Estimation of density level sets with a given probability content

Benoît CADRE[a], Bruno PELLETIER[b], and Pierre PUDLO[c]

[a] IRMAR, ENS Cachan Bretagne, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman, 35170 Bruz, France
cadre@bretagne.ens-cachan.fr

[b] IRMAR, Université Rennes 2, CNRS, UEB
Campus Villejean
Place du Recteur Henri Le Moal
35043 Rennes Cedex, France
bruno.pelletier@univ-rennes2.fr

[c] I3M, UMR CNRS 5149
Université Montpellier II, CC 051
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
pudlo@math.univ-montp2.fr

Abstract

Given a random vector X valued in $\mathbb{R}^d$ with density f and an arbitrary probability number $p \in (0;1)$, we consider the estimation of the upper level set $\{f \geq t^{(p)}\}$ of f corresponding to probability content p, i.e., such that the probability that X belongs to $\{f \geq t^{(p)}\}$ is equal to p. Based on an i.i.d. random sample $X_1, \ldots, X_n$ drawn from f, we define the plug-in level set estimate $\{\hat{f}_n \geq t_n^{(p)}\}$, where $t_n^{(p)}$ is a random threshold depending on the sample, and where $\hat{f}_n$ is a nonparametric kernel density estimate based on the same sample. We establish the exact convergence rate of the Lebesgue measure of the symmetric difference between the estimated and actual level sets.

Index Terms — Kernel estimate; Density level sets; Nonparametric statistics.
AMS 2000 Classification — 62H30, 62H12.

1 Introduction

Let X be a $\mathbb{R}^d$-valued random variable with density f. For any $t \geq 0$, the t-upper-level set $\mathscr{L}(t)$ of f is defined by

$$\mathscr{L}(t) = \{f \geq t\} = \{x \in \mathbb{R}^d : f(x) \geq t\}. \tag{1.1}$$

Density level sets are used in a variety of scientific areas, including statistics and machine learning, medical imaging, computer vision, or remote sensing, and with applications to unsupervised classification/clustering, pattern recognition, anomaly or novelty detection for instance. Motivated by these applications, the theory behind their estimation has developed significantly in the recent years. Excess-mass level set estimates are studied in Hartigan (1987), Muller and Sawitzki (1991), Nolan (1991), Polonik (1995, 1997). Other popular level set estimates are the plug-in level set estimates, formed by replacing the density f with a density estimate $\hat{f}_n$ in (1.1). Under some assumptions, consistency and rates of convergence (for the volume of the symmetric difference) have been established in Baillo et al. (2000, 2001), Baillo (2003), and an exact convergence rate is obtained in Cadre (2006). Recently, Mason and Polonik (2009) derive the asymptotic normality of the volume of the symmetric difference for kernel plug-in level set estimates; see also related works in Molchanov (1998), Cuevas et al. (2006).

So far, most theoretical works on the subject have focused on the estimation of a density level set at a fixed threshold t in (1.1). In practice, this threshold has to be interpreted as a resolution level of the analysis for the application under consideration, but its meaning is more easily understood in terms of probability content as follows: given a probability number $p \in (0;1)$, define $t^{(p)}$ as the largest threshold such that the probability of $\mathscr{L}(t^{(p)})$ is greater than p, i.e.,

$$t^{(p)} = \sup\left\{t \geq 0 : \mathbb{P}\big(X \in \mathscr{L}(t)\big) \geq p\right\}. \tag{1.2}$$

Note that $\mathbb{P}(X \in \mathscr{L}(t^{(p)})) = p$ whenever $\mathbb{P}(f(X) = t^{(p)}) = 0$. Hence when p is close to one, the upper level set is close to the support of the distribution. To the contrary, when p is small, $\mathscr{L}(t^{(p)})$ is a small domain concentrated around the largest modes of f.

In Cadre (2006), a consistent estimate of $t^{(p)}$ is defined as a solution in t of the following equation

$$\int_{\{\hat{f}_n \geq t\}} \hat{f}_n(x)\mathrm{d}x = p, \tag{1.3}$$

where $\hat{f}_n$ is a nonparametric density estimate of f based on an i.i.d random sample $X_1, \cdots, X_n$ drawn from f. From a numerical perspective, solving for t in (1.3) at a fixed p would require multiple evaluations of integrals which would become intractable in practice, especially in large dimensions. On the other hand, equation (1.2) hints at a definition of the threshold $t^{(p)}$ as a quantile of the distribution of the (real-valued) random variable $f(X)$. Based on that, and following an idea that goes back to Hyndman (1996), we propose in this paper to consider the estimate $t_n^{(p)}$ defined as the $(1-p)$-quantile of the empirical distribution of $\hat{f}_n(X_1), \dots, \hat{f}_n(X_n)$. This estimate may be computed easily from an order statistic. We focus on the case where $\hat{f}_n$ is a nonparametric kernel density estimate, and we consider the plug-in level set estimate $\mathscr{L}_n(t_n^{(p)})$ defined by

$$\mathscr{L}_n(t_n^{(p)}) = \{\hat{f}_n \geq t_n^{(p)}\}.$$

Our main result (Theorem 2.1) states that, under suitable conditions, $\mathscr{L}_n(t_n^{(p)})$ is consistent in the sense that

$$\sqrt{nh_n^d}\,\lambda\left(\mathscr{L}_n(t_n^{(p)}) \Delta \mathscr{L}(t^{(p)})\right) \xrightarrow{\mathbb{P}} C_f^{(p)},$$

where $C_f^{(p)}$ is an explicit constant depending on f and p, where $A \Delta B$ denotes the symmetric difference of two sets A and B, and where λ denotes the Lebesgue measure on $\mathbb{R}^d$. We also show that the limit constant $C_f^{(p)}$ may also be consistently estimated. An analogous result is obtained in Corollary 2.1 in Cadre (2006) by using the threshold estimate defined in (1.3), but the proofs of our consistency results require refined arguments.

