Confidence Regions for Level Sets

Enno Mammen
Department of Economics
University of Mannheim
L 7, 3-5
68131 Mannheim
Germany

Wolfgang Polonik
Department of Statistics
University of California
One Shields Ave.
Davis, CA 95616-8705
USA

November 30, 2012

Abstract

This paper discusses a universal approach to the construction of confidence regions for level sets $\{h(x) \geq 0\} \subset \mathbb{R}^d$ of a function h of interest. The proposed construction is based on a plug-in estimate of the level sets using an appropriate estimate $\widehat{h}_n$ of h. The approach provides finite sample upper and lower confidence limits. This leads to generic conditions under which the constructed confidence regions achieve a prescribed coverage level asymptotically. The construction requires an estimate of quantiles of the distribution of $\sup_{\Delta_n} |\widehat{h}_n(x) - h(x)|$ for appropriate sets $\Delta_n \subset \mathbb{R}^d$. In contrast to related work from the literature, the existence of a weak limit for an appropriately normalized process $\{\widehat{h}_n(x), x \in D\}$ is not required. This adds significantly to the challenge of deriving asymptotic results for the corresponding coverage level. Our approach is exemplified in the case of a density level set utilizing a kernel density estimator and a bootstrap procedure.

This research was support by a grant of the DFG and the NSF

AMS 2000 subject classifications. Primary 62G07, Secondary 62G08, 62G09.

Keywords and phrases. Level sets, nonparametric curve estimation, kernel density estimation, smooth bootstrap.

1 Introduction

We consider the following problem. For a continuous function $h : \mathbb{R}^d \to \mathbb{R}$ let the zero-level set of h be defined as

$$C = \{x \in \mathbb{R}^d : h(x) \geq 0\}.$$

We assume throughout this paper that C is compact. Level sets are the central objects of this work. Relevant examples for h are (transformations of) regression functions, distribution functions, density functions, and others; see also below. In particular, the case $h(x) = f(x) - \lambda$, with f being a density function and λ a (strictly) positive constant leads to density level sets at level λ. Based on a sample $X_1, \ldots, X_n$ from a distribution F on $\mathbb{R}^d$, our interest is to derive a valid confidence region for the set C. To be a little more precise, let

$$C^- = \{x \in \mathbb{R}^d : h(x) > 0\}.$$

Then our interest is to find (random) sets $\widehat{C}_\ell = \widehat{C}_\ell(X_1, \ldots, X_n)$ and $\widehat{C}_u = \widehat{C}_u(X_1, \ldots, X_n)$ with

$$P[\widehat{C}_\ell \subset C^- \text{ and } C \subset \widehat{C}_u] \to 1 - \alpha \quad \text{as } n \to \infty, \tag{1.1}$$

where $1 - \alpha$, $\alpha \in (0, 1)$ is a given confidence level. What is needed for the inference is a 'good' estimator $\widehat{h}_n(\cdot) = \widehat{h}_n(\cdot, X_1, \ldots, X_n)$ for h.

In general one has to distinguish between two different scenarios. In one of the scenarios the appropriately standardized process $\{\widehat{h}_n(x),\, x \in D\}$ converges weakly to a tight limit process for an appropriate index set D. This is the situation considered in the literature on partially identified models, see below. See also Molchanov (1998), Vogel (2008), Jankowski and Stanberry (2011). In the second scenario, there exists no such limit process, as in the case of $\widehat{h}_n(x) = \widehat{f}_n(x) - \lambda$ with $\widehat{f}_n$ being a kernel density estimator. In this case, $\widehat{h}_n(x)$ does not have a tight limit distribution on the set $\{x : h(x) = 0\}$. The lack of a 'nice' limit makes this case more challenging. To our knowledge the so far only distributional result dealing with this situation is derived by Mason and Polonik (2009). There a central limit theorem for the symmetric difference of a level set and plug-in level set estimator based on a kernel estimator is derived.

In Section 2 we present a general treatment of this problem containing both of these scenarios as special cases. Non-asymptotic upper and lower bounds for the confidence level are derived. While

these bounds in general depend on unknown quantities, they indicate what is needed for confidence level consistency of the constructed confidence regions. This then is applied in Section 3 to the case of level sets of a density estimator where $\widehat{h}(x) = \widehat{f}_n(x) - \lambda$ with $\widehat{f}_n(x)$ being a kernel density estimator. This provides an instance of the second scenario discussed above. The challenge is that the limit distribution of $\sup_{h(x)=0} |\widehat{h}(x)|$ can only be handled under strong assumptions on the geometric structure of $\{x : h(x) = 0\}$. Such limit theorems are available for Gaussian processes, see e.g. Adler and Taylor (2007). To apply such results one could use strong Gaussian approximations for nonparametric curve estimators (e.g. for kernel estimators by using results of Rio, 1994). We will use another approach that avoids such technical geometrical assumptions. Our approach is based on a bootstrap method. Strong approximations are used to show the asymptotic consistency of the coverage level of the bootstrap method without studying limiting distributions.

Level set estimation and its applications have received significant attention in the recent literature. Statistical problems considered include the estimation of level sets including support estimation (e.g. Polonik, 1995, Cavalier, 1997, Tsybakov, 1997, Walther 1997, Cuevas and Fraiman, 1997, Molchanov, 1998, Baillo et al, 2001, Baillo, 2003, Gayraud and Rousseau, 2005, Cadre, 2006, Cuevas et al. 2006, Scott and Davenport, 2006, Scott and Nowak, 2006, Willett and Nowak, 2006, Biau et al, 2008, Mason and Polonik, 2009, Rigollet and Vert, 2009), optimal bandwidth selection (Samworth and Wand, 2010), clustering (e.g. Cuevas et al, 2000, Rinaldo et al. 2010), discrimination / classification (Mammen and Tsybakov, 1999, Hall and Kang, 2005, Steinwart et al. 2005, Audibert and Tsybakov, 2007), and visualization (Stuetzle, 2003, Klemelä, 2004, 2006, 2009, Stuetzle and Nugent, 2010). Level sets have applications in various fields such as engineering (e.g. anomaly detection, Desforges et al, 1998), flow cytometry (Duong et al, 2009), medical imaging (Willett and Nowak, 2005), astronomy (Jang, 2006). Other applications arise in econometrics in the context of partially identified models where the target set is a subset of the parameter space (see e.g. Chernozhukov et al., 2007; Bugni 2010, and references therein).

2 Construction of confidence regions and bounds for coverage probabilities

Let $\widehat{h}_n(x)$ denote an estimator for $h(x)$ based on a sample $X_1, \ldots, X_n$. The basic idea underlying our approach is to construct lower and upper confidence sets of the form

$$\widehat{C}_\ell = \widehat{C}^-(s) = \{x \in \mathbb{R}^d : \widehat{h}_n(x) > s\}, \qquad \widehat{C}_u = \widehat{C}(-t) = \{x \in \mathbb{R}^d : \widehat{h}_n(x) \geq -t\}$$

with $s, t \geq 0$. The question is, how to choose s and t in order to ensure good properties of the resulting confidence region? In this paper we will only consider the case $s = t$. First we introduce some notation. For $\beta \in \mathbb{R}$ let $C(\beta) = \{x \in \mathbb{R}^d : h(x) \geq \beta\}$ and $C^-(\beta) = \{x \in \mathbb{R}^d : h(x) > \beta\}$ (such that $C = C(0)$ and $C^- = C^-(0)$). For a given level $\alpha \in (0, 1)$ and a sequence $\beta_n \geq 0$ let

$$\Delta_n = \Delta(\beta_n) = C(-\beta_n) \setminus C^-(\beta_n) = \{x \in \mathbb{R}^d : -\beta_n \leq h(x) \leq \beta_n\},$$

and define

$$Z_n = Z_n(\beta_n) = \sup_{x \in \Delta_n} \big|\widehat{h}_n(x) - h(x)\big|, \tag{2.1}$$

where $\widehat{h}_n(x)$ denotes an estimator for $h(x)$. In the special case of $\beta_n = 0$ we denote

$$Z_n^0 = Z_n(0) = \sup_{\{x \in \partial C\}} \big|\widehat{h}_n(x) - h(x)\big|. \tag{2.2}$$

where we define $\partial C = \{x \in \mathbb{R}^d : h(x) = 0\}$. We assume without further mention that both Z_n and Z_n^0 are measurable. The distribution functions of $Z_n = Z_n(\beta_n)$ and Z_n^0 are denoted as

$$F_{Z_n}(t) = P\big(Z_n \leq t\big) \quad \text{and} \quad F_{Z_n^0}(t) = P(Z_n^0 \leq t).$$

