n(x,t)h^{-d}} \sqrt{\tilde{K}w(x)} \left(\sqrt{V_n(x,t)h^{-d}} + \sqrt{\tilde{K}w(x)} \right)} + ch^2 \frac{1}{\sqrt{\tilde{K}w(x)}} \leq ch^2 \text{ by (5), A1 and A3.} \end{aligned}$$

Finally, using Lipschitz property of Φ , $\left| \sqrt{nh^d} \left(\int_{V_n^t} \bar{\Phi}(\tau_n(x)) dx - \int_{V_n^t} \bar{\Phi}(\tilde{\tau}_n(x)) \right) dx \right| \leq c \sqrt{nh^{d+4} \log n} \rightarrow 0$.

This convergence and (4) conclude the lemma. \square

Lemma 5. Let $a_n(x) = (tf(x) - \phi(x)) \sqrt{\frac{nh^d}{\tilde{K}f(x)q(x)}} = (t - r(x)) \sqrt{\frac{nh^d f(x)}{\tilde{K}q(x)}}$. Then, under the assumptions of Theorem 1, $\sqrt{nh^d} \left(\int_{V_n^t} \bar{\Phi}(\tilde{\tau}_n(x)) dx - \int_{V_n^t} \bar{\Phi}(a_n(x)) dx \right) \rightarrow 0$.

Proof. We have

$$\begin{aligned} & \left| \sqrt{nh^d} \left[\int_{V_n^t} \bar{\Phi}(\tilde{\tau}_n(x)) dx - \int_{V_n^t} \bar{\Phi}(a_n(x)) dx \right] \right| \\ & \leq c\sqrt{nh^d} |V_n^t| \sup_{V_n^t} \left[|tf - \phi| \sqrt{\frac{nh^d}{\tilde{K}f}} \left| \frac{1}{\sqrt{q + (r-t)^2}} - \frac{1}{\sqrt{q}} \right| \right] \text{ by Lipschitz property of } \bar{\Phi} \\ & \leq c\sqrt{nh^d \log n} \sup_{V_n^t} (r-t)^2 \text{ using that } |V_n^t| \leq c\sqrt{\frac{\log n}{nh^d}} \text{ and } \inf_{r^{-1}(\Theta)} q > 0 \\ & \leq c\sqrt{\frac{\log^3 n}{nh^d}} \text{ by the definition of } V_n^t, \end{aligned}$$

which tends to 0 by assumptions on h , hence the lemma. \square

Lemma 6. Let $\epsilon^* > 0$ such that $[t - 4\epsilon^*, t + 4\epsilon^*] \subset \Theta$. Under the assumptions of Theorem 1, there exists a function $b \in C^2(\mathbb{R}^d, \mathbb{R})$ bounded between two strictly positive numbers \underline{b} and \bar{b} such that $b(x) = \frac{\tilde{K}q(x)}{f(x)}$ on $r^{-1}[t - \epsilon^*, t + \epsilon^*]$ and $\lim_{x \rightarrow \infty} \nabla b(x) = 0$. Furthermore $\forall n$ and $\forall u \leq \Gamma\sqrt{\frac{\log n}{\bar{b}}}$, we have $\{x | u \geq a_n(x), x \in V_n^t\} = \left\{ x | r(x) \in \left[t - u\sqrt{\frac{b(x)}{nh^d}}, t \right] \right\}$.

Proof. By the continuity of r , $r^{-1}[t - \epsilon^*, t + \epsilon^*]$ is a compact set contained in the open set $r^{-1}[t - 2\epsilon^*, t + 2\epsilon^*]$. By proposition 2.26 in Lee (2003), there exists a nonnegative function p such that $p = 1$ on $r^{-1}[t - \epsilon^*, t + \epsilon^*]$ and the support of p is contained in $r^{-1}[t - 2\epsilon^*, t + 2\epsilon^*]$. By A1 and A3, the function $\frac{\tilde{K}q}{f}$ is bounded on $r^{-1}(\Theta)$ between two strictly positive numbers denoted \underline{b} and \bar{b} . Then the function $b(x)$ defined as follows satisfies our requirements:

$$b(x) = \begin{cases} \left(\frac{\tilde{K}q(x)}{f(x)} - \underline{b} \right) p(x) + \underline{b} & \text{if } x \in r^{-1}[t - 3\epsilon^*, t + 3\epsilon^*]; \\ \underline{b} & \text{otherwise.} \end{cases}$$