The paper is organized as follows. In section 2, we introduce the estimates of the level set along with some notation. The main consistency result is stated in section 3. Proofs are gathered in section 4. Finally, several auxiliary results for the proofs are postponed in the Appendices, at the end of the paper.

2 Level set estimation

2.1 Notation and definitions

For any $t \geq 0$, let $\mathscr{L}(t) = \{f \geq t\}$ be the t-upper level set of f. Let H be the function defined for all $t \geq 0$ by

$$H(t) = \mathbb{P}(f(X) \leq t).$$

Given a real number p in $(0;1)$, let $t^{(p)}$ be the $(1-p)$-quantile of the law of $f(X)$, i.e.

$$t^{(p)} = \inf\{t \in \mathbb{R} : H(t) \geq 1-p\}. \tag{2.1}$$

By definition, the set $\mathscr{L}(t^{(p)})$ has μ-coverage no more than p, where μ denotes the law of X. Consequently, whenever H is continuous at $t^{(p)}$, $\mathscr{L}(t^{(p)})$ has μ-coverage equal to p, i.e., we have $\mu(\mathscr{L}(t^{(p)})) = p$ in this case.

Given a density estimate $\hat{f}_n$ based on an i.i.d. random sample $X_1,\ldots,X_n$ drawn from f, plug-in estimates of $\mathscr{L}(t)$, $H(t)$, and $t^{(p)}$ may be defined by replacing f with $\hat{f}_n$ in the definitions above. For any $t \geq 0$, define the estimates $\mathscr{L}_n(t)$ and $H_n(t)$ by

$$\mathscr{L}_n(t) = \{\hat{f}_n \geq t\} \quad \text{and} \quad H_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\hat{f}_n(X_i) \leq t\},$$

and the estimate $t_n^{(p)}$, for any p in $(0;1)$, by

$$t_n^{(p)} = \inf\{t \in \mathbb{R} : H_n(t) \geq 1-p\}. \tag{2.2}$$

Combining these estimates leads to the estimate $\mathscr{L}_n(t_n^{(p)})$ of $\mathscr{L}(t^{(p)})$ at fixed probability level p. Note that all these plug-in estimates are based on the same and whole sample $X_1,\ldots,X_n$, without splitting. In comparison with the estimator of $t^{(p)}$ defined as a solution of (1.3), the estimate $t_n^{(p)}$ is easily computed, by considering the order statistic induced by the sample $\hat{f}_n(X_1),\ldots, \hat{f}_n(X_n)$.

2.2 Main result

First of all, whenever f is of class C^1, let $\mathscr{T}_0$ be the subset of the range of f defined by

$$\mathscr{T}_0 = \left\{ t \in (0;\sup f) : \inf_{\{f=t\}} \|\nabla f\| = 0 \right\}.$$

This set naturally arises when considering the distribution of $f(X)$. Indeed, the Implicit Function Theorem implies that $\mathscr{T}_0$ contains the set of points in $(0;\sup f)$ which charges the distribution of $f(X)$. We shall assume throughout that the density f satisfies the following conditions.

Assumption 1 [on f]

(i) The density f is of class C^2 with bounded second-order partial derivatives, and $f(x) \to 0$ as $\|x\| \to \infty$.

(ii) $\mathscr{T}_0$ has Lebesgue measure 0.

(iii) $\lambda(\{f = t\}) = 0$ for all $t > 0$.

By Assumption 1-*(i)*, the upper t-level set $\mathscr{L}(t)$ is compact for all $t > 0$, as well as its boundary $\{f = t\}$. Assumption 1-*(iii)*, which ensures the continuity of H, roughly means that each flat part of f has a null volume ; it was first introduced in Polonik (1995). Moreover, it is proved in Lemma A.1 that under Assumption 1-*(i)*, we have $\mathscr{T}_0 = f(\mathscr{X}) \setminus \{0; \sup f\}$, where $\mathscr{X} = \{\nabla f = 0\}$ is the set of critical points of f. Suppose in addition that f is of class C^k, with $k \geq d$. Then, Sard's Theorem (see, e.g., Aubin, 2000) ensures that the Lebesgue measure of $f(\mathscr{X})$ is 0, hence implying Assumption 1-*(ii)*.

In the sequel, $\hat{f}_n$ is the nonparametric kernel density estimate of f with kernel K and bandwidth sequence (h_n) defined by

$$\hat{f}_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right). \tag{2.3}$$

We shall assume that the kernel K satisfies the following assumption.

Assumption 2 [on K]
K is a density on $\mathbb{R}^d$ with radial symmetry:

$$K(x) = \Phi\left(\|x\|\right),$$

where $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a decreasing function with compact support.

Under Assumption 2, sharp almost-sure convergence rates on $\hat{f}_n - f$ can be established ; see for instance Giné and Guillou (2002) and Einmahl and Mason (2005).

Denote by $\mathscr{H}$ the $(d-1)$-dimensional Hausdorff measure (see, e.g., Evans and Gariepy, 1992), which agrees with the ordinary $(d-1)$-dimensional surface measure on nice sets, and by $\|.\|_2$ the L_2-norm. We are now in a position to state our main result.