For any $\beta_n \geq 0$ and $0 \leq \alpha \leq 1$ let

$$b_n^- = b_n^-(1-\alpha) = \inf\{t : F_{Z_n}(t) \geq 1 - \alpha\} \quad \text{and} \quad b_n^+ = b_n^+(1-\alpha) = \sup\{t : F_{Z_n}(t) \leq 1 - \alpha\}. \tag{2.3}$$

Given an estimator $\widehat{b}_n \geq 0$, define the upper and lower confidence bounds as

$$\widehat{C}_\ell := \{x \in \mathbb{R}^d : \widehat{h}_n(x) > \widehat{b}_n\}, \qquad \widehat{C}_u := \{x \in \mathbb{R}^d : \widehat{h}_n(x) \geq -\widehat{b}_n\}.$$

2.1 Non-asymptotic bounds for coverage probabilities

For $Z_n = Z_n(\beta_n)$ and $Z_n^0 = Z_n^0(\beta_n)$ as above, denote

$$p_n = p_n(\beta_n) = P(Z_n \le \widehat{b}_n) \quad \text{and} \quad p_{n0} = p_{n0}(\beta_n) = P(Z_n^0 \le \widehat{b}_n).$$

We have the following upper and lower bounds for the coverage probability:

Lemma 2.1 *Let $\beta_n \ge 0$. With $A_n = \{ \sup_{x\in\mathbb{R}^d} |\widehat{h}_n(x) - h(x)| \le \beta_n \}$ we have*

$$(1-\alpha) - P(A_n^{\complement}) + \big(p_n - F_{Z_n}(b_n^+)\big) \;\le\; P\big[\, \widehat{C}_\ell \subset C^- \text{ and } C \subset \widehat{C}_u \,\big] \;\le\; (1-\alpha) + \big(p_{n0} - F_{Z_n}(b_n^-)\big).$$

The proof of Lemma 2.1 is given in the Appendix.

This lemma implies that the sets $\widehat{C}_\ell$ and $\widehat{C}_u$ provide approximately valid lower and upper confidence sets if the following conditions (P) hold with sequences α_{n1}, α_{n2}, α_{n3} all being of order $o(1)$ as $n \to \infty$:

Conditions (P):

$$P(A_n^{\complement}) = O(\alpha_{n1}); \tag{P1}$$

$$r_n^{\pm} := |p_n - F_{Z_n}(b_n^{\pm})| = O(\alpha_{n2}); \tag{P2}$$

$$s_n := |p_{n0} - p_n| = O(\alpha_{n3}). \tag{P3}$$

Remarks. (i) Notice that all the above quantities depend on β_n. In particular this applies to both A_n and b_n. In fact, the smaller β_n, the smaller will be b_n and consequently, the smaller our confidence region. Also the quantity s_n is decreasing with decreasing β_n. On the other hand, in order to get $P(A_n^{\complement})$ small, we want β_n to be large. A good choice of β_n will thus strike a balance between these conflicting goals.

(ii) By definition $r_n^{\pm} = |P(Z_n \le b_n^{\pm}) - P(Z_n \le \widehat{b}_n)|$. Thus, if $b_n^{\pm}$ and $\widehat{b}_n$ are close, and we have control over small increments of the cdf of Z_n, then we can control $r_n^{\pm}$. This is being made precise in Theorem 2.2 below.

(iii) Not surprisingly, conditions similar to (P1) - (P3) also show up implicitly in the proofs of Chernozhukov et al. (2007) and Bugni (2010) in the context of partially identified models, see e.g. Condition C.5 on page 1257 of Chernozhukov et al. (2007) and formula (A4) in the supplemental material of Bugni (2010). However, in the situation considered by Bugni and Chernozhukov et

al. the process $a_n(\widehat{h}_n(x) - h(x))$ converges weakly to a tight limiting process for some sequence $a_n \to \infty$. The case without an existing tight limit is considered in Chernozhukov et al. (2012), where the distribution of $\sup_{x\in C} \widehat{h}_n(x)$ is analyzed via strong Gaussian approximations. In contrast to our case the set C is assumed to be known there.

2.1.1 A generic approach for controlling $\mathbf{r}_n^{\pm}$ (cf. (P2))

Recall the definitions of $b_n^{\pm}(\beta)$ given in (2.3). Let $\widehat{b}_n \geq 0$ denote an estimator.

Lemma 2.2 *Let γ_{n1} and τ_{n1} denote real numbers such that*

$$\sup_{t\in\mathbb{R}} P(Z_n \in [t,\, t+\gamma_{n1}]) \leq \tau_{n1}. \tag{2.4}$$

Assume further that the estimator $\widehat{b}_n$ satisfies the following property. There exist sequences $\{\eta_n\}$ and $\{\delta_{n1}\}$ of positive real numbers with

$$P\Big(b_n^-(1-\alpha-\eta_n) - \gamma_{n1} \,\leq\, \widehat{b}_n \,\leq\, b_n^+(1-\alpha+\eta_n) + \gamma_{n1}\Big) \geq 1 - \delta_{n1}. \tag{2.5}$$

Then we have $r_n^{\pm} \leq 2\,\tau_{n1} + 2\eta_n + \delta_{n1}$.

The proof of the lemma is deferred to the Appendix.

Application to a bootstrap estimate:

We will apply Lemma 2.2 to bootstrap estimates of the quantile, i.e. $\widehat{b}_n = b_n^* = b_n^*(X_1, \ldots, X_n)$ is constructed by using bootstrap processes Z_n^*. Since we have to account for possible flat parts of the distribution function of Z_n^*, we define $\widehat{b}_n$ as a random variable satisfying

$$b_n^{*^-} \leq \widehat{b}_n \leq b_n^{*^+}, \tag{2.6}$$

where $b_n^{*^-}$ and $b_n^{*^+}$ are defined as

$$\begin{aligned} b_n^{*^-} &= b_n^{*^-}(1-\alpha) = \inf\{t : P^*(Z_n^* \leq t) \geq 1-\alpha\} \\ b_n^{*^+} &= b_n^{*^+}(1-\alpha) = \sup\{t : P^*(Z_n^* \leq t) \leq 1-\alpha\}, \end{aligned} \tag{2.7}$$

and P^* denotes the conditional probability, given the sample. Further, $\mathcal{L}(V)$ denotes the unconditional distribution of a random variable V and $\mathcal{L}^*(V)$ denotes the conditional distribution of V, given the sample.

The following theorem shows how a strong approximation enters the crucial condition (2.5) for the estimator $\widehat{b}_n$ just defined.

Lemma 2.3 *Let Z_n^* be such that there exist real numbers γ_{n1}, β_n and δ_n' with*

$$E\left[P^*\big(\,\big|Z_n^+ - Z_n^*\big| \le \gamma_{n1}\big)\right] = P\big(\,\big|Z_n^+ - Z_n^*\big| \le \gamma_{n1}\big) \ge 1 - \delta_n' \tag{2.8}$$

for appropriately constructed random variables Z_n^+ with

$$\mathcal{L}^*(Z_n^+) = \mathcal{L}(Z_n).$$

Assume further that there exists τ_{n1} such that (2.4) holds with γ_{n1} as in (2.8). Then (2.5) holds with γ_{n1}, $\eta_n = 2\tau_{n1} + 2\sqrt{\delta_n'}$ and $\delta_{n1} = \sqrt{\delta_n'}$ and thus

$$r_n^{\pm} = |P(Z_n \le \widehat{b}_n) - P(Z_n \le b_n^{\pm})| \le 6\tau_{n1} + 5\sqrt{\delta_n'}.$$

2.1.2 A generic approach for controlling $\mathbf{s}_n$ (cf. (P3))

Lemma 2.4 *Let γ_{n2}, τ_{n2} and δ_{n2} denote real numbers such that*

$$\sup_{t\in\mathbb{R}} P(Z_n \in [t,\, t+\gamma_{n2}]) \le \tau_{n2}. \tag{2.9}$$

and

$$P\Big(|Z_n - Z_n^0| \le \gamma_{n2}\Big) \ge 1 - \delta_{n2}. \tag{2.10}$$

Suppose further that (2.4) and (2.5) hold with η_n, δ_{n1} and τ_{n1}, respectively. Then we have

$$s_n \le 2\,\tau_{n1} + 2\tau_{n2} + 2\eta_n + \delta_{n1} + \delta_{n2}.$$

3 Confidence sets for level sets of a probability density via a kernel density estimator and the bootstrap

To exemplify the universal approach for constructing confidence regions for level sets developed above, we now consider the special case of density level sets. As an estimator for the density we use

a kernel density estimator. Our main result will provide rates of convergence for the corresponding coverage probability.