For the second point, recall that on V_n^t , $a_n(x) = (t - r(x)) \sqrt{\frac{nh^d}{b(x)}}$ so on V_n^t , we have $u \geq a_n(x) \iff r(x) \geq t - u\sqrt{\frac{b(x)}{nh^d}}$. So we have easily $\{x | u \geq a_n(x), x \in V_n^t\} \subset \left\{ x | r(x) \in \left[t - u\sqrt{\frac{b(x)}{nh^d}}, t \right] \right\}$. Inversely, let x be such that $r(x) \in \left[t - u\sqrt{\frac{b(x)}{nh^d}}, t \right]$. Then $r(x) \geq t - u\sqrt{\frac{b(x)}{nh^d}} \geq t - \Gamma\sqrt{\frac{b(x) \log n}{\bar{b}nh^d}}$ (by assumptions on u) $\geq t - \Gamma\sqrt{\frac{\log n}{nh^d}}$. This means that $x \in V_n^t$ and thus $\left\{ x | r(x) \in \left[t - u\sqrt{\frac{b(x)}{nh^d}}, t \right] \right\} \subset \{x | u \geq a_n(x), x \in V_n^t\}$. \square

Proposition 7. Under the assumptions of Theorem 1, we have $\sqrt{nh^d} \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx \rightarrow \sqrt{\frac{\tilde{K}}{2\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}$ and $\sqrt{nh^d} \mathbb{E}[\lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t))] \rightarrow \sqrt{\frac{2\tilde{K}}{\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}$.

Proof. We have

$$\begin{aligned} \sqrt{nh^d} \int_{V_n^t} \bar{\Phi}(a_n(x)) dx &= \sqrt{nh^d} \int_0^{\Gamma \sqrt{\frac{\log n}{b}}} \Phi'(u) \int_{\mathbb{R}} \mathbb{1}_{u \geq a_n(x)} \mathbb{1}_{x \in V_n^t} dx du \\ &+ \sqrt{nh^d} \int_{\Gamma \sqrt{\frac{\log n}{b}}}^{\infty} \Phi'(u) \int_{\mathbb{R}} \mathbb{1}_{u \geq a_n(x)} \mathbb{1}_{x \in V_n^t} dx du := I_1 + I_2. \end{aligned} \quad (7)$$

Using Lemma 6, a generalization of Lemma 3.2 and Propositions A.1 and A.2 of Cadre (2006), Lebesgue's dominated convergence theorem gives

$$I_1 \rightarrow \int_0^{\infty} \Phi'(u) du \int_{r^{-1}\{t\}} \frac{\sqrt{b}}{\|\nabla r\|} d\mathcal{H} = \sqrt{\frac{1}{2\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{b}}{\|\nabla r\|} d\mathcal{H} = \sqrt{\frac{\tilde{K}}{2\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}.$$

We also have that $I_2 \leq \sqrt{nh^d} \int_{\Gamma \sqrt{\frac{\log n}{b}}}^{\infty} \Phi'(u) du \int_{\mathbb{R}} \mathbb{1}_{x \in V_n^t} dx \leq c\sqrt{\log n} \int_{c\sqrt{\log n}}^{\frac{1}{u^3}} du \rightarrow 0$. Using (7) this leads us to $\sqrt{nh^d} \int_{V_n^t} \bar{\Phi}(a_n(x)) dx \rightarrow \sqrt{\frac{\tilde{K}}{2\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}$. Lemma 3, 4 and 5 end the proof. \square

