

Estimating level sets of a distribution function using a plug-in method: a multidimensional extension

Elena Di Bernardino¹, Thomas Laloë²

Abstract

This paper deals with the problem of estimating the level sets $L(c) = \{F(x) \geq c\}$, with $c \in (0, 1)$, of an unknown distribution function F on \mathbb{R}_+^d . A plug-in approach is followed. That is, given a consistent estimator F_n of F , we estimate $L(c)$ by $L_n(c) = \{F_n(x) \geq c\}$. We state consistency results with respect to the Hausdorff distance and the volume of the symmetric difference. These results can be considered as generalizations of results previously obtained, in a bivariate framework, in Di Bernardino *et al.* (2011). Finally we investigate the effects of scaling data on our consistency results.

Keywords: Level sets, multidimensional distribution function, plug-in estimation, Hausdorff distance.

Introduction

In this present paper, we consider the problem of estimating the level sets of a d -variate distribution function. To this aim, we generalize the results obtain in a previous paper (Di Bernardino *et al.*, 2011).

As yet remarked in Di Bernardino *et al.* (2011), considering the level sets of a distribution function, the commonly assumed property of compactness for these sets is no more reasonable. Then, differing from the classical literature (Cavalier, 1997; Cuevas and Fraiman, 1997; Baíllo *et al.*, 2001; Baíllo, 2003; Cuevas *et al.*, 2006; Biau *et al.*, 2007; Laloë, 2009), we need to work in a non-compact setting and this requires special attention in the statement of our problem.

We follow the same general approach than in Di Bernardino *et al.* (2011), and we will keep as much as possible the same notation. Considering a consistent estimator F_n of the distribution function F , we propose a plug-in approach (e.g. see Baíllo *et al.*, 2011; Rigollet and Vert, 2009; Cuevas *et al.*, 2006) to estimate the level set

$$L(c) = \{x \in \mathbb{R}_+^d : F(x) \geq c\}, \quad \text{for } c \in (0, 1),$$

by

$$L_n(c) = \{x \in \mathbb{R}_+^d : F_n(x) \geq c\}, \quad \text{for } c \in (0, 1).$$

The regularity properties of F and F_n as well as the consistency properties of F_n will be specified in the statements of our theorems.

As in Di Bernardino *et al.* (2011) our consistency results are stated with respect to two criteria of “physical proximity” between sets: the Hausdorff distance and volume of the symmetric difference.

¹Université de Paris X Ouest, UFR SEGMI, 200 Avenue de la République, 92000 Nanterre, France. elena.dibernardino@univ-lyon1.fr, <http://isfaserveur.univ-lyon1.fr/elena.dibernardino/>.

²Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné, Parc Valrose, 06108 Nice Cedex 02. Thomas.Laloe@unice.fr, <http://math.unice.fr/~laloe/>

If the consistency in term of the Hausdorff distance is a trivial extension of Theorem 2.1 in Di Bernardino *et al.* (2011) (see Theorem 2.1 below), things are a little more complex for the volume of the symmetric difference. In particular, in this latter case the convergence rate suffers from the well-known *curse of dimensionality* (see Theorem 3.1).

A second aim of this paper is to analyze the effects of scaling data on our consistency results (see Theorem 4.1).

The paper is organized as follows. We introduce some notation, tools and technical assumptions in Section 1. Consistency and asymptotic properties of our estimator of $L(c)$ are given in Sections 2 and 3. Section 4 is devoted to investigate the effects of scaling data on our consistency results. Finally, proofs are postponed to Section 5.

1. Notation and preliminaries

In this section we introduce some notation and tools which will be useful later.

Let $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\mathbb{R}_+^* = \mathbb{R}_+ \setminus \{0\}$ and $\mathbb{R}_+^{d*} = \mathbb{R}_+^d \setminus \{0\}$. Let \mathcal{F} be the set of continuous distribution functions $\mathbb{R}_+^d \rightarrow [0, 1]$ and $\mathbf{X} := (X_1, X_2, \dots, X_d)$ a random vector with distribution function $F \in \mathcal{F}$. Given an *i.i.d* sample $\{\mathbf{X}_i\}_{i=1}^n$ in \mathbb{R}_+^d with distribution function $F \in \mathcal{F}$, we denote by F_n an estimator of F based on this finite sample.