Let $h(x) = f(x) - \lambda$ with f a continuous pdf, so that in this section

$$C = \{x \in \mathbb{R}^d : f(x) - \lambda \geq 0\}$$

is the level set of f at level λ. Assume that we have available a random sample $X_1, \ldots, X_n$ from f. Define the kernel density estimator as

$$\widehat{f}_{n,h}(x) = \frac{1}{n\, h^d} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h} \Big). \tag{3.1}$$

Here the kernel K is a symmetric probability density function, and h is the bandwidth. (There should be no confusion with the notation $h(x)$ used in the previous sections for the underlying function of interest.) Consequently, in this section $\widehat{h}_n(x) = \widehat{f}_{n,h}(x) - \lambda$ and thus

$$\widehat{C}(b) = \{x \in \mathbb{R}^d : \widehat{f}_{n,h}(x) - \lambda \geq b\},$$

and we will construct an estimate b_n^* with

$$P\big[\, \widehat{C}^-(b_n^*) \subset C^- \text{ and } C \subset \widehat{C}(-b_n^*) \,\big] \to 1 - \alpha.$$

To this end we will utilize the smooth bootstrap procedure that goes back to Efron (1979). Here we draw bootstrap samples $X_1^{*,g}, \ldots, X_n^{*,g}$ from $\widehat{f}_{n,g}$ where the bandwidth g can be different from h. For simplicity we use the same kernel K for the bootstrap procedure as for the original density estimator. Denote by $\widehat{f}^*_{n,h}(x)$ the kernel density estimator based on the bootstrap sample, defined as in (3.1) but with the sample $\{X_1, \ldots, X_n\}$ replaced by the bootstrap sample $\{X_1^{*,g}, \ldots, X_n^{*,g}\}$. (The dependence of $\widehat{f}^*_{n,h}(x)$ on the bandwidth g is dropped in this notation.) Let further $\widehat{\Delta}_n = \widehat{C}(-\beta_n) \setminus \widehat{C}^-(\beta_n)$ and with

$$Z_n^* = \sup_{x \in \widehat{\Delta}_n} \big| \widehat{f}^*_{n,h}(x) - \widehat{f}_{n,g}(x) \big|, \tag{3.2}$$

define

$$\begin{aligned} b_n^{*^-} &= \inf\{t : P(Z_n^* \leq t |\, X_1, \ldots, X_n) \geq 1 - \alpha\} = \inf\{t : P^*(Z_n^* \leq t) \geq 1 - \alpha\} \\ b_n^{*^+} &= \sup\{t : P(Z_n^* \leq t |\, X_1, \ldots, X_n) \leq 1 - \alpha\} = \sup\{t : P^*(Z_n^* \leq t) \leq 1 - \alpha\} \end{aligned}$$

and $\widehat{b}_n$ is a random variable with

$$b_n^{*^-} \;\le\; \widehat{b}_n \;\le\; b_n^{*^+}. \tag{3.3}$$

We now derive explicit rates for the upper and lower bounds of the coverage probability for the corresponding confidence regions for the level sets.

(A) Assumptions on f and the kernel K:

(*A.i*) f is bounded, Lipschitz continuous, and twice differentiable.

(*A.ii*) There exists $\epsilon_0, \delta_0 > 0$ with $\|\mathrm{grad} f(x)\| > \epsilon_0 > 0$ for all $x \in \Delta(\delta_0)$.

(*A.iii*) K is a pdf, symmetric about zero, twice continuously differentiable with support contained in $[-1,1]^d$.

Theorem 3.1 *Suppose that assumption (A) holds. Put $\alpha_n = g^2 + (ng^d)^{-1/2}(\log n)^{1/2}$ and let ϵ_n be such that*

$$\sup_{x\in\mathbb{R}^d} \Big| \frac{\partial^2}{\partial x_i \partial x_j} \widehat{f}_{n,g}(x) - \frac{\partial^2}{\partial x_i \partial x_j} f(x) \Big| = O_P(\epsilon_n) \quad \text{for all } \; i,j = 1,\dots,d.$$

Suppose that $\sqrt{\frac{\log n}{nh^d}} + h^2 = o(\beta_n)$ and

$$\rho_n = \log n \Big[\sqrt{\frac{\log n}{nh^{d+2}}} + \frac{\beta_n}{h} + h + \sqrt{\alpha_n} + \sqrt{h^d \log n} + \sqrt{\frac{n\, h^{d+4}(\epsilon_n^2 + \beta_n^2)}{\log n}} + \sqrt[4]{\frac{\log n}{nh^d}} \Big] = o(1)$$

as $n \to \infty$. Then, for $L > 0$ arbitrary,

$$\Big| P\big[\, \widehat{C}^-(\widehat{b}_n) \subset C^- \text{and } \; C \subset \widehat{C}(-\widehat{b}_n) \,\big] - (1-\alpha) \Big| = O\big(\rho_n + n^{-L}\big).$$

Under typical assumptions we have $\epsilon_n = \sqrt{\frac{\log n}{ng^{d+4}}} + g^\mu$, where g^μ is the order of the bias of $\frac{\partial^2}{\partial x_i \partial x_j}\widehat{f}_{n,g}(x)$ and depends on the smoothness of $\frac{\partial^2}{\partial x_i \partial x_j} f_{n,g}(x)$. For Lipschitz continuous second derivatives we have $\mu = 1$ and the optimal choice of $h \sim n^{-1/(d+4)}$ and $g \sim n^{-1/(d+6)}$ we obtain the rate $\rho_n = g \log n$.

Proof. We will show that (P1) - (P3) hold (see discussion given right after Lemma 2.1) with certain rates α_{n1}, α_{n2}, and α_{n3} the sum of which is of the order given in the theorem. Lemma 2.1 implies the assertion. To simplify the notation further we write $\widehat{f}_n(x)$ instead of $\widehat{f}_{n,h}(x)$.

Verification of condition (P1). It is well-known that under our assumptions we have

$$P\{\sup_{x\in\mathbb{R}^d}|\widehat{f}_n(x)-f(x)|\geq\beta_n\}=O(n^{-L}) \tag{3.4}$$

with L arbitrarily large. In other words, (P1) holds with rate $\alpha_{n1}=n^{-L}$.

Verification of conditions (P2) and (P3).

We will utilize Theorems 2.3 and 2.4. To this end we have to verify the conditions of these results. This is done in the following.

Verification of conditions (2.4) and (2.9). Here we argue that for $-\infty<c<d<\infty$ we have

$$\begin{aligned}&P\big(\sup_{x\in\Delta(\beta_n)}|\widehat{f}_n(x)-f(x)|\in[c,d]\,\big)\\&\qquad=O\Big((d-c)\sqrt{nh^d\,\log n}+h\,\log n+\log n\,\Big(\sqrt[4]{\frac{\log n}{nh^d}}\,+\,\sqrt{h^d\log n}\,\Big)\Big),\end{aligned} \tag{3.5}$$

with the constants involved in the O-term not depending on c,d, and that the same holds for β_n replaced by 0. An application of this result to $[c,d]=[t,t+\gamma_{ni}]$, $i=1,2$ where the explicit form of γ_{ni} is given below, shows both (2.4) and (2.9) with τ_{ni} given through plugging in $d-c=\gamma_{ni}$ into the right hand side of (3.5), i.e.

$$\tau_{ni}=\gamma_{ni}\sqrt{n\,h^d\,\log n}+\log n\,\Big(h+\sqrt[4]{\frac{\log n}{nh^d}}+\sqrt{h^d\,\log n}\Big),\quad i=1,2. \tag{3.6}$$