Lemma 8. Let $x, y \in \mathbb{R}^d$ such that $\|x - y\| \geq 2h$ and put $N_x = \sum_{i=1}^n \mathbb{1}_{X_i \in B(x;h)}$. Let $D_x = \mathbb{P}(X \in B(x, h))$, then we know that $D_x = \int_{B(x,h)} f(u) du \leq \alpha h^d$ where α does not depend neither on x nor on n . Then, under the assumptions of Theorem 1, we have $\lim_{n \rightarrow \infty} J_n = 0$ where $J_n = \sup_{\substack{x, y \in \mathbb{R}^d \\ \|x - y\| \geq 2h \\ u, v \leq 2\alpha n h^d}} \left| \frac{\mathbb{P}(N_x = u, N_y = v)}{\mathbb{P}(N_x = u)\mathbb{P}(N_y = v)} - 1 \right|$.

Proof. Let $(x_n), (y_n), (u_n)$ and (v_n) be four sequences such that $\|x_n - y_n\| \geq 2h$ and $u_n, v_n \leq 2\alpha n h^d$. It is sufficient to show that $\frac{\mathbb{P}(N_{x_n} = u_n, N_{y_n} = v_n)}{\mathbb{P}(N_{x_n} = u_n)\mathbb{P}(N_{y_n} = v_n)} \rightarrow 1$. Using the definitions of N_x and D_x , this quotient can be rewritten as

$$\frac{(n - u_n)!(n - v_n)!}{n!(n - u_n - v_n)!} \frac{(1 - D_{x_n} - D_{y_n})^{n - u_n - v_n}}{(1 - D_{x_n})^{n - u_n} (1 - D_{y_n})^{n - v_n}}.$$

From the Stirling formula, the first term is equivalent to $\frac{(1 - \frac{u_n}{n})^{n - u_n} (1 - \frac{v_n}{n})^{n - v_n}}{(1 - \frac{u_n + v_n}{n})^{n - u_n - v_n}}$ when $n \rightarrow \infty$. Thus, one only needs to prove the following limits:

$$\frac{(1 - \frac{u_n}{n})^{n - u_n} (1 - \frac{v_n}{n})^{n - v_n}}{(1 - \frac{u_n + v_n}{n})^{n - u_n - v_n}} \rightarrow 1 \text{ and } \frac{(1 - D_{x_n})^{n - u_n} (1 - D_{y_n})^{n - v_n}}{(1 - D_{x_n} - D_{y_n})^{n - u_n - v_n}} \rightarrow 1.$$

The left hand of the first limit can be written as

$$\frac{(1 - \frac{u_n + v_n}{n})^{u_n + v_n}}{(1 - \frac{u_n}{n})^{u_n} (1 - \frac{v_n}{n})^{v_n}} \left(1 + \frac{u_n v_n / n^2}{1 - \frac{u_n + v_n}{n}}\right)^n := \frac{A_1}{A_2 \times A_3} \times A_4.$$

By using straightforward calculations and the fact that $u_n, v_n \leq 2\alpha n h^d$, one can show that $A_i \rightarrow 1, \forall i \in \{1, 2, 3, 4\}$, thus finish the demonstration of the lemma. \square

Proposition 9. Under the assumptions of Theorem 1, we have $\text{Var} \left(\sqrt{nh^d} \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx \right) \rightarrow 0$.

Proof. First, one has

$$\text{Var} \left(\sqrt{nh^d} \int_{V_n^t} \mathbb{1}_{r_n(x) \geq t} dx \right) = nh^d \iint_{(V_n^t)^2} \mathbb{P}(r_n(x) \geq t, r_n(y) \geq t) - \mathbb{P}(r_n(x) \geq t) \mathbb{P}(r_n(y) \geq t) dx dy. \quad (8)$$