Theorem 2.1. *Suppose that f satisfies Assumption 1 and that $d \geq 2$. Let $\hat{f}_n$ be the nonparametric kernel density estimate (2.3) with kernel K satisfying Assumption 2, and bandwidth sequence (h_n) satisfying $\frac{nh_n^d}{(\log n)^{16}} \to \infty$ and $nh_n^{d+4}(\log n)^2 \to 0$ as $n \to \infty$. Then, for almost all $p \in (0;1)$, we have*

$$\sqrt{nh_n^d}\, \lambda\left(\mathscr{L}_n(t_n^{(p)}) \Delta \mathscr{L}(t^{(p)})\right) \xrightarrow{\mathbb{P}} \sqrt{\frac{2}{\pi}}\, \|K\|_2\, t^{(p)} \int_{\{f = t^{(p)}\}} \frac{1}{\|\nabla f\|} \mathrm{d}\mathscr{H}.$$

Under the conditions on (h_n), the convergence rate established in Theorem 2.1 is equal to $1/\sqrt{nh_n^d}$. Hence if (h_n) is restricted to a sequence of the form n^{-s} for

$s > 0$, then the best convergence rate must be necessarily slower than $O(n^{-2/(d+4)})$, which is the same convergence rate as the one obtained in Corollary 2.1 in Cadre (2006) when the estimate of $t^{(p)}$ is defined as a solution of (1.3).

Note that the deterministic limit in Theorem 2.1 depends on the unknown density f. However, one can prove that if (α_n) is a sequence of positive numbers tending to 0 and such that $\alpha_n^2 n h_n^d / (\log n)^2 \to \infty$, then, for almost all $p \in (0;1)$,

$$\frac{t_n^{(p)}}{\alpha_n} \lambda \left(\mathscr{L}_n(t_n^{(p)}) \setminus \mathscr{L}_n(t_n^{(p)} + \alpha_n) \right) \xrightarrow{\mathbb{P}} t^{(p)} \int_{\{f=t^{(p)}\}} \frac{1}{\|\nabla f\|} \mathrm{d}\mathscr{H}.$$

The proof of the above result is similar to the one of Lemma 4.6 in Cadre (2006), using our Proposition 3.4. Combined with Theorem 2.1, we then have, for almost all $p \in (0;1)$,

$$\frac{\alpha_n \sqrt{n h_n^d}}{t_n^{(p)} \lambda \left(\mathscr{L}_n(t_n^{(p)}) \setminus \mathscr{L}_n(t_n^{(p)} + \alpha_n) \right)} \lambda \left(\mathscr{L}_n(t_n^{(p)}) \Delta \mathscr{L}(t^{(p)}) \right) \xrightarrow{\mathbb{P}} \sqrt{\frac{2}{\pi}} \|K\|_2,$$

which yields a feasible way to estimate $\lambda(\mathscr{L}_n(t_n^{(p)}) \Delta \mathscr{L}(t^{(p)}))$.

Remark 2.2. *As pointed out by a referee, by inspecting the proofs, the statement of Theorem 2.1 remains valid for any estimate of the threshold $t^{(p)}$ provided that it converges at the rate given by Proposition 3.4.*

Remark 2.3. *According to Proposition A.2 in Appendix A, on any interval $I \subset (0; \sup f)$ with $I \cap \mathscr{T}_0 = \emptyset$, the random variable $f(X)$ has a density on I, which is given by*

$$g(t) = t \int_{\{f=t\}} \frac{1}{\|\nabla f\|} \mathrm{d}\mathscr{H}, \quad t \in I.$$

Thus the normalized distance between $\mathscr{L}_n(t_n^{(p)})$ and $\mathscr{L}(t^{(p)})$ in Theorem 2.1 corresponds to the density g at point $t^{(p)}$, up to a multiplicative constant.

3 Proofs

In all the proofs, we assume that Assumption 1 on the underlying density f and Assumption 2 on the kernel K are satisfied.

3.1 Auxiliary results

First note that under Assumption 1, H is a bijection from $(0;\sup f)$ to $(0;1)$. Indeed, Assumption 1-*(iii)* implies that H is a continuous function. Moreover, under Assumption 1-*(i)*, H is increasing: for suppose it were not, then for some $t \geq 0$ and some $\varepsilon > 0$,

$$0 = H(t+\varepsilon) - H(t) = \int_{\{t<f\leq t+\varepsilon\}} f \mathrm{d}\lambda,$$

which is impossible, because $\lambda(\{t < f < t+\varepsilon\}) > 0$. Then we denote by G the inverse of H restricted to $(0;\sup f)$.

Lemma 3.1. *The function G is almost everywhere differentiable.*

Proof. As stated above, H is increasing. Hence, by the Lebesgue derivation Theorem, for almost all t, H is differentiable with derivative $H'(t) > 0$. Thus, G is almost everywhere differentiable. □

The Levy metric $d_{\mathscr{L}}$ between any non-decreasing bounded real-valued functions φ_1, φ_2 on $\mathbb{R}$ is defined by

$$d_{\mathscr{L}}(\varphi_1,\varphi_2) = \inf\big\{\theta > 0 : \forall x \in \mathbb{R}, \varphi_1(x-\theta) - \theta \leq \varphi_2(x) \leq \varphi_1(x+\theta) + \theta\big\},$$

(see, e.g., Billingsley, 1995, 14.5). Recall that convergence in distribution is equivalent to convergence of the underlying distribution functions for the metric $d_{\mathscr{L}}$.