To show (3.5) we utilize the proof of Proposition 3.1 from Neumann (1998). Neumann's Proposition 3.1 is similar to (2.4) and (2.9), respectively, but with the supremum being extended over the entire $\mathbb{R}^d$. In fact, Neumann's result says (under assumptions that in the present iid setting are implied by our assumptions) that $P(\sup_{x\in\mathbb{R}^d}|\widehat{f}_n(x)-\mathrm{E}\widehat{f}_n(x)|\in[c,d]\,)=O\big((d-c)\sqrt{nh^d\,\log n}+\log n\,\big(h+\sqrt[4]{\frac{\log n}{nh^d}}+\sqrt{h^d\log n}\big)\,\big)$. (Even though it is not needed here, it is worth pointing out that Neumann's result even holds under certain dependence assumptions on the X_i.) To verify (3.5) we need to show that in Neumann's result $\mathbb{R}^d$ can be replaced by $\Delta(\beta_n)$ and $\mathrm{E}\widehat{f}_n(x)$ replaced by $f(x)$. A close inspection of Neumann's proof in fact reveals that such a result continues to hold for $\mathbb{R}^d$ replaced by either the set $\Delta(\beta_n)$ or $\Delta(0)=\partial C$, i.e. we have (3.5), and the same holds for β_n replaced by 0. While details are omitted, we briefly outline the changes to Neumann's proof that are in order. On

page 2043, Neumann argues that the supremum in his result is attained in areas where $\mathrm{E}\widehat{f}_n(x)$ is bounded away from zero. In our case, i.e. for the supremum restricted to $\Delta(\beta_n)$ we have $f(x) > \lambda/2$ on $\Delta(\epsilon)$ for $0 \le \epsilon < \epsilon_0$ and ϵ_0 sufficiently small. Thus, we automatically have for n large enough that $\int h^{-d} K(\frac{x-z}{h})\, f(z)\, dz > \lambda/2$ for all $x \in \Delta(\beta_n)$. To take care of the centering around $f(x)$ instead of $\mathrm{E}\widehat{f}_n(x)$ we have to replace $T_{k2}(x)$ in Neumann's proof by $T_{k2}(x) + \mathrm{E}\widehat{f}_n(x) - f(x)$. Having observed that, the only further changes to Neumann's proof are now to restrict the supremum in the definition of the quantities Z_ℓ and $T_{k,2}$ from Neumann's proof to $x \in I_k \cap \Delta(\beta_n)$. With this change, the arguments in the Neumann's proof can be followed to derive (3.5). Further details are omitted. As indicated above, (2.4) and (2.9) follow from (3.5) with the corresponding $\tau_{ni}, i = 1, 2$ given in (3.6).

Verification of condition (2.10). For a set A denote

$$V_n(A) = \sup_{x \in A} |\widehat{f}_n(x) - f(x)|,$$

and for $\delta > 0$ let A_δ denote the δ-enlarged set A, i.e. $A_\delta = \bigcup_{x \in A} U_\delta(x)$ where $U_\delta(x)$ denotes the closed ball of radius δ with midpoint x. Recall that $\Delta(\beta_n) = C(-\beta_n) \setminus C^-(\beta_n)$, and notice that $V_n(\Delta(t)) = Z_n(t)$. In other words, we have $V_n(\Delta(\beta_n)) = Z_n$ and $V_n(\Delta(0)) = Z_n^0$.
Now observe that by using assumption $(A.ii)$ there exists a constant $c > 0$ such that for large enough n we have $\Delta(\beta_n) \subset \big(\partial C\big)_{c\beta_n}$. In other words, each $x \in \Delta(\beta_n)$ can be written as $x = y + b$ with $y \in \partial C = \Delta(0)$ and $b \in B_c = \{x \in \mathbb{R}^d : \|x\| \le c\beta_n\}$. Therefore we can write

$$\sup_{x \in \Delta(\beta_n)} |\widehat{f}_n(x) - f(x)| \le \sup_{y \in \partial C} |\widehat{f}_n(y) - f(y)| + \sup_{\|x-y\| \le c\,\beta_n} |(\widehat{f}_n(x) - f(x)) - (\widehat{f}_n(y) - f(y))|,$$

or,

$$0 \le V_n(\Delta(\beta_n)) - V_n(\Delta(0)) \le \sup_{\|x-y\| \le c\,\beta_n} |(\widehat{f}_n(x) - f(x)) - (\widehat{f}_n(y) - f(y))| =: R_n(c\,\beta_n).$$

We will show that $R_n(c\beta_n)$ satisfies

$$P\big(R_n(c\,\beta_n) \ge \gamma_{n2}\big) = O(\delta_{n2}) \tag{3.7}$$

with $\gamma_{n2} = C\big(\frac{\sqrt{\log n}}{\sqrt{n\, h^{d+2}}} + h^2\big)\, \beta_n$, $C > 0$ an appropriate constant and $\delta_{n2} = n^{-L}$ for $L > 0$ arbitrary. Assume for the moment that this is true. Since $V_n(\Delta(\beta_n)) - V_n(\Delta(0)) = Z_n - Z_n^0$ we obtain from (3.7) that (2.10) holds with γ_{n2} and δ_{n2} as just specified.

To show (3.7) one might use empirical process theory as follows. First observe that

$$\nu_n(x) = \sqrt{n}\, h^d(\widehat{f}_n(x) - \mathrm{E}\widehat{f}_n(x)) = \sqrt{n}\,\Big(\frac{1}{n}\sum_{i=1}^n K\Big(\frac{X_i - x}{h}\Big) - \mathrm{E}K\Big(\frac{X_i - x}{h}\Big)\Big),\ \ x \in \mathbb{R}^d$$

can be viewed as an empirical process indexed by $\mathcal{S}_n = \big\{K\big(\frac{\cdot - x}{h}\big),\, x \in \mathbb{R}^d\big\}$. Under our assumptions on the bandwidth, the class $\mathcal{S}_n$ is known to be a VC-class with bound on the uniform covering number not depending on n (but only on K); for details see Rio (1994), for instance. The same applies to the process of differences $\nu_n(x) - \nu_n(y)$ with $\|x - y\| \le c\beta_n$ (regarded as an empirical process indexed by the class of differences $\mathcal{D}_n = \big\{K\big(\frac{\cdot - x}{h}\big) - K\big(\frac{\cdot - y}{h}\big),\, \|x - y\| \le c\beta_n,\, x, y \in \mathbb{R}^d\big\}$). Further, using our assumptions on K, we have for n sufficiently large (such that $\beta_n \le h$)

$$\begin{aligned}
\mathrm{Var}\Big(K\Big(\frac{X - x}{h}\Big) - K\Big(\frac{X - y}{h}\Big)\Big) &\le \int_{\{\max_i |x_i - u_i| \le h\}} \Big(K\big(\frac{u - x}{h}\big) - K\big(\frac{u - y}{h}\big)\Big)^2 f(u)\, du \\
&= h^d \int_{\max |v_i| \le 1} \Big(K(v) - K\Big(v + \frac{x - y}{h}\Big)\Big)^2 f(x + vh)\, dv \\
&\le C_0^2 \Big(\frac{\|x - y\|}{h} \wedge 1\Big)^2 h^d \le C_0^2\, h^d \Big(\frac{\beta_n}{h}\Big)^2
\end{aligned}$$

for some constant $C_0^2 > 0$ depending on c, d and on our assumptions on f and K. As for the bias notice that the assumed regularity assumptions on our kernel and on f assure that

$$\mathrm{E}\big[\big(\widehat{f}_n(x) - f(x)\big) - \big(\widehat{f}_n(y) - f(y)\big)\big] = O\Big(h^2\, \|x - y\|\Big) \le C_1\, h^2\, \beta_n. \tag{3.8}$$