Define, for $c \in (0, 1)$, the *upper c-level set* of $F \in \mathcal{F}$ and its plug-in estimator

$$L(c) = \{x \in \mathbb{R}_+^d : F(x) \geq c\}, \quad L_n(c) = \{x \in \mathbb{R}_+^d : F_n(x) \geq c\},$$

and

$$\{F = c\} = \{x \in \mathbb{R}_+^d : F(x) = c\}.$$

In addition, given $T > 0$, we set

$$L(c)^T = \{x \in [0, T]^d : F(x) \geq c\}, \quad L_n(c)^T = \{x \in [0, T]^d : F_n(x) \geq c\},$$

$$\{F = c\}^T = \{x \in [0, T]^d : F(x) = c\}.$$

Given a set $A \subset \mathbb{R}_+^d$ we denote by ∂A its boundary, and by βA the scaled set $\{\beta x, \text{ with } x \in A\}$.

Note that, in the presence of a plateau at level c , $\{F = c\}$ can be a portion of quadrant \mathbb{R}_+^d instead of a set of Lebesgue measure null in \mathbb{R}_+^d . In the following we introduce suitable conditions in order to avoid this situation.

We denote by $B(x, \rho)$ the closed ball centered on $x \in \mathbb{R}_+^d$ and with positive radius ρ . Let $B(S, \rho) = \bigcup_{x \in S} B(x, \rho)$, with S a closed set of \mathbb{R}_+^d .

For $r > 0$ and $\zeta > 0$, define

$$E = B(\{x \in \mathbb{R}_+^d : |F(x) - c| \leq r\}, \zeta),$$

and, for a twice differentiable function F ,

$$m^\nabla = \inf_{x \in E} \|(\nabla F)_x\|, \quad M_H = \sup_{x \in E} \|(HF)_x\|,$$

where $(\nabla F)_x$ is the gradient vector of F evaluated at x and $\|(\nabla F)_x\|$ its Euclidean norm, $(HF)_x$ the Hessian matrix evaluated in x and $\|(HF)_x\|$ its matrix norm induced by the Euclidean norm.

For sake of completeness, we recall that if A_1 and A_2 are compact sets in \mathbb{R}_+^d , the Hausdorff distance between A_1 and A_2 is defined by

$$d_H(A_1, A_2) = \max \left\{ \sup_{x \in A_1} d(x, A_2), \sup_{x \in A_2} d(x, A_1) \right\},$$

where $d(x, A_2) = \inf_{y \in A_2} \|x - y\|$.

The above expression is well defined even when A_1 and A_2 are just closed (not necessarily compact) sets but, in this case, the value $d_H(A_1, A_2)$ could be infinity. Then in our setting, in order to avoid these situations, we introduce the following assumption.

H: There exist $\gamma > 0$ and $A > 0$ such that, if $|t - c| \leq \gamma$ then $\forall T > 0$ such that $\{F = c\}^T \neq \emptyset$ and $\{F = t\}^T \neq \emptyset$,

$$d_H(\{F = c\}^T, \{F = t\}^T) \leq A |t - c|.$$

For further details about this assumption the interest reader is referred to Di Bernardino *et al.* (2011), Cuevas *et al.* (2006), Tsybakov (1997). Remark that a sufficient condition for Assumption **H** can be obtained in terms of the differentiability properties of F . Proposition 1.1 below is a trivial extension in d -variate setting of Proposition 1.1 in Di Bernardino *et al.* (2011).

Proposition 1.1 *Let $c \in (0, 1)$. Let $F \in \mathcal{F}$ be twice differentiable on \mathbb{R}_+^{d*} . Assume there exist $r > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Then F satisfies Assumption **H**, with $A = \frac{2}{m^\nabla}$.*

Remark 1 Under assumptions of Proposition 1.1, $\{F = t\}$ is a set of Lebesgue measure null in \mathbb{R}_+^d . Furthermore we obtain $\partial L(c)^T = \{F = c\}^T = \{F = c\} \cap [0, T]^d$ (we refer for details to Remark 1 in Di Bernardino *et al.*, 2011 and Theorem 3.2 in Rodríguez-Casal, 2003).