Now, set $S_n = \{x, y \in V_n^t \mid \|x - y\| \geq 2h\}$ and $\widetilde{S}_n = \{x, y \in V_n^t \mid \|x - y\| \leq 2h\}$. Then

$$\left| nh^d \iint_{\widetilde{S}_n} \mathbb{P}(r_n(x) \geq t, r_n(y) \geq t) - \mathbb{P}(r_n(x) \geq t) \mathbb{P}(r_n(y) \geq t) dx dy \right| \leq cnh^d \sqrt{\frac{\log n}{nh^d}} h^d \rightarrow 0, \quad (9)$$

whereas, by using straightforward calculations, one can rewrite the integral over S_n as

$$\left| nh^d \iint_{S_n} \sum_{u,v} Q_{x,u} Q_{y,v} (\mathbb{P}(N_x = u, N_y = v) - \mathbb{P}(N_x = u) \mathbb{P}(N_y = v)) dx dy \right| \quad (10)$$

with $Q_{x,u} = \mathbb{P} \left(\frac{\sum_{i=1}^u Y_i K(\frac{x-X_i}{h})}{\sum_{i=1}^u K(\frac{x-X_i}{h})} \geq t \mid X_1, \dots, X_u \in B(x; h) \right)$. Moreover

$$\begin{aligned} & nh^d \iint_{S_n} \sum_{u \vee v \geq 2\alpha nh^d} |\mathbb{P}(N_x = u, N_y = v) - \mathbb{P}(N_x = u) \mathbb{P}(N_y = v)| dx dy \\ & \leq n^2 h^d \iint_{S_n} \mathbb{P}(N_x \geq 2\alpha nh^d) + \mathbb{P}(N_y \geq 2\alpha nh^d) dx dy. \end{aligned} \quad (11)$$

Observe that

$$\begin{aligned} \mathbb{P}(N_x \geq 2\alpha nh^d) & \leq \mathbb{P}(N_x \geq nD_x + \alpha nh^d) \text{ where } D_x \text{ is defined in Lemma 8} \\ & \leq \exp \left(-\frac{\alpha^2 n^2 h^{2d}}{2nD_x + \frac{2}{3}\alpha nh^d} \right) \text{ by using } N_x \sim \text{Binomial}(n, D_x) \\ & \leq e^{-\frac{1}{3}\alpha nh^d} \end{aligned}$$

using Bernstein's inequality (see Theorem 1 in [Janson \(1999\)](#)). With (11), this leads us to

$$nh^d \iint_{S_n} \sum_{u \vee v \geq 2\alpha nh^d} |\mathbb{P}(N_x = u, N_y = v) - \mathbb{P}(N_x = u) \mathbb{P}(N_y = v)| dx dy \leq cn \log n e^{-\frac{1}{3}\alpha nh^d} \rightarrow 0 \quad (12)$$

as $\frac{nh^d}{\log n} \rightarrow \infty$. We also have that

$$\begin{aligned} & nh^d \iint_{S_n} \sum_{u,v \leq 2\alpha nh^d} Q_{x,u} \mathbb{P}(N_x = u) Q_{y,v} \mathbb{P}(N_y = v) \left| \frac{\mathbb{P}(N_x = u, N_y = v)}{\mathbb{P}(N_x = u) \mathbb{P}(N_y = v)} - 1 \right| dx dy \\ & = J_n \left(\sqrt{nh^d} \int_{V_n^t} \mathbb{P}(r_n(x) \geq t) dx \right)^2. \end{aligned}$$

with J_n defined in Lemma 8. This last quantity tends to 0 according to Lemma 8 and Proposition 7. Using that convergence and (8), (9), (10) and (12) end the proof. \square

4.2 Proof of Theorem 2

Proof of Theorem 2. The theorem follows from Propositions 10, 11 and 13. \square

Proposition 10. *Let $P(u) = \int_{\mathcal{L}(t^-)} f(x) \mathbb{1}_{r(x) \geq u} dx$ for u in $[t^-, \sup r]$ and let $\Theta_s = [s_1, s_2]$ be a strict subinterval of Θ . Then, under the assumptions of Theorem 2, for every p in $P(\Theta_s)$ and almost surely for $n > n_0 = n_0(\omega)$, there exists a unique t_n such that $\int_{\mathcal{L}(t^-)} f_n(x) \mathbb{1}_{r_n(x) \geq t_n} dx = p$. Moreover, $t_n \in \Theta$.*