Let $\sigma = C\,\beta_n\, h^{\frac{d}{2} - 1}$ for some $C \ge C_0$ (chosen sufficiently large). Our assumptions now imply that for n large enough we can apply Theorem 2.8 of Alexander (1984) (with $\alpha^{1/2} = \sigma$ and $M = \sigma\,\sqrt{\widetilde{c}\,\log n}\,$, $\widetilde{c} > 0$, in Alexander's notation), and we obtain

$$\begin{aligned}
& P\Big[\sup_{|x-y| \le c\beta_n} \big|\widehat{f}_n(x) - \widehat{f}_n(y) - \big(f(x) - f(y)\big)\big| \ge \frac{M}{h^d\,\sqrt{n}} + C_1 h^2 \beta_n\Big] \\
\le\ & P\Big[\sup_{|x-y| \le c\beta_n} \big|\widehat{f}_n(x) - \widehat{f}_n(y) - \mathrm{E}\big(\widehat{f}_n(x) - \widehat{f}_n(y)\big)\big| \ge \frac{M}{h^d\,\sqrt{n}}\Big] \\
=\ & P\Big[\sup_{|x-y| \le c\beta_n} \big|\nu_n(x) - \nu_n(y))\big| \ge M\Big] \\
\le\ & 16\, \exp\Big\{-\frac{1}{4}\,\frac{M^2}{\sigma^2 + \frac{M}{3\sqrt{n}}}\Big\} \\
=\ & 16\, \exp\Big\{-\frac{1}{4}\,\frac{\widetilde{c}\,\sigma^2 \log n}{\sigma^2 + \frac{\sigma\sqrt{\widetilde{c}\,\log n}}{3\sqrt{n}}}\Big\}.
\end{aligned}$$

By assumption $\sqrt{\frac{n}{\log n}}\,\sigma \to \infty$. Thus, for n large enough we have $\frac{\sigma\sqrt{\widetilde{c}\log n}}{3\sqrt{n}} \le \sigma^2$ so that our bound can be further estimated by

$$\le\ 16\,\exp\big\{-\frac{1}{8}\,\widetilde{c}\,\log n\big\} \le \widetilde{C}\,n^{-L}, \tag{3.9}$$

where the last inequality holds for $L>0$ arbitrarily large by choosing $\widetilde{c}>0$ large enough. It follows that for some $\tilde{C}>0$ large enough and $L>0$ arbitrarily large,

$$P\Big(\sup_{|x-y|\le c\,\beta_n} |\big(\widehat{f}_n(x)-\widehat{f}_n(y)\big)-\big(f(x)-f(y)\big)| \ge \tilde{C}\Big(\frac{\sqrt{\log n}}{\sqrt{n\,h^{d+2}}}+h^2\Big)\,\beta_n\Big) = O\big(n^{-L}\big). \tag{3.10}$$

This is (3.7), which in turn implies (2.10).

Verification of condition (2.8). Recall the definition of $V_n(A)$ given above at the beginning of the derivation of the rate of (P2). Similarly define for $A\subset\mathbb{R}^d$,

$$V_n^*(A) = \sup_{x\in A} |\widehat{f}_n^*(x) - f_n^*(x)|,$$

and let $\widehat{\Delta}(\beta_n) = \widehat{C}(-\beta_n)\setminus\widehat{C}^-(\beta_n) = \{x\in\mathbb{R}^d : -\beta_n \le \widehat{f}_n(x)-\lambda \le \beta_n\}$. We now argue that (2.8) holds for $Z_n^* = V_n^*(\widehat{\Delta}(\beta_n))$.

Again we start with a result by Neumann (1998). This paper constructs a pairing of the random variables $\{X_1,\dots,X_n\}$ and $\{X_1^{*,g},\dots,X_n^{*,g}\}$ such that the following results hold on a rich enough probability space. In the current density estimation context, Theorem 3.1 in Neumann (1998) provides a result, closely related to condition (2.8) when ignoring the bias. The difference is that Neumann considers suprema over entire $\mathbb{R}^d$, i.e. he considers the quantity $M_n = \sup_{x\in\mathbb{R}^d}|\widehat{f}_n(x)-\mathrm{E}\widehat{f}_n(x)|$ and $M_n^* = \sup_{x\in\mathbb{R}^d}|\widehat{f}_n^*(x)-\mathrm{E}\widehat{f}_n^*(x)|$. Under assumptions that are implied by our assumptions Neumann shows that on an appropriate probability space there exists a pairing of the corresponding random variables such that for an arbitrarily large $L>0$

$$P(|M_n^*-M_n| > c\,\widetilde{\gamma}_n) = O(n^{-L}) \qquad \text{as}\ \ n\to\infty,$$

with

$$\widetilde{\gamma}_n = \log n\,[n^{-1/2}+(nh^d)^{-1}] + \alpha_n^{1/2}(nh^d)^{-1/2}(\log n)^{1/2}\,), \tag{3.11}$$

where $\alpha_n = g^2+(ng^d)^{-1/2}(\log n)^{1/2}$, and g denotes the bandwidth of the bootstrap density, namely the kernel density estimator with bandwidth g. In this proof we denote this bootstrap density by

$f_n^*(x)$. In fact, inspecting the proof of Neumann's result shows, that it is actually shown that for $C > 0$ large enough we have

$$P\big(\sup_{x\in\mathbb{R}^d} \big| \big(\widehat{f}_n^*(x) - \mathrm{E}\widehat{f}_n^*(x)\big) - \big(\widehat{f}_n(x) - \mathrm{E}\widehat{f}_n(x)\big) \big| \geq C\,\widetilde{\gamma}_n \big) = O(n^{-L}). \tag{3.12}$$

It is straightforward to see that our assumptions imply that

$$\begin{aligned}\sup_{x\in R^d} \big|(E\widehat{f}_n^*(x) - E\widehat{f}_n(x)) - (f_n^*(x) - f(x))\big| &= O_P\Big(h^2 \sum_{i,j=1}^d \sup_{x\in\mathbb{R}^d} \big|\frac{\partial^2}{\partial x_i \partial x_j}\widehat{f}_{n,g}(x) - \frac{\partial^2}{\partial x_i \partial x_j} f(x)\big|\Big) \\ &= O_P(h^2\epsilon_n),\end{aligned}$$

which together with (3.12) gives

$$P\big(\sup_{x\in\mathbb{R}^d} \big| \big(\widehat{f}_n^*(x) - f_n^*(x)\big) - \big(\widehat{f}_n(x) - f(x)\big) \big| \geq C\,\big(\widetilde{\gamma}_n + h^2\epsilon_n\big) \big) = O(n^{-L}). \tag{3.13}$$

Obviously, (3.13) implies that for any sequence A_n of (measurable) subsets, whether random or not, we have

$$P\big(\sup_{x\in A_n} \big| \big(\widehat{f}_n^*(x) - f_n^*(x)\big) - \big(\widehat{f}_n(x) - f(x)\big) \big| \geq C\,\big(\widetilde{\gamma}_n + h^2\epsilon_n\big) \big) = O(n^{-L}).$$

Thus we have

$$\sup_{0\leq t\leq\beta_n} P\big(\big|V_n^*(\widehat{\Delta}(t)) - V_n(\widehat{\Delta}(t))\big| > C\,\big(\widetilde{\gamma}_n + h^2\epsilon_n\big) \big) = O\big(n^{-L}\big). \tag{3.14}$$

We now show that for $C > 0$ large enough

$$\sup_{0\leq t\leq\beta_n} P\big(\big|V_n(\widehat{\Delta}(t)\big) - V_n(\Delta(t))\big| > C\,\gamma_{n2}\big) = O(n^{-L}) \tag{3.15}$$

with γ_{n2} as in (3.7) and $L > 0$ arbitrary. The two properties (3.14) and (3.15) imply the desired result, i.e. for $C > 0$ sufficiently large we have

$$\sup_{0\leq t\leq\beta_n} P\big(\big|V_n^*(\widehat{\Delta}(t)\big) - V_n\big(\Delta(t)\big)\big| > C\,(\widetilde{\gamma}_n + \gamma_{n2} + h^2\epsilon_n) \big) = O\big(n^{-L}\big). \tag{3.16}$$

This then implies that (2.8) holds with $\gamma_{n1} = \widetilde{\gamma}_n + \gamma_{n2} + h^2\epsilon_n$ and $\delta_n' = n^{-L}$.