Lemma 3.2. *Let x_0 be a real number, and let φ_1 be an increasing function with a derivative at point x_0. There exists $C > 0$ such that, for any increasing function φ_2 with $d_{\mathscr{L}}(\varphi_1,\varphi_2) \leq 1$,*

$$|\varphi_1(x_0) - \varphi_2(x_0)| \leq C d_{\mathscr{L}}(\varphi_1,\varphi_2).$$

Proof. Let θ be any positive number such that, for all $x \in \mathbb{R}$,

$$\varphi_1(x-\theta) - \theta \leq \varphi_2(x) \leq \varphi_1(x+\theta) + \theta. \tag{3.1}$$

Since φ_1 is differentiable at x_0,

$$\varphi_1(x_0 \pm \theta) = \varphi_1(x_0) \pm \theta\varphi_1'(x_0) + \theta\psi_{\pm}(\theta) \tag{3.2}$$

where each function $\psi_\pm$ satisfies $\psi_\pm(\theta) \to 0$ when $\theta \to 0^+$. Using (3.1) and (3.2), we obtain

$$-\theta(\varphi_1'(x_0)+1)+\theta\psi_-(\theta) \leq \varphi_2(x_0)-\varphi_1(x_0) \leq \theta(\varphi_1'(x_0)+1)+\theta\psi_+(\theta).$$

Taking the infimum over θ satisfying (3.1) gives the announced result with any C such that, for all $\delta \leq 1$,

$$\left|\varphi_1'(x_0)+1\right| + \max\left(|\psi_-(\delta)|, |\psi_+(\delta)|\right) \leq C. \qquad \square$$

Let $\mathscr{L}^\ell(t)$ denote the lower t-level set of the unknown density f, i.e., $\mathscr{L}^\ell(t) = \{x \in \mathbb{R}^d : f(x) \leq t\}$. Moreover, we set

$$V_n = \sup_{t \geq 0} \left|\mu_n\left(\mathscr{L}^\ell(t)\right) - \mu\left(\mathscr{L}^\ell(t)\right)\right|, \tag{3.3}$$

where $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ is the empirical measure indexed by the sample, δ_x denoting the Dirac measure at point x. The next lemma borrows elements from the Vapnik-Chervonenkis theory; we refer the reader to Devroye et al. (1996) for materials on the subject.

Lemma 3.3. *There exists a constant C such that, for all $\eta > 0$, we have*

$$\mathbb{P}(V_n \geq \eta) \leq Cn \exp\left(-n\eta^2/32\right).$$

Proof. Let $\mathscr{A}$ be the collection of lower level sets, namely

$$\mathscr{A} = \{\mathscr{L}^\ell(t), t \geq 0\}.$$

Observe that the Vapnik-Chervonenkis dimension (see, e.g., Devroye et al., 1996, Definition 12.1) of $\mathscr{A}$ is 1: for any set with two elements $\{x_1, x_2\}$, where, without loss of generality, $f(x_1) \leq f(x_2)$, it is impossible to express the subset $\{x_2\}$ as an intersection of $\{x_1, x_2\}$ with an element of $\mathscr{A}$. Then, by the Vapnik-Chervonenkis inequality (see, e.g., Devroye et al., 1996, Theorem 12.5), combined with Theorem 13.3 in Devroye et al. (1996), we obtain the stated result. $\square$

3.2 Asymptotics for the threshold estimate

Proposition 3.4. *Under the conditions of Theorem 2.1, for almost all $p \in (0;1)$, we have*

$$\frac{\sqrt{nh_n^d}}{\log n}\left|t_n^{(p)} - t^{(p)}\right| \xrightarrow{\mathbb{P}} 0.$$

Proof. We first proceed to bound $d_{\mathscr{L}}(H,H_n)$. We have $H_n(t) = \mu_n\big(\mathscr{L}_n^\ell(t)\big)$, and $H(t) = \mu\big(\mathscr{L}^\ell(t)\big)$ where $\mathscr{L}_n^\ell(t) = \{x \in \mathbb{R}^d : \hat{f}_n(x) \leq t\}$ and $\mathscr{L}^\ell(t) = \{x \in \mathbb{R}^d : f(x) \leq t\}$. The triangle inequality gives

$$\mathscr{L}^\ell\left(t - \|\hat{f}_n - f\|_\infty\right) \subset \mathscr{L}_n^\ell(t) \subset \mathscr{L}^\ell\left(t + \|\hat{f}_n - f\|_\infty\right),$$

which, applying μ_n, yields

$$\mu_n\left(\mathscr{L}^\ell\left(t - \|\hat{f}_n - f\|_\infty\right)\right) \leq H_n(t) \leq \mu_n\left(\mathscr{L}^\ell\left(t + \|\hat{f}_n - f\|_\infty\right)\right).$$

Moreover, by definition of V_n in (3.3), we have

$$H(s) - V_n \leq \mu_n\big(\mathscr{L}^\ell(s)\big) \leq H(s) + V_n,$$

for any real number s. The two last inequalities give

$$H(t - \|\hat{f}_n - f\|_\infty) - V_n \leq H_n(t) \leq H(t + \|\hat{f}_n - f\|_\infty) + V_n.$$

Using the fact that H is non-decreasing, we obtain

$$d_{\mathscr{L}}(H,H_n) \leq \max\left(\|\hat{f}_n - f\|_\infty, V_n\right). \tag{3.4}$$

By Lemma 3.1, G is almost everywhere differentiable. Let us fix $p \in (0;1)$ such that G is differentiable at $1-p$, and observe that $G(1-p) = t^{(p)}$. Denote by G_n the pseudo-inverse of H_n, i.e.