2. Consistency in terms of the Hausdorff distance

In this section we study the consistency properties of $L_n(c)^T$ with respect to the Hausdorff distance between $\partial L_n(c)^T$ and $\partial L(c)^T$.

From now on we note, for $n \in \mathbb{N}^*$,

$$\|F - F_n\|_\infty = \sup_{x \in \mathbb{R}_+^d} |F(x) - F_n(x)|,$$

and for $T > 0$

$$\|F - F_n\|_\infty^T = \sup_{x \in [0, T]^d} |F(x) - F_n(x)|.$$

The following result can be considered a trivially adapted version of Theorem 2.1 in Di Bernardino *et al.* (2011).

Theorem 2.1 *Let $c \in (0, 1)$. Let $F \in \mathcal{F}$ be twice differentiable on \mathbb{R}_+^{d*} . Assume that there exist $r > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Let $T_1 > 0$ such that for all $t : |t - c| \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Let $(T_n)_{n \in \mathbb{N}^*}$ be an increasing sequence of positive values. Assume that, for each n and for almost all samples of size n , F_n is a continuous function and that*

$$\|F - F_n\|_\infty \rightarrow 0, \quad a.s.$$

Then, for n large enough,

$$d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) \leq 6 A \|F - F_n\|_\infty^{T_n} \quad a.s.,$$

where $A = \frac{2}{m^\nabla}$. Therefore we have

$$d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) = O(\|F - F_n\|_\infty) \quad a.s.$$

Under assumptions of Theorem 2.1, $d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n})$ converges to zero and the quality of our plug-in estimator is obviously related to the quality of the estimator F_n . For comments and discussions about this result we refer the interested reader to Remark 2 in Di Bernardino *et al.* (2011).

3. L_1 consistency

The previous section was devoted to the consistency of $L_n(c)$ in terms of the Hausdorff distance. We consider now another consistency criterion: the consistency of the volume (in the Lebesgue measure sense) of the symmetric difference between $L(c)^{T_n}$ and $L_n(c)^{T_n}$. This means that we define the distance between two subsets A_1 and A_2 of \mathbb{R}_d^+ by

$$d_\lambda(A_1, A_2) = \lambda(A_1 \triangle A_2),$$

where λ stands for the Lebesgue measure on \mathbb{R}^d and \triangle for the symmetric difference.

Let us introduce the following assumption:

A1 There exist positive increasing sequences $(v_n)_{n \in \mathbb{N}^*}$ and $(T_n)_{n \in \mathbb{N}^*}$ such that

$$v_n \int_{[0, T_n]^d} |F - F_n|^p \lambda(dx) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

for some $1 \leq p < \infty$.

We now establish our consistency result with convergence rate, in terms of the volume of the symmetric difference. We can interpret the following theorem as an extension of Theorem 3.1 in Di Bernardino *et al.* (2011), in the case of a d -variate distribution function F .

Theorem 3.1 *Let $c \in (0, 1)$. Let $F \in \mathcal{F}$ be a twice differentiable distribution function on \mathbb{R}_+^{d*} . Assume that there exist $r > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Assume that for each n , with probability one, F_n is measurable. Let $(v_n)_{n \in \mathbb{N}^*}$ and $(T_n)_{n \in \mathbb{N}^*}$ positive increasing sequences such that Assumption **A1** is satisfied and that for all $t : |t - c| \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Then, it holds that*

$$p_n d_\lambda(L(c)^{T_n}, L_n(c)^{T_n}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

with p_n an increasing positive sequence such that $p_n = o\left(\frac{1}{v_n^{p+1}} / T_n^{\frac{(d-1)p}{p+1}}\right)$.

The proof is postponed to Section 5. This demonstration is basically based on the proof of Theorem 3.1 in Di Bernardino *et al.* (2011).