In order to derive (3.15) recall that by our assumptions for $C > 0$ large enough, $P\big[\sup_{x\in\mathbb{R}^d} |\widehat{f}_n(x) - f(x)| \geq C\,\beta_n\big] = O(n^{-L})$. This implies that

$$P\Big[\ \widehat{\Delta}(t) \subset \Delta\big(t + C\,\beta_n\big)\ \text{ for all }\ 0 \leq t \leq \beta_n\ \Big] > 1 - O(n^{-L}).$$

Utilizing assumption $(A.ii)$ we can find a constant $c_1 > 0$ such that for n large enough we have $\Delta\big(t + C\,\beta_n\big) \subset \big(\,\Delta(t)\,\big)_{c_1\,\beta_n}$. Consequently, we have that on a set with probability $1 - O(n^{-L})$ we can write $x \in \widehat{\Delta}(t)$ as $x = y + b$ with $y \in \Delta(t)$ and $\|b\| \le c_1\beta_n$, and thus

$$\big|\,V_n\big(\widehat{\Delta}(t)\big) - V_n\big(\Delta(t)\big)\,\big| \le \sup_{\|x-y\|\le c_1\beta_n} \big|\,\big(\,\widehat{f}_n(x) - f(x)\big) - \big(\,\widehat{f}_n(y) - f(y)\,\big)\big|,$$

and (3.10) implies (3.15). Thus (2.8) is verified with $\gamma_n = \widetilde{\gamma}_n + \gamma_{n1} + h^2\epsilon_n$, where $\widetilde{\gamma}_n$ is given in (3.11), and $\delta'_n = n^{-L}$ with L arbitrarily large.

After we have verified their conditions we are now in a position to be able to apply Lemmas 2.2 - 2.4 to show (P2) and (P3) and to determine the corresponding rates α_{n2} and α_{n3}, respectively.

Applications of Lemmas 2.2 - 2.4 to verify (P2) and (P3).

First we use Lemma 2.3 to derive the rate α_{n2} from (P2). The above verification of (2.8) gives us a rate $\gamma_{n1} = \widetilde{\gamma}_n + \gamma_{n2} + h^2\,\epsilon_n$ and $\delta'_n = O(n^{-L})$ for $L > 0$ arbitrary. Using (3.6) gives an expression for τ_{n1} (see below). Lemma 2.3 implies that $r_n^{\pm} = O(\tau_{n1} + \sqrt{\delta'_n}) = O(\tau_{n1} + n^{-L})$, $L > 0$ arbitrary. In other words we have $\alpha_{n2} = \tau_{n1} + n^{-L}$ with

$$\begin{aligned}
\tau_{n1} &= (\widetilde{\gamma}_n + \gamma_{n2} + h^2\epsilon_n)\,\sqrt{nh^d\,\log n} + \sqrt{nh^d\,\log n} + h\,\log n + \log n\,\Big(\sqrt[4]{\frac{\log n}{nh^d}} + \sqrt{h^d \log n}\Big) \\
&= \Big[\,\log n\,[n^{-1/2} + (nh^d)^{-1}] + \alpha_n^{1/2}(nh^d)^{-1/2}(\log n)^{1/2}\,)\Big]\,\sqrt{nh^d\,\log n} + \frac{\beta_n\,\log n}{h} \\
&\qquad + \sqrt{nh^{d+4}\,(\epsilon_n^2 + \beta_n^2)\,\log n} + h\,\log n + \log n\,\Big(\sqrt[4]{\frac{\log n}{nh^d}} + \sqrt{h^d \log n}\Big) \\
&= O\Big(\,\log n\,\Big[\sqrt{\frac{\log n}{nh^{d+2}}} + \frac{\beta_n}{h} + h + \sqrt{\alpha_n} + \sqrt{h^d\,\log n} + \sqrt{\frac{n\,h^{d+4}(\epsilon_n^2 + \beta_n^2)}{\log n}} + \sqrt[4]{\frac{\log n}{nh^d}}\,\Big]\Big).
\end{aligned} \tag{3.17}$$

In order to derive the rate for $s_n^{\pm}$ we apply Lemma 2.4. To this end observe that we already have derived explicit expressions for γ_{n1} and τ_{n1}. Lemma 2.8 implies that (2.5) holds with γ_{n1} and $\eta_n = O(\tau_{n1} + \sqrt{\delta'})$. The derivation of α_{n2} shows that $\eta_n = O(\alpha_{n2})$. Further, we have seen that (2.10) holds with $\gamma_{n2} = O\big(\big(\sqrt{\frac{\log n}{nh^{d+2}}} + h^2\big)\,\beta_n\big)$ and $\delta_{n2} = O(n^{-L})$, $L > 0$ arbitrary. Since γ_{n1} is not of smaller order than γ_{n2}, and (3.7) implies the same relation between τ_{n1} and τ_{n2}, we obtain by using Lemma 2.4 that

$$s_n^{\pm} = O(\tau_{n1} + \tau_{n2} + \eta_n + \delta_{n1} + \delta_{n2}) = O(\tau_{n2} + \alpha_{n2} + n^{-L}) = O(\alpha_{n2}).$$

4 Further examples and outlook.

Various interesting examples fall into the general framework discussed in section **??**. For instance, one can consider level sets of the form $\{x \in \mathbb{R}^d : h(x) \geq \lambda(x)\}$ where $\lambda(x)$ is a known function. Another class of examples comes from discrimination and binary classification. E.g. the estimation of the binary classifier of the form $\{x \in \mathbb{R}^d : p\, f(x) - (1-p)g(x) \geq 0\}$ where f, g denote the two class densities and p is the probability of observing from f. (See Duong et al. 2009). When using kernel density estimators, both of these instances can be treated by using similar techniques as in section 3. Another research problem is the construction of confidence regions for intersections of level sets of several different functions h. Such types of problems arise naturally in econometrics (e.g. see Bugni 2010).

Promising future research problems on level set estimation related to our work include in particular applications to dependent data, where the level sets also might vary with time. We have also not addressed optimality issues. For instance, for constructing even smaller confidence regions it might be interesting to construct confidence regions via asymmetric thresholds, i.e. to consider upper and lower 'confidence bounds' of the form $\{x : \widehat{f}_n(x) \geq -\widehat{b}_{n1}\}$ and $\{x : \widehat{f}_n(x) \geq \widehat{b}_{n2}\}$, respectively, with $0 \leq \widehat{b}_{n1}, \widehat{b}_{n2}$ not necessarily equal, rather than using $\widehat{b}_{n1} = \widehat{b}_{n2}$ as we did in this paper. While the asymptotic behavior of the corresponding confidence region might is expected to be similar to the one from this paper, the finite sample performance might be improved.

5 Appendix

5.1 Proof of Lemma 2.1.

The proof immediately follows from the following two results.

Lemma 5.1 (lower bound) *For any values of $\beta_n, b \geq 0$ we have with A_n as in Lemma 2.1 that*

$$\big[\,\{Z_n \leq b\} \cap A_n\,\big] \subset \big[\,\{\,\widehat{C}^-(b) \subset C^-\,\} \cap \{\,C \subset \widehat{C}(-b)\,\}\,\big]. \tag{5.1}$$

Consequently, for any choice of $b \geq 0$ (whether dependent on n and/or random or not), we have

$$P\big(\widehat{C}^-(b) \subset C^- \text{and } C \subset \widehat{C}(-b)\big) \geq P\big(\,Z_n \leq b\,\big) - P(A_n^{\complement}). \tag{5.2}$$

Proof. Let $S_n = \sup_{x\in\Delta_n}(\widehat{h}_n(x) - h(x))$ and $I_n = \inf_{x\in\Delta_n}(\widehat{h}_n(x) - h(x))$. First consider the set

$$\big\{S_n \le b\big\} = \big\{\, \widehat{h}_n(x_0) \le h(x_0) + b \ \text{ for all } \ x_0 \in \{x : -\beta_n \le h(x) \le \beta_n\}\,\big\}.$$

We show first that

$$\big[\,\big\{S_n \le b\big\} \cap A_n\,\big] \subset \{\widehat{C}^-(b) \subset C^-\}. \tag{5.3}$$

To this end, we separately consider the two cases, $x_0 \in \widehat{C}^-(b) \cap \{x : -\beta_n \le h(x) \le \beta_n\}$ and $x_0 \in \widehat{C}^-(b) \cap \{x : -\beta_n \le h(x) \le \beta_n\}^{\complement}$. In the first case we have on $\{S_n \le b\}$ that $x_0 \in C^-$ because

$$h(x_0) > h(x_0) - \widehat{h}_n(x_0) + b \ge b - b = 0.$$

As for the second case, observe that $\{-\beta_n \le h(x) \le \beta_n\}^{\complement} = \{h(x) < -\beta_n\} \cup \{h(x) > \beta_n\}$, and on A_n we have $\{h(x) < -\beta_n\} = \{\widehat{h}_n(x) < \widehat{h}_n(x) - h(x) - \beta_n\} \subset \{\widehat{h}_n(x) < 0\}$ so that

$$\widehat{C}^-(b) \cap \{-\beta_n \le h(x) \le \beta_n\}^{\complement} = \widehat{C}^-(b) \cap \{h(x) > \beta_n\} \subset C^-,$$

and (5.3) is verified. The fact that

$$\big[\,\{I_n \ge -b\} \cap A_n\big] \subset \{C \subset \widehat{C}(-b)\} \tag{5.4}$$

can be seen similarly. First recall that

$$\{I_n \ge -b\} = \Big\{\, \widehat{h}_n(x_0) \ge h(x_0) - b \ \text{ for all } \ x_0 \in \{-\beta_n \le h(x) \le \beta_n\}\,\Big\}.$$