$$G_n(s) = \inf\{t \geq 0 : H_n(t) \geq s\},$$

and remark that $G_n(1-p) = t_n^{(p)}$. Moreover, we always have $d_{\mathscr{L}}(H,H_n) \leq 1$ because $0 \leq H(t) \leq 1$ and $0 \leq H_n(t) \leq 1$ for all $t \in \mathbb{R}$. Hence, since $d_{\mathscr{L}}(H,H_n) = d_{\mathscr{L}}(G,G_n)$, we obtain from Lemma 3.2 that for some constant C,

$$\left|t_n^{(p)} - t^{(p)}\right| = |G_n(1-p) - G(1-p)| \leq C d_{\mathscr{L}}(H,H_n).$$

Then, from (3.4) and Lemma 3.3, it follows that, for almost all $p \in (0;1)$ and for all $\eta > 0$, we have

$$\mathbb{P}\left(|t_n^{(p)} - t^{(p)}| \geq \eta\right) \leq \mathbb{P}\left(\|\hat{f}_n - f\|_\infty \geq C_1\eta\right) + C_2 n \exp\left(-nC_1\eta^2\right), \tag{3.5}$$

where C_1 and C_2 are positive constants.

Under Assumption 2 on the kernel, and since f is bounded under Assumption 1-(i), the conditions of Theorem 1 in Einmahl and Mason (2005) are satisfied, implying that $\nu_n\|\hat{f}_n - \mathbb{E}\hat{f}_n\|_\infty \to 0$ almost surely as $n \to \infty$, where (v_n) is any sequence satisfying $v_n = o(\sqrt{nh_n^d/\log n})$. Moreover, under Assumption 1-*(i)*, it may be easily shown that $\|\mathbb{E}\hat{f}_n - f\|_\infty = O(h_n^2)$. Hence it follows that for all $\eta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(v_n\|\hat{f}_n - f\|_\infty \geq \eta\big) = 0 \tag{3.6}$$

Then, by the concentration inequality (3.5), we obtain the desired result. □

3.3 Asymptotics for level sets

Lemma 3.5. *For almost all* $p \in (0;1)$*, we have*

(i) $(\log n)\lambda\left(\mathscr{L}_n(t^{(p)})\Delta\mathscr{L}(t^{(p)})\right) \xrightarrow{\mathbb{P}} 0$ *and*

(ii) $(\log n)\lambda\left(\mathscr{L}_n(t_n^{(p)})\Delta\mathscr{L}(t^{(p)})\right) \xrightarrow{\mathbb{P}} 0.$

Proof. We only prove (ii). Set $\varepsilon_n = \log n/\sqrt{nh_n^d}$, which tends to 0 as n goes to infinity under Assumption 2. Moreover, let $\mathscr{N}_1$, $\mathscr{N}_2$ be defined as

$$\begin{aligned} \mathscr{N}_1^c &= \left\{p \in (0;1) : \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\lambda\left(\left\{t^{(p)} - \varepsilon \leq f \leq t^{(p)} + \varepsilon\right\}\right) \text{ exists}\right\}; \\ \mathscr{N}_2^c &= \left\{p \in (0;1) : \frac{1}{\varepsilon_n}|t_n^{(p)} - t^{(p)}| \xrightarrow{\mathbb{P}} 0\right\}. \end{aligned}$$

Both $\mathscr{N}_1$ and $\mathscr{N}_2$ have a null Lebesgue measure: the first property is a consequence of the Lebesgue derivation Theorem and the fact that H is a bijection from $(0;\sup f)$ onto $(0;1)$. The second one is a direct consequence of Proposition 3.4.

Hence, one only needs to prove the lemma for all $p \in \mathscr{N}_1^c \cap \mathscr{N}_2^c$. We now fix p in this set, and we denote by Ω_n the event

$$\Omega_n = \{\|\hat{f}_n - f\|_\infty \leq \varepsilon_n\} \cap \{|t_n^{(p)} - t^{(p)}| \leq \varepsilon_n\}.$$

Since $\mathbb{P}(\Omega_n) \to 1$ by (3.6), it suffices to show that the stated convergence holds on the event Ω_n. Simple calculations yield

$$\begin{aligned} &\lambda\left(\mathscr{L}_n(t_n^{(p)})\Delta\mathscr{L}(t^{(p)})\right) \\ &\quad = \lambda\left(\left\{\hat{f}_n \geq t_n^{(p)}; f < t^{(p)}\right\}\right) + \lambda\left(\left\{\hat{f}_n < t_n^{(p)}; f \geq t^{(p)}\right\}\right). \end{aligned}$$

But, on the event Ω_n, we have $\hat{f}_n + \varepsilon_n \geq f \geq \hat{f}_n - \varepsilon_n$ and $t_n^{(p)} - \varepsilon_n \leq t^{(p)} \leq t_n^{(p)} + \varepsilon_n$. Consequently, if n is large enough,

$$\begin{aligned}
&\lambda\left(\mathscr{L}_n(t_n^{(p)})\Delta\mathscr{L}(t^{(p)})\right) \\
&\quad \leq \lambda\left(\left\{t^{(p)} - 2\varepsilon_n \leq f < t^{(p)}\right\}\right) + \lambda\left(\left\{t^{(p)} \leq f \leq t^{(p)} + 2\varepsilon_n\right\}\right) \\
&\quad = \lambda\left(\left\{t^{(p)} - 2\varepsilon_n \leq f \leq t^{(p)} + 2\varepsilon_n\right\}\right) \\
&\quad \leq C\varepsilon_n,
\end{aligned}$$

for some constant C, because $p \in \mathscr{N}_1^c$ and $\varepsilon_n \to 0$. The last inequality proves the lemma, since under the conditions of Theorem 2.1, we have $\varepsilon_n \log n \to 0$ as n goes to infinity. □