Theorem 3.1 provides a convergence rate, which is closely related to the choice of the sequence T_n . Note that, as in Theorem 3 in Cuevas *et al.* (2006), Theorem 3.1 above does not require any continuity assumption on F_n . Furthermore, as in Theorem 3.1 in Di Bernardino *et al.* (2011),

we remark that a sequence T_n , whose divergence rate is large, implies a convergence rate p_n quite slow. Moreover, this phenomenon is emphasized by the dimension d of the data, and we face here the well-known *curse of dimensionality*. In the following we will illustrate this aspect by giving convergence rate in the case of the empirical distribution function (see Example 1). Firstly, from Theorem 3.1 we can derive the following result.

Corollary 3.1 *Under the assumptions and notations of Theorem 3.1. Assume that there exists a positive increasing sequence $(v_n)_{n \in \mathbb{N}^*}$ such that $v_n \|F - F_n\|_\infty \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$. Then, it holds that*

$$p_n d_\lambda(L(c)^{T_n}, L_n(c)^{T_n}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

with p_n an increasing positive sequence such that $p_n = o\left(v_n \frac{p}{p+1} / T_n^{\frac{d+(d-1)p}{p+1}}\right)$.

This result comes trivially from Theorem 3.1 and the fact that that $v_n \|F - F_n\|_\infty \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$ implies

$$\forall p \geq 1, \quad w_n \int_{[0, T_n]^d} |F - F_n|^p \lambda(dx) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0, \quad \text{with} \quad w_n = \frac{v_n^p}{T_n^d}.$$

Let us now present a more practical example.

Example 1 (The empirical distribution function case)

Let F_n the d -variate empirical distribution function. Then, it holds that $v_n \|F - F_n\|_\infty \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$, with $v_n = o(\sqrt{n})$. From Theorem 3.1, with $p = 2$, we obtain for instance:

$$p_n = o\left(\frac{n^{1/3}}{T_n^{7/3}}\right), \quad \text{for } d = 3; \quad p_n = o\left(\frac{n^{1/3}}{T_n^{10/3}}\right), \quad \text{for } d = 4.$$

The next section is dedicated to study the effects of scaling data.

4. About the effects of scaling data

Suppose now to scale our data using a scale parameter $a \in \mathbb{R}_+^*$. In our case, the scaled random vector will be $(a X_1, a X_2, \dots, a X_d) := a \mathbf{X}$. From now on we denote $F_{a \mathbf{X}}$ (resp. $F_{\mathbf{X}}$) the distribution function associated to $a \mathbf{X}$ (resp. to \mathbf{X}). Using notation of Section 1, let

$$L_a(c) = \{x \in \mathbb{R}_+^d : F_{a \mathbf{X}}(x) \geq c\}.$$

It is easy to prove (see for instance Section 3 in Tibiletti, 1993) that

$$L_a(c) = a L(c),$$

and

$$E_a = B(\{x \in \mathbb{R}_+^d : |F_{a \mathbf{X}}(x) - c| \leq r\}, \zeta) = a E.$$

Define now

$$m_a^\nabla = \inf_{x \in E_a} \|\nabla F_{a \mathbf{X}}(x)\|.$$

First, we can obtain the following result whose proof is postponed to Section 5.

Lemma 4.1 *It holds that*

$$m_a^\nabla = \frac{1}{a} m^\nabla, \quad \forall a \in \mathbb{R}_+^*.$$

Furthermore, if

$$M_H = \sup_{x \in E} \|(HF_{\mathbf{X}})_x\| < +\infty \quad \text{then} \quad M_{H,a} = \sup_{x \in aE} \|(HF_{a\mathbf{X}})_x\| < +\infty, \quad \text{with } a \in \mathbb{R}_+^*.$$

We can now consider the effects of scaling data on Theorem 2.1 and 3.1.