Let $x_0 \in C \cap \{-\beta_n \le h(x) \le \beta_n\} = \{0 \le h(x) \le \beta_n\}$. Then, on $\{I_n \ge -b\}$ we have $\widehat{h}_n(x_0) \ge h(x_0) - b \ge -b$, and thus $x_0 \in \widehat{C}(-b)$. If $x_0 \in C \cap \{-\beta_n \le h(x) \le \beta_n\}^{\complement} = \{h(x) \ge \beta_n\}$, then we have on A_n that

$$\widehat{h}_n(x) = \widehat{h}_n(x_0) - h(x_0) + h(x_0) \ge -\beta_n + h(x_0) \ge 0 \ge -b,$$

and (5.4) follows. Putting together (5.3) and (5.4) we obtain

$$\begin{aligned} P\big[Z_n \le b\big] = P\big[\{S_n \le b \ \text{ and } \ I_n \ge -b\}\big] &\le P\big[\{S_n \le b \ \text{ and } \ I_n \ge -b\} \cap A_n\big] + P(A_n^{\complement}) \\ &\le P\big[\,\widehat{C}(b) \subset C^- \text{and } \ C \subset \widehat{C}(-b)\,\big] + P(A_n^{\complement}). \end{aligned}$$

□

Lemma 5.2 (Upper bound) *For any choice of $b \geq 0$ we have*

$$\big[\, \widehat{C}^-(b) \subset C^- \text{and } C \subset \widehat{C}(-b) \,\big] \subset \big\{ \sup_{x \in \partial C} \big|\widehat{h}_n(x) - h(x)\big| \leq b \big\}. \tag{5.5}$$

Proof. The assertion is straightforward to see by observing that $\{C \subset \widehat{C}(-b)\}$ in particular means that $\{\partial C \subset \widehat{C}(-b)\}$. Similarly, $\{\widehat{C}^-(b) \subset C^-\} = \{\big(C^-\big)^{\complement} \subset \big(\widehat{C}^-(b)\big)^{\complement}\}$ means that $\partial C \subset \big(\widehat{C}^-(b)\big)^{\complement}$. By definition of both $\widehat{C}(-b)$ and $\widehat{C}^-(b)$ the assertion follows.

□

5.2 Proof of Lemma 2.2.

Let $B_n = \Big\{ b_n^-(1-\alpha-\eta_n) - \gamma_{n1} \,\leq\, \widehat{b}_n \,\leq\, b_n^+(1-\alpha+\eta_n) + \gamma_{n1} \Big\}$. We have

$$\begin{aligned}
& \big|\, P\big(Z_n \leq \widehat{b}_n\big) - P\big(Z_n \leq b_n^+\big) \,\big| \\
\leq\; & \big|\, P\big[(Z_n \leq \widehat{b}_n) \cap\, B_n\big] - P\big[Z_n \leq b_n^+)\big] \,\big| + P\Big(\big(B_n\big)^{\complement}\Big) \\
\leq\; & P\big[b_n^-(1-\alpha-\eta_n) - \gamma_{n1} \leq Z_n < b_n^-(1-\alpha-\eta_n)\big] \\
& \quad + P\big[b_n^-(1-\alpha-\eta_n) \leq Z_n < b_n^+(1-\alpha+\eta_n)\big] \\
& \quad + P\big[b_n^+(1-\alpha+\eta_n) \leq Z_n \leq b_n^+(1-\alpha+\eta_n) + \gamma_{n1}\big] + P\Big(\big(B_n\big)^{\complement}\Big) \\
\leq\; & 2\tau_{n1} + 2\eta_n + \delta_{n1}
\end{aligned}$$

where the last identity is using (2.4) and (2.5). The statement of the theorem follows from this inequality and the fact that $P(b_n^- \leq Z_n < b_n^+) = 0$.

□

5.3 Proof of Lemma 2.3.

Observe that

$$A_n^* := \big\{ |Z_n^+ - Z_n^*| \leq \gamma_{n1} \big\} \subset \Big[\big\{ Z_n^+ \leq \widehat{b}_n - \gamma_{n1} \big\} \subset \big\{ Z_n^* \leq \widehat{b}_n \big\} \subset \big\{ Z_n^+ \leq \widehat{b}_n + \gamma_{n1} \big\} \Big]. \tag{5.6}$$

Let $d_n = 1 - P^*\big(A_n^*\big)$ and define d_n^* such that $P^*(Z_n^* \leq \widehat{b}_n) = 1 - \alpha + d_n^*$. First observe that $|d_n^*|$ is bounded by the maximum jump-size of the distribution of Z_n^*. We have the following bound for

the maximum jump-sizes:

$$\begin{aligned} & P^*(Z_n^* = t) \\ \leq\ & P^*(Z_n^+ \in [t - |Z_n^* - Z_n^+|,\, t + |Z_n^* - Z_n^+|) \\ \leq\ & P^*(Z_n^+ \in [t - \gamma_{n1},\, t + \gamma_{n1}],\, A_n^*) + d_n \\ \leq\ & 2\tau_{n1} + d_n. \end{aligned}$$

The last inequality uses assumption (2.4). Now we show that (2.5) holds.

$$\begin{aligned} P^*(Z_n^+ \leq \widehat{b}_n - \gamma_{n1}) &\leq P^*\big(\{Z_n^+ \leq \widehat{b}_n - \gamma_{n1}\} \cap A_n^*\big) + d_n \\ &\leq P^*\big(\{Z_n^* \leq \widehat{b}_n\} \cap\ A_n^*\big) + d_n \\ &\leq P^*(Z_n^* \leq \widehat{b}_n) + d_n, \\ &\leq 1 - \alpha + 2\tau_{n1} + 2d_n, \end{aligned}$$

which implies that

$$b_n^+(1 - \alpha + 2\tau_{n1} + 2d_n) \geq \widehat{b}_n - \gamma_{n1}. \tag{5.7}$$

Similarly, we obtain

$$\begin{aligned} P^*\big(Z_n^+ \leq \widehat{b}_n + \gamma_{n1}\big) &\geq P^*\big(\{Z_n^+ \leq \widehat{b}_n + \gamma_{n1}\} \cap A_n^*\big) \\ &\geq P^*(Z_n^* \leq \widehat{b}_n) + P^*\big(\{Z_n^* \leq \widehat{b}_n\} \cap A_n^*\big) - P^*(Z_n^* \leq \widehat{b}_n), \\ &\geq 1 - \alpha + d_n^* - d_n \\ &\geq 1 - \alpha - 2\tau_{n1} - 2d_n, \end{aligned}$$

and thus we have

$$b_n^-(1 - \alpha - 2\tau_{n1} - 2d_n) \leq \widehat{b}_n + \gamma_{n1}. \tag{5.8}$$

Further, $0 \leq d_n \leq 1$ and by assumption $Ed_n = P(|Z_n^+ - Z_n^*| \geq \gamma_{n1}) \leq \delta_n'$ for n sufficiently large. Thus, $P(d_n \geq \sqrt{\delta_n'}) \leq \sqrt{\delta_n'}$. Using the monotonicity of $\beta \to b_n(\beta)$ we obtain that with probability $\geq 1 - \sqrt{\delta_n'}$

$$b_n^-\big(\,1 - \alpha - 2\tau_{n1} - 2\sqrt{\delta_n'}\,\big) - \gamma_{n1}\ \leq\ \widehat{b}_n\ \leq\ b_n^+\big(\,1 - \alpha + 2\tau_{n1} + 2\sqrt{\delta_n'}\,\big) + \gamma_{n1}.$$

In other words, (2.5) holds with γ_{n1}, $\eta_n = 2\tau_{n1} + 2\sqrt{\delta_n'}$ and $\delta_{n1} = \sqrt{\delta_n'}$. Lemma 2.2 implies the asserted estimate of $r_n^{\pm}$.