In the sequel, $\tilde{\mu}_n$ denotes the smoothed empirical measure, which is the random measure with density $\hat{f}_n$, defined for all Borel set $A \subset \mathbb{R}^d$ by

$$\tilde{\mu}_n(A) = \int_A \hat{f}_n \mathrm{d}\lambda.$$

Lemma 3.6. *For almost all $p \in (0;1)$,*

(i) $\sqrt{nh_n^d}\left\{\tilde{\mu}_n(\mathscr{L}_n(t^{(p)})) - \mu(\mathscr{L}_n(t^{(p)}))\right\} \xrightarrow{\mathbb{P}} 0$ *and*

(ii) $\sqrt{nh_n^d}\left\{\tilde{\mu}_n(\mathscr{L}_n(t_n^{(p)})) - \mu(\mathscr{L}_n(t_n^{(p)}))\right\} \xrightarrow{\mathbb{P}} 0.$

Proof. We only prove (ii). Fix $p \in (0;1)$ such that the result in Lemma 3.5 holds. Observe that

$$\begin{aligned}
&\left|\tilde{\mu}_n(\mathscr{L}_n(t_n^{(p)})) - \mu(\mathscr{L}_n(t_n^{(p)}))\right| \\
&\quad \leq \int_{\mathscr{L}_n(t_n^{(p)})\Delta\mathscr{L}(t^{(p)})} |\hat{f}_n - f|\mathrm{d}\lambda + \left|\int_{\mathscr{L}(t^{(p)})} (\hat{f}_n - f)\mathrm{d}\lambda\right| \\
&\quad \leq \lambda\left(\mathscr{L}_n(t_n^{(p)})\Delta\mathscr{L}(t^{(p)})\right)\|\hat{f}_n - f\|_\infty + \left|\int_{\mathscr{L}(t^{(p)})} (\hat{f}_n - f)\mathrm{d}\lambda\right|. \quad (3.7)
\end{aligned}$$

Recall that K is a radial function with compact support. Since $nh_n^{d+4} \to 0$ and $\mathscr{L}(t^{(p)})$ is compact for all $p \in (0;1)$, it is a classical exercise to prove that for all $p \in (0;1)$,

$$\sqrt{nh_n^d}\int_{\mathscr{L}(t^{(p)})} (\hat{f}_n - f)\mathrm{d}\lambda \xrightarrow{\mathbb{P}} 0. \quad (3.8)$$

(see, e.g., Cadre, 2006, Lemma 4.2). Moreover, by (3.6) and Lemma 3.5,

$$\sqrt{nh_n^d}\,\lambda\left(\mathscr{L}_n(t_n^{(p)})\Delta\mathscr{L}(t^{(p)})\right)\|\hat{f}_n - f\|_\infty \xrightarrow{\mathbb{P}} 0. \tag{3.9}$$

The inequalities (3.7), (3.8) and (3.9) prove the assertion of the lemma. □

Lemma 3.7. *For almost all* $p \in (0;1)$,

$$\sqrt{nh_n^d}\Big\{\mu\left(\mathscr{L}_n(t_n^{(p)})\right) - \mu\left(\mathscr{L}(t^{(p)})\right)\Big\} \xrightarrow{\mathbb{P}} 0.$$

Proof. Let $\varepsilon_n = \log n/\sqrt{nh_n^d}$ and $\mathscr{N}$ be the set defined by

$$\mathscr{N}^c = \left\{p \in (0;1) : \frac{1}{\varepsilon_n}|t_n^{(p)} - t^{(p)}| \xrightarrow{\mathbb{P}} 0\right\}.$$

By Proposition 3.4, $\mathscr{N}$ has a null Lebesgue measure. If $p \in \mathscr{N}^c$, then we have $t^{(p)} - \varepsilon_n \le t_n^{(p)} \le t^{(p)} + \varepsilon_n$ on an event A_n such that $\mathbb{P}(A_n) \to 1$ as n goes to infinity. But, on A_n:

$$\mathscr{L}_n(t^{(p)} + \varepsilon_n) \subset \mathscr{L}_n(t_n^{(p)}) \subset \mathscr{L}_n(t^{(p)} - \varepsilon_n).$$

Consequently, one only needs to prove that for almost all $p \in \mathscr{N}^c$, the two following results hold:

$$\sqrt{nh_n^d}\left\{\mu\left(\mathscr{L}_n(t^{(p)} \pm \varepsilon_n)\right) - \mu\left(\mathscr{L}(t^{(p)})\right)\right\} \xrightarrow{\mathbb{P}} 0. \tag{3.10}$$

For the sake of simplicity, we only prove the "+" part of (3.10).