Proof. Define $P_n(u) = \int_{\mathcal{L}(t^-)} f_n(x) \mathbb{1}_{r_n(x) \geq u} dx$ and let $\Theta_z = [z_1, z_2]$ be an interval such that $\Theta_s \subsetneq \Theta_z \subsetneq \Theta$. We have $|P(u) - P_n(u)| \leq \int_{\mathcal{L}(t^-)} f(x) |\mathbb{1}_{r(x) \geq u} - \mathbb{1}_{r_n(x) \geq u}| dx + \int_{\mathcal{L}(t^-)} |f(x) - f_n(x)| dx \leq I_1 + I_2$. As almost surely $f_n \rightarrow f$ uniformly on $\mathcal{L}(t^-)$, we have $\int_{\mathcal{L}(t^-)} |f(x) - f_n(x)| dx \rightarrow 0$. For I_1 , we note that $\mathbb{1}_{r \geq u} - \mathbb{1}_{r_n \geq u}$ is only nonzero when $r \geq u > r_n$ or $r_n \geq u > r$. In the first case, we have $u - \epsilon_n \leq u \leq r = r_n + r - r_n \leq u + r - r_n \leq u + \epsilon_n$ where $\epsilon_n = \sup_{\mathcal{L}(t^-)} |r - r_n|$. In the second case, we also have $u - \epsilon_n \leq r \leq u + \epsilon_n$. Thus $\int_{\mathcal{L}(t^-)} f(x) |\mathbb{1}_{r(x) \geq u} - \mathbb{1}_{r_n(x) \geq u}| dx \leq \int_{\mathcal{L}(t^-)} f(x) \mathbb{1}_{r(x) \in [u - \epsilon_n, u + \epsilon_n]} dx \leq \lambda_f (r^{-1} [u - \epsilon_n, u + \epsilon_n]) \rightarrow 0$ by using Cadre (2006) and the fact that $\epsilon_n \rightarrow 0$ by Theorem II.3, Chapter 5 of Bosq and Lecoutre (1987). So, we conclude that $P_n(u) \rightarrow P(u)$ almost surely for every u in Θ_z .

Let t be such that $P(t) = p$. To show unique existence of a value t_n such that $P_n(t_n) = p$, we need to justify three points: a/ P_n is strictly decreasing on $[z_1, z_2]$; b/ $P_n(z_1) \geq p$ and $P_n(z_2) \leq p$ and c/ P_n is continue on $[z_1, z_2]$. Let $u < u'$ be two points on the segment $[z_1, z_2]$. Put $M_1 = r^{-1}]t^-, \sup r[$, then M_1 is an open subset of $\mathcal{L}(t^-)$. By continuity of r , there exists $x \in M_1$ and $\epsilon > 0$ such that $B(x, \epsilon) \subset M_1$ and $r(B(x, \epsilon)) \in]\frac{2u+u'}{3}, \frac{u+2u'}{3}[$. Uniform convergence of r_n to r implies that $r_n(B(x, \epsilon)) \in]u, u'[$. Because $\inf_{\mathcal{L}(t^-)} f_n > 0$ for n large enough, this leads to $\int_{\mathcal{L}(t^-)} f_n(x) \mathbb{1}_{r_n(x) \in [u, u'[} dx > 0$. Therefore, $P_n(u) > P_n(u')$ by definition of P_n , and a/ is proved. Using the same argument, $P(t)$ is a strictly decreasing function on $[z_1, z_2]$. Because $t \in \Theta_s \subsetneq \Theta_z$, we have $P(z_1) > P(t) = p$ and $P(z_2) < p$. With the previously proved convergence of P_n , one can deduce b/ and c/ can be proved using dominated convergence theorem and straightforward calculations which completes the proof. \square