□

5.4 Proof of Lemma 2.4.

Let $B'_n = \big\{|Z_n - Z_n^0| \leq \gamma_{n2}\big\}$ and let B_n be defined as the proof of Theorem 2.2. We have

$$
\begin{aligned}
& |\, P(Z_n \leq \widehat{b}_n) - P(Z_n^0 \leq \widehat{b}_n)| \\
\leq\ & \Big| P\big(\, \{Z_n \leq \widehat{b}_n\} \cap B'_n \,\big) - P\big(\, \{Z_n^0 \leq \widehat{b}_n\} \cap B'_n \,\big) \Big| + \delta_{n2} \\
\leq\ & P\big(\, \widehat{b}_n - \gamma_{n2} \leq Z_n \leq \widehat{b}_n + \gamma_{n2} \,\big) + \delta_{n2} \\
\leq\ & P\big(\, \{\widehat{b}_n - \gamma_{n2} \leq Z_n \leq \widehat{b}_n + \gamma_{n2}\} \cap\ B_n \,\big) + \delta_{n1} + \delta_{n2} \\
\leq\ & P\big(\, b_n^-(1-\alpha-\eta_n) - \gamma_{n1} - \gamma_{n2} \ \leq Z_n \ \leq\ b_n^+(1-\alpha+\eta_n) + \gamma_{n1} + \gamma_{n2}) \,\big) + \delta_{n1} + \delta_{n2} \\
\leq\ & 2\tau_{n1} + 2\tau_{n2} + 2\eta_n + \delta_{n1} + \delta_{n2}.
\end{aligned}
$$

The last inequality uses a similar reasoning as in the proof of Lemma 2.2.

□

References

[1] Adler, R. J. and Taylor, J. (2007): *Random Fields and Geometry.* Springer, Berlin.

[2] Alexander, K.S. (1984): Probability inequalities for empirical processes and a law of iterated logarithm. *Ann. Probab.* **12**, 1041 – 1067.

[3] Audibert, J-Y. and Tsybakov, A. (2007): Fast learning rates for plug-in classifiers. *Ann. Statist.* **35**, 608-633.

[4] Baillo, A., Cuevas, A. and Justel A. (2000). Set estimation and nonparametric detection. *Canad. J. Statist.* **28**, 765-782.

[5] Baillo A., Cuestas-Albertos, J.A., and Cuevas, A. (2001). Convergence rates in nonparametric estimation of level sets. *Stat. Probab. Lett.* **53**, 27-35.

[6] Baillo, A. (2003). Total error in a plug-in estimator of level sets. *Stat. Probab. Lett.***65**, 411-417.

[7] Biau, G., Cadre, B. and Pelletier, B. (2008). Exact rates in density support estimation.*J. Multivariate Anal.* **99**, 2185–2207.

[8] Bugni, F. (2010): Bootstrap inference in partially identified models defined by moment inequalities: coverage of the identified set. *Econometrica* **76**, 735-753.

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

[10] Cavalier, L. (1997). Nonparametric estimation of regression level sets. *Statistics* **29**, 131-160.

[11] Chernozhukov, V., Hong, H. and Tamer, E. (2007): Estimation and confidence regions for parameter sets in econometric models. *Econometrica* **75**, 1243-1284.

[12] Chernozhukov, V., Lee, S. and Rosen, A.M. (2012): Intersection bounds estimation and inference. To appear in *Econometrica*.

[13] Cuevas, A. and Fraiman, R. (1997): A plug-in approach to support estimation. *Ann. Statist.* **25**, 2300-2312.

[14] Cuevas, A., Febrero, M., and Fraiman, R. (2000). Estimating the number of clusters. *Canad. J. Statist.* **28**, 367-382.

[15] Cuevas, A., Gonzalez-Manteiga, W., and Rodriguez-Casal, A. (2006). Plug-in estimation of general level sets. *Australian and New Zealand J. Statist.* **48**(1), 7-19.

[16] Desforges, M.J., Jacob, P.J. and Cooper, J.E. (1998). Application of probability density estimation to the detection of abnormal conditions in engineering. In: *Proceedings of the Institute of Mechanical Engineering* **212**, 687-703.

[17] Duong, T., Koch, I. and Wand, M.P. (2009): Highest density difference region estimation with application to flow cytometry data. *Biometrical Journal* **51**, 504-521.

[18] Efron, B. (1979): Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

[19] Gayraud, G. and Rousseau, J. (2005). Rates of convergence for a Bayesian level set estimation. *Scand. J. Statist.* **32**(4), 639 - 660.

[20] Klemelä, J. (2004). Visualization of multivariate density estimates with level set trees. *J. Comput. Graph. Statist.* **13**, 599-620.

[21] Klemelä, J. (2006). Visualization of multivariate density estimates with shape trees. *J. Comput. Graph. Statist.* **15**, 372-397.

[22] Klemelä, J. (2009). *Smoothing of Multivariate Data: Density Estimation and Visualization.* Wiley, Hoboken, New Jersey.

[23] Jang, W. (2006). Nonparametric density estimation and clustering in astronomical sky survey. *Comp. Statist. & Data Anal.* **50**, 760 - 774.

[24] Jankowski, H.K. and Stanberry, L.I. (2011): Confidence regions for means of random sets using oriented distance functions. *arXiv:0903:1869v2*

[25] Hall, P. and Kang, K-H. (2005). Bandwidth choice for nonparametric classification. *Ann. Statist.* **33**, 284-306.

[26] Mammen, E. and Tsybakov, A.B. (1999): Smooth discrimination analysis. *Ann. Statist.* **27**, 1808-1829.

[27] Mason, D. and Polonik, W. (2009): Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.* **19**, 1108-1142.

[28] Molchanov, I. (1998): A limit theorem for solutions of inequalities. *Scan. J. Statist.* **25**, 235 - 242.

[29] Neumann, M.H. (1998): Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.* **26**, 2014 – 2048.

[30] Polonik, W. (1995). Measuring mass concentration and estimating density contour clusters: an excess mass approach. *Ann. Statist.* **23** 855-881.

[31] Rigollet, P. and Vert, R. (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15**, 1154-1178.

[32] Rinaldo, A., Singh, A., Nugent, R. and Wasserman, L. (2010): Stability of density-based clustering. arXiv:1011.2771v1

[33] Rio, E. (1994): Local invariance principles and their application to density estimation. *Probab. Theory Relat. Fields* **98**, 21 – 45

[34] Samworth R. and Wand, M.P. (2010): Asymptotics and optimal bandwidth selection for highest density region estimation. *Ann. Statist.* **38**, 1767-1792.

[35] Scott, C.D. and Davenport, M. (2007). Regression level set estimation via cost-sensitive classification. *IEEE Trans. Signal Proc.* **55**(6), 2752-2757.

[36] Scott, C.D. and Nowak, R.D. (2006). Learning minimum volume sets. *J. Machine Learning Research* **7**, 665-704.

[37] Steinwart, I., Hush, D. and Scovel, C. (2005). A classification framework for anomaly detection. *J. Machine Learning Research* **6**, 211-232.

[38] Stuetzle, W. (2003). Estimating the cluster tree of a density by analyzing the minimum spanning tree of a sample. *J. Classification* **20**(5), 25 - 47.

[39] Stuetzle, W. and Nugent, R. (2010): A generalized single linkage method for estimating the cluster tree of a density. *J. Comput. Graph. Statist.* **19**, 397-418.

[40] Tsybakov, A.B. (1997): Nonparametric estimation of density level sets. *Ann. Statist.* **25**, 948-969.

[41] Vogel, S. (2008): Confidence sets and convergence of random functions. In: *Festschrift in Celebration of Prof. Dr. Wilfried Grecksch's 60th Birthday; C. Tammer, F. Heyde (edit.)*, Shaker, Aachen.

[42] Walther, G. (1997). Granulometric smoothing. *Ann. Statist.* **25**, 2273 - 2299.

[43] Willett, R.M. and Nowak, R.D. (2005). Level Set Estimation in Medical Imaging, *Proceedings of the IEEE Statistical Signal Processing, Vol. 5*, 1089 - 1092.

[44] Willett, R.M. and Nowak, R.D. (2006). Minimax optimal level set estimation. Minimax optimal level set estimation. *IEEE Trans. Image Proc.* **16**(12), 2965-2979.