Theorem 4.1

1. *Under same notation and assumptions of Theorem 2.1, for n large enough, it holds that*

$$d_H(\partial L_a(c)^{aT_n}, \partial L_{n,a}(c)^{aT_n}) \leq 6 A a \|F - F_n\|_\infty^{T_n}, \quad \text{a.s.}$$

2. *Under same notation and assumptions of Theorem 3.1 it holds that*

$$p_{n,a} d_\lambda(L_a(c)^{aT_n}, L_{n,a}(c)^{aT_n}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

with $p_{n,a}$ an increasing positive sequence such that $p_{n,a} = o\left(v_n^{\frac{1}{p+1}} / \left(a^{\frac{dp}{p+1}} T_n^{\frac{(d-1)p}{p+1}}\right)\right)$.

Remark 2

1. The first result of Theorem 4.1 states that a change of scale of the data implies the same change of scale for the Hausdorff distance.
2. The second result states that a change of scale of the data implies a rate in

$$o\left(v_n^{\frac{1}{p+1}} / \left(a^d T_n^{(d-1)}\right)^{p/(p+1)}\right)$$

instead of

$$o\left(v_n^{\frac{1}{p+1}} / \left(T_n^{(d-1)}\right)^{\frac{p}{p+1}}\right).$$

So, we see logically that the scale factor a impacts the volume in \mathbb{R}^d with an exponent d .

Conclusion

Starting from previous results obtained in Di Bernardino *et al.* (2011), we propose in this paper a generalization to the estimation of level sets in the case of a d -variate distribution function. The consistency results are stated in term of Hausdorff distance and volume of the symmetric difference. We propose a rate of convergence for this second criterion. Moreover, we analyze the impact of scaling data on our results. As a future work, a complete simulation study and an R-package are in preparation.

5. Proofs

Proof of Theorem 3.1

Under assumptions of Theorem 3.1, we can always take $T_1 > 0$ such that for all $t : |t - c| \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Then for each n , for all $t : |t - c| \leq r$, $\partial L(t)^{T_n}$ is a non-empty (and compact) set on \mathbb{R}_+^d .

We consider a positive sequence ε_n such that $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$. For each $n \geq 1$ the random sets $L(c)^{T_n} \triangle L_n(c)^{T_n}$, $Q_{\varepsilon_n} = \{x \in [0, T_n]^d : |F - F_n| \leq \varepsilon_n\}$ and $\tilde{Q}_{\varepsilon_n} = \{x \in [0, T_n]^d : |F - F_n| > \varepsilon_n\}$ are measurable and

$$\lambda(L(c)^{T_n} \triangle L_n(c)^{T_n}) = \lambda(L(c)^{T_n} \triangle L_n(c)^{T_n} \cap Q_{\varepsilon_n}) + \lambda(L(c)^{T_n} \triangle L_n(c)^{T_n} \cap \tilde{Q}_{\varepsilon_n}).$$

Since $L(c)^{T_n} \triangle L_n(c)^{T_n} \cap Q_{\varepsilon_n} \subset \{x \in [0, T_n]^d : c - \varepsilon_n \leq F < c + \varepsilon_n\}$ we obtain

$$\lambda(L(c)^{T_n} \triangle L_n(c)^{T_n}) \leq \lambda(\{x \in [0, T_n]^d : c - \varepsilon_n \leq F < c + \varepsilon_n\}) + \lambda(\tilde{Q}_{\varepsilon_n}).$$

From Assumption **H** (Section 1) and Proposition 1.1, if $2\varepsilon_n \leq \gamma$ then

$$d_H(\partial L(c + \varepsilon_n)^{T_n}, \partial L(c - \varepsilon_n)^{T_n}) \leq 2\varepsilon_n A.$$

From assumptions on first derivatives of F (see Assumption **H** and Proposition 1.1) and Propriety 1 in Imlahi *et al.* (1999), we can write

$$\lambda(\{x \in [0, T_n]^d : c - \varepsilon_n \leq F < c + \varepsilon_n\}) \leq (2\varepsilon_n A) d T_n^{d-1}.$$

If we now choose

$$\varepsilon_n = o\left(\frac{1}{p_n T_n^{d-1}}\right), \quad (1)$$

we obtain that, for n large enough, $2\varepsilon_n \leq \gamma$ and

$$p_n \lambda(\{x \in [0, T_n]^d : c - \varepsilon_n \leq F < c + \varepsilon_n\}) \xrightarrow{n \rightarrow \infty} 0.$$