One can follow the arguments of the proofs of Propositions 3.1 and 3.2 in Cadre (2006), to obtain that for almost all $p \in \mathscr{N}^c$, there exists $J = J(p)$ with

$$\sqrt{nh_n^d}\,\mu\left(\mathscr{L}_n(t^{(p)} + \varepsilon_n) \cap \mathscr{V}_n\right) \xrightarrow{\mathbb{P}} J \quad \text{and}$$
$$\sqrt{nh_n^d}\,\mu\left(\mathscr{L}_n(t^{(p)} + \varepsilon_n)^c \cap \bar{\mathscr{V}}_n\right) \xrightarrow{\mathbb{P}} J,$$

where we set

$$\mathscr{V}_n = \left\{t^{(p)} - \varepsilon_n \le f < t^{(p)}\right\} \quad \text{and} \quad \bar{\mathscr{V}}_n = \left\{t^{(p)} \le f < t^{(p)} + 3\varepsilon_n\right\}.$$

Thus, for almost all $p \in \mathscr{N}^c$

$$\sqrt{nh_n^d}\Big\{\mu\left(\mathscr{L}_n(t^{(p)} + \varepsilon_n) \cap \mathscr{V}_n\right) - \mu\left(\mathscr{L}_n(t^{(p)} + \varepsilon_n)^c \cap \bar{\mathscr{V}}_n\right)\Big\} \xrightarrow{\mathbb{P}} 0. \tag{3.11}$$

Now let $p \in \mathscr{N}^c$ satisfying the above result, and set $\Omega_n = \{\|\hat{f}_n - f\|_\infty \leq 2\varepsilon_n\}$. By (3.6), $\mathbb{P}(\Omega_n) \to 1$ hence one only needs to prove that the result holds on the event Ω_n. But, on Ω_n,

$$\begin{aligned}
&\mu\left(\mathscr{L}_n(t^{(p)}+\varepsilon_n)\right) - \mu\left(\mathscr{L}(t^{(p)})\right) \\
&\quad = \mu\left(\left\{\hat{f}_n \geq t^{(p)}+\varepsilon_n; f < t^{(p)}\right\}\right) - \mu\left(\left\{\hat{f}_n < t^{(p)}+\varepsilon_n; f \geq t^{(p)}\right\}\right) \\
&\quad = \mu\left(\mathscr{L}_n(t^{(p)}+\varepsilon_n) \cap \mathscr{V}_n\right) - \mu\left(\mathscr{L}_n(t^{(p)}+\varepsilon_n)^c \cap \bar{\mathscr{V}}_n\right).
\end{aligned}$$

Consequently, by (3.11), we have on Ω_n

$$\sqrt{nh_n^d}\left\{\mu\left(\mathscr{L}_n(t^{(p)}+\varepsilon_n)\right) - \mu\left(\mathscr{L}(t^{(p)})\right)\right\} \xrightarrow{\mathbb{P}} 0.$$

This proves the "+" part of (3.10). The "−" part is obtained with similar arguments. □

3.4 Proof of Theorem 2.1

Let $t_0 \in \mathscr{T}_0^c$. Since f is of class C^2, there exists an open set $I(t_0)$ containing t_0 such that

$$\inf_{\{f \in I(t_0)\}} \|\nabla f\| > 0.$$

Thus, by Theorem 2.1 in Cadre (2006), we have, for almost all $t \in I(t_0)$,

$$\sqrt{nh_n^d}\,\lambda\left(\mathscr{L}_n(t)\Delta\mathscr{L}(t)\right) \xrightarrow{\mathbb{P}} \sqrt{\frac{2}{\pi}}\|K\|_2\, t \int_{\{f=t\}} \frac{1}{\|\nabla f\|} \mathrm{d}\mathscr{H}.$$

Recalling now that the Lebesgue measure of $\mathscr{T}_0$ is 0, and that H is a bijection from $(0;\sup f)$ onto $(0;1)$, it follows that the above result remains true for almost all $p \in (0;1)$, with $t^{(p)}$ instead of t. As a consequence, one only needs to prove that for almost all $p \in (0;1)$, $\sqrt{nh_n^d}D_n(p) \to 0$ in probability, where

$$D_n(p) = \lambda\left(\mathscr{L}_n(t_n^{(p)})\Delta\mathscr{L}(t^{(p)})\right) - \lambda\left(\mathscr{L}_n(t^{(p)})\Delta\mathscr{L}(t^{(p)})\right).$$

After some calculations, $D_n(p)$ may be expressed as

$$D_n(p) = \int_{\mathbb{R}^d} \mathbf{1}\{t_n^{(p)} \leq \hat{f}_n < t^{(p)}\} g\, \mathrm{d}\lambda - \int_{\mathbb{R}^d} \mathbf{1}\{t^{(p)} \leq \hat{f}_n < t_n^{(p)}\} g\, \mathrm{d}\lambda,$$

where $g = 1 - 2\mathbf{1}\{f \geq t^{(p)}\}$. For simplicity, we assume that $0 < t_n^{(p)} \leq t^{(p)}$. Recall that $\tilde{\mu}_n$ is the random measure with density $\hat{f}_n$. Thus,

$$D_n(p) \leq \lambda\left(\left\{t_n^{(p)} \leq \hat{f}_n < t^{(p)}\right\}\right) \leq \frac{1}{t_n^{(p)}}\tilde{\mu}_n\left(\left\{t_n^{(p)} \leq \hat{f}_n < t^{(p)}\right\}\right).$$

The factor $1/t_n^{(p)}$ in the right-hand side of the last inequality might be asymptotically bounded by some constant C, using Proposition 3.4. Hence, for all n large enough, and for almost all $p \in (0;1)$,

$$D_n(p) \leq C\left|\tilde{\mu}_n\left(\mathscr{L}_n(t_n^{(p)})\right) - \tilde{\mu}_n\left(\mathscr{L}_n(t^{(p)})\right)\right|. \tag{3.12}$$