Proposition 11. *Let $t \in \Theta_s$ be such that $\int_{\mathcal{L}(t^-)} f(x) \mathbb{1}_{r(x) \geq t} dx = p$ where $p \in P(\Theta_s)$. Then, under the assumptions of Theorem 2, we have $t_n \rightarrow t$ a.s.*

Proof. According to Proposition 10, t_n is contained in Θ for n large enough. We extract a subsequence (t_{n_k}) from (t_n) that converges to some t^* . To show that (t_{n_k}) converges to t , it is sufficient to show that $P(t_{n_k})$ converges to $P(t)$ because $P(u) = \int_{\mathcal{L}(t^-)} f(x) \mathbb{1}_{r(x) \geq u} dx$ is a continuous strictly decreasing function. We have $|P(t_{n_k}) - P(t)| \leq \int_{\mathcal{L}(t^-)} f(x) |\mathbb{1}_{r(x) \geq t_{n_k}} - \mathbb{1}_{r_{n_k}(x) \geq t_{n_k}}| dx + \int_{\mathcal{L}(t^-)} |f(x) - f_{n_k}(x)| dx$. Using the same

arguments as in the previous proof one can show that it converges to 0. Therefore we get $P(t_{n_k}) \rightarrow P(t)$ which leads to $t_{n_k} \rightarrow t$. As this conclusion is valid for any convergent subsequence of (t_n) , we have $t_n \rightarrow t$. \square

Lemma 12. *Under assumptions of Theorem 2, we have $\sqrt{nh^d} \int_{\mathcal{L}(t^-)} f(x) (\mathbb{1}_{r(x) \geq t} - \mathbb{1}_{r_n(x) \geq t}) dx \xrightarrow{\mathbb{P}} 0$.*

Proof. Write

$$\sqrt{nh^d} \int_{\mathcal{L}(t^-)} f(x) (\mathbb{1}_{r(x) \geq t} - \mathbb{1}_{r_n(x) \geq t}) dx = \sqrt{nh^d} \int_{\mathcal{L}(t^-)} f(x) (\mathbb{1}_{r(x) \geq t > r_n(x)} - \mathbb{1}_{r_n(x) \geq t > r(x)}) dx.$$

By Lemma 4.1 of Laloë and Servien (2013), the event $\left\{ \sup_{\mathcal{L}(t^-)} |r_n - r| \leq \Gamma \sqrt{\frac{\log n}{nh^d}} \right\}$ happens with probability tending to 1. Note that on this event, we have $\int_{\mathcal{L}(t^-)} f(x) \mathbb{1}_{r(x) \geq t > r_n(x)} dx = \int_{V_n^t} f(x) \mathbb{1}_{r_n(x) < t} dx$ so the lemma will be proved if we have that $\sqrt{nh^d} \int_{V_n^t} f(x) \mathbb{1}_{r_n(x) < t} dx - \sqrt{nh^d} \int_{V_n^t} f(x) \mathbb{1}_{r_n(x) \geq t} dx \xrightarrow{\mathbb{P}} 0$. Using the same arguments as in Proposition 7 and 9 with integrating with respect to the measure induced by f , we can show that each of these two terms tends in probability to the same constant. \square

Proposition 13. *Under assumptions of Theorem 2, we have $\sqrt{nh^d} [\lambda(\mathcal{L}_n(t_n) \Delta \mathcal{L}(t)) - \lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t))] \xrightarrow{\mathbb{P}} 0$, which implies that $\sqrt{nh^d} \lambda[\mathcal{L}_n(t_n) \Delta \mathcal{L}(t)] \xrightarrow{\mathbb{P}} \sqrt{\frac{2\bar{K}}{\pi}} \int_{r^{-1}\{t\}} \frac{\sqrt{q/f}}{\|\nabla r\|} d\mathcal{H}$.*