The right-hand term in (3.12) may be bounded from above by

$$\begin{aligned}\left|\tilde{\mu}_n\left(\mathscr{L}_n(t_n^{(p)})\right) - \tilde{\mu}_n\left(\mathscr{L}_n(t^{(p)})\right)\right| \quad \leq \quad & \left|\tilde{\mu}_n\left(\mathscr{L}_n(t_n^{(p)})\right) - \mu\left(\mathscr{L}_n(t_n^{(p)})\right)\right| \\ & + \left|\mu\left(\mathscr{L}_n(t_n^{(p)})\right) - \mu\left(\mathscr{L}(t^{(p)})\right)\right| \\ & + \left|\mu\left(\mathscr{L}(t^{(p)})\right) - \tilde{\mu}_n\left(\mathscr{L}_n(t^{(p)})\right)\right|.\end{aligned}$$

By Lemma 3.6 and Lemma 3.7, we obtain, for almost all $p \in (0;1)$,

$$\sqrt{nh_n^d}\left\{\tilde{\mu}_n\left(\mathscr{L}_n(t_n^{(p)})\right) - \tilde{\mu}_n\left(\mathscr{L}_n(t^{(p)})\right)\right\} \xrightarrow{\mathbb{P}} 0,$$

which, according to (3.12), gives the stated result. □

A Auxiliary results on f and H

In this Appendix, we only assume that Assumption 1-*(i)* holds. Recall that $\mathscr{X}$ is the subset of $\mathbb{R}^d$ composed of the critical points of f, i.e.,

$$\mathscr{X} = \{\nabla f = 0\}.$$

The following lemma characterizes the set $\mathscr{T}_0$.

Lemma A.1. *We have* $f(\mathscr{X}) \setminus \{0; \sup f\} = \mathscr{T}_0$.

Proof. Let $x \in \mathscr{X}$. If $f(x) \neq 0$ or $f(x) \neq \sup f$, then obviously $f(x) \in \mathscr{T}_0$ and hence, $f(\mathscr{X}) \setminus \{0; \sup f\} \subset \mathscr{T}_0$. Conversely, $\mathscr{T}_0 \subset f(\mathscr{X})$ by continuity of ∇f and because the set $\{f = t\}$ is compact whenever $t \neq 0$. □

The next proposition describes the absolutely continuous part of the random variable $f(X)$.

Proposition A.2. *Let I be a compact interval of $\mathbb{R}_+^\star$ such that $I \cap \mathscr{T}_0 = \emptyset$. Then, the random variable $f(X)$ has a density g on I, which is given by*

$$g(t) = t \int_{\{f=t\}} \frac{1}{\|\nabla f\|} \mathrm{d}\mathscr{H}, \quad t \in I.$$

Proof. Since $\{f \in I\}$ is compact and $\{f \in I\} \cap \{\nabla f = 0\} = \emptyset$, we have

$$\inf_{\{f \in I\}} \|\nabla f\| > 0.$$

Now, let J be any interval included in I. Observe that f is a locally Lipschitz function and that $\mathbf{1}\{f \in J\}$ is integrable. According to Theorem 2, Chapter 3 in Evans and Gariepy (1992),

$$\mathbb{P}(f(X) \in J) = \int_{\{f \in J\}} f \mathrm{d}\lambda = \int_J \left(\int_{\{f=s\}} \frac{f}{\|\nabla f\|} \mathrm{d}\mathscr{H} \right) \mathrm{d}s,$$

hence the lemma. □

Acknowledgements

This work was partially supported by a grant from the French National Research Agency (ANR 09-BLAN-0051-01). We thank two anonymous referees and the Associate Editor for valuable comments and insightful suggestions.

References

Aubin, T. (2000). *A course in Differential Geometry*. Graduate Studies in Mathematics, American Mathematical Society, Providence, Rhode Island.

Baillo, A. (2003). Total error in a plug-in estimator of level sets. *Statist. and Probab. Lett.*, 65(4):411–417.

Baillo, A., Cuestas-Alberto, J., and Cuevas, A. (2001). Convergence rates in nonparametric estimation of level sets. *Statist. and Probab. Lett.*, 53(1):27–35.

Baillo, A., Cuevas, A., and Justel, A. (2000). Set estimation and nonparametric detection. *Canad. J. of Statist.*, 28(4):765–782.

Billingsley, P. (1995). *Probability and Measure*. Wiley-Intersciences.

Cadre, B. (2006). Kernel estimation of density level sets. *J. of Multivariate Anal.*, 97(4):999–1023.

Cuevas, A., Gonzalez-Manteiga, W., and Rodriguez-Casal, A. (2006). Plug-in estimation of general level sets. *Aust. N. Z. J. of Stat.*, 48(1):7–19.

Devroye, L., Gyorfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New-York.

Einmahl, U. and Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3):1380–1403.

Evans, L. and Gariepy, R. (1992). *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton, FL.

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Prob. Statist.*, 38(6):907–921.

Hartigan, J. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.*, 82(397):267–270.

Hyndman, R. (1996). Computing and graphing highest density regions. *The American Statistician*, 50:120–126.

Mason, D. and Polonik, W. (2009). Asymptotic normality of plug-in level set estimates. *Annals of Applied Probability*, 19:1108-1142.

Molchanov, I. (1998). A limit theorem for solutions of inequalities. *Scand. J. of Statist.*, 25(1):235–242.

Muller, D. and Sawitzki, G. (1991). Excess mass estimates and tests of multimodality. *J. Amer. Assoc.*, 86(415):738–746.

Nolan, D. (1991). The excess-mass elipsoid. *J. Multivariate Anal.*, 39(2):348–371.

Polonik, W. (1995). Measuring mass concentration and estimating density contour clusters—an excess mass approach. *Ann. Statist.*, 23(3):855–881.

Polonik, W. (1997). Minimum volume sets and generalized quantile processes. *Stochastic Process. Appl.*, 69(1):1–24.