Proof. Straightforward calculations show that $\sqrt{nh^d} |\lambda(\mathcal{L}_n(t_n) \Delta \mathcal{L}(t)) - \lambda(\mathcal{L}_n(t) \Delta \mathcal{L}(t))| \leq c\sqrt{nh^d} \left| \int_{\mathcal{L}(t^-)} (f(x) - f_n(x)) \mathbb{1}_{r(x) \geq t} dx \right| + c\sqrt{nh^d} \left| \int_{\mathcal{L}(t^-)} f_n(x) (\mathbb{1}_{r(x) \geq t} - \mathbb{1}_{r_n(x) \geq t}) dx \right|$. The first term tends in probability to 0 by Lemma 4.2 in Cadre (2006). For the second term, we have $c\sqrt{nh^d} \left| \int_{\mathcal{L}(t^-)} f_n(x) (\mathbb{1}_{r(x) \geq t} - \mathbb{1}_{r_n(x) \geq t}) dx \right| \leq c \sup_{\mathcal{L}(t^-)} |f_n - f| \sqrt{nh^d} \int_{\mathcal{L}(t^-)} (\mathbb{1}_{r(x) \geq t > r_n(x)} + \mathbb{1}_{r_n(x) \geq t > r(x)}) dx + c\sqrt{nh^d} \left| \int_{\mathcal{L}(t^-)} f(x) (\mathbb{1}_{r(x) \geq t} - \mathbb{1}_{r_n(x) \geq t}) dx \right|$. Theorem 1 and Lemma 12 end the proof. \square

Bibliography

- Bosq, D. and Lecoutre, J. (1987). *Théorie de l'estimation fonctionnelle*. Collection Economie et statistiques avancées. Série Ecole nationale de la statistique et de l'administration et du Centre d'études des programmes économiques. Économica.
- Cadre, B. (2006). Kernel estimation of density level sets. *J. Multivariate Anal.*, 97(4):999–1023.
- Cavalier, L. (1997). Nonparametric estimation of regression level sets. *Statistics*, 29(2):131–160.
- Chen, Y.-C., Genovese, C. R., and Wasserman, L. (2017). Density level sets: Asymptotics, inference, and visualization. *Journal of the American Statistical Association*, 112(520):1684–1696.
- Di Bernardino, E., Laloë, T., Maume-Deschamps, V., and Prieur, C. (2013). Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory. *ESAIM Probab. Stat.*, 17:236–256.

- Di Bernardino, E., Laloë, T., and Servien, R. (2015). Estimating covariate functions associated to multivariate risks: a level set approach. *Metrika*, 78(5):497–526.
- Doss, C. R. and Weng, G. (2018). Bandwidth selection for kernel density estimators of multivariate level sets and highest density regions. *Electron. J. Statist.*, 12(2):4313–4376.
- Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3):1380–1403.
- Evans, L. C. and Gariepy, R. F. (2015). *Measure theory and fine properties of functions*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised edition.
- Hartigan, J. A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.*, 82(397):267–270.
- Janson, S. (1999). On concentration of probability. *Combinatorics, Probability and Computing*, 11:2002.
- Laloë, T. and Servien, R. (2013). Nonparametric estimation of regression level sets using kernel plug-in estimator. *J. Korean Statist. Soc.*, 42(3):301–311.
- Lee, J. M. (2003). *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.
- Mason, D. M. and Polonik, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.*, 19(3):1108–1142.
- Polonik, W. and Wang, Z. (2005). Estimation of regression contour clusters: an application of the excess mass approach to regression. *Journal of multivariate analysis*, 94:227–249.
- Qiao, W. (2018). Asymptotics and optimal bandwidth selection for nonparametric estimation of density level sets. *arXiv:1707.09697v2*.
- Qiao, W. (2019). Nonparametric estimation of surface integrals on level sets. *arXiv:1804.03601v2*.
- Scott, C. and Davenport, M. (2007). Regression level set estimation via cost-sensitive classification. *IEEE Trans. Signal Process.*, 55(6, part 1):2752–2757.
- Tsybakov, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.*, 25(3):948–969.
- Willett, R. M. and Nowak, R. D. (2007). Minimax optimal level-set estimation. *IEEE Trans. Image Process.*, 16(12):2965–2979.