Let us now prove that $p_n \lambda(\tilde{Q}_{\varepsilon_n}) \xrightarrow{n \rightarrow \infty} 0$. To this end, we write

$$p_n \lambda(\tilde{Q}_{\varepsilon_n}) = p_n \int 1_{\{x \in [0, T_n]^d : |F - F_n| > \varepsilon_n\}} \lambda(dx) \leq \frac{p_n}{\varepsilon_n^p} \int_{[0, T_n]^d} |F - F_n|^p \lambda(dx).$$

Take ε_n such that

$$\varepsilon_n = \left(\frac{p_n}{v_n}\right)^{\frac{1}{p}}. \quad (2)$$

So, from Assumption **A1** in Section 3, we obtain $p_n \lambda(\tilde{Q}_{\varepsilon_n}) \xrightarrow{n \rightarrow \infty} 0$. As $p_n = o\left(\frac{1}{v_n^{\frac{1}{p+1}} / T_n^{\frac{(d-1)p}{p+1}}}\right)$ we can choose ε_n that satisfies (1) and (2). Hence the result. \square

Proof of Lemma 4.1

First, we remark that

$$F_{a\mathbf{X}}(x) = F_{\mathbf{X}}\left(\frac{x}{a}\right).$$

Then, we obtain

$$\begin{aligned}
m_a^\nabla &= \inf_{x \in aE} \left\| \left(\frac{\partial}{\partial x_1} F_{\mathbf{X}} \left(\frac{x}{a} \right), \dots, \frac{\partial}{\partial x_d} F_{\mathbf{X}} \left(\frac{x}{a} \right) \right) \right\|, \\
&= \inf_{x \in aE} \left\| \frac{1}{a} \left(\frac{\partial F_{\mathbf{X}}}{\partial x_1} \left(\frac{x}{a} \right), \dots, \frac{\partial F_{\mathbf{X}}}{\partial x_d} \left(\frac{x}{a} \right) \right) \right\|, \\
&= \frac{1}{a} \inf_{x \in E} \left\| \left(\frac{\partial F_{\mathbf{X}}}{\partial x_1}(x), \dots, \frac{\partial F_{\mathbf{X}}}{\partial x_d}(x) \right) \right\|. \\
&= \frac{1}{a} m^\nabla.
\end{aligned}$$

Second part of Lemma 4.1 comes down from trivial calculus. Hence the result. \square

Proof of Theorem 4.1

Proof of 1.

Following the proof of Theorem 2.1, it holds that

$$d_H(\partial L_{a\mathbf{X}}(c)^{aT_n}, \partial L_{n,a}(c)^{aT_n}) \leq 6 \frac{2}{m_a^\nabla} \sup_{x \in [0, aT_n]^d} |F_{\mathbf{X}} \left(\frac{x}{a} \right) - F_n \left(\frac{x}{a} \right)|.$$

Using Lemma 4.1 and the fact that

$$\sup_{x \in [0, aT_n]^d} |F_{\mathbf{X}} \left(\frac{x}{a} \right) - F_n \left(\frac{x}{a} \right)| = \sup_{x \in [0, T_n]^d} |F_{\mathbf{X}}(x) - F_n(x)|,$$

we get the result. \square

Proof of 2.

As in the proof of Theorem 3.1 and using same notation, we can write

$$\lambda(\{x \in [0, aT_n]^d : c - \varepsilon_n \leq F_{a\mathbf{X}} < c + \varepsilon_n\}) \leq (2\varepsilon_n A a) d a^{d-1} T_n^{d-1}.$$

If we now choose

$$\varepsilon_n = o\left(\frac{1}{p_{n,a} a^d T_n^{d-1}}\right) \quad (3)$$

we obtain that for n large enough $2\varepsilon_n \leq \gamma$ and

$$p_{n,a} \lambda(\{x \in [0, aT_n]^d : c - \varepsilon_n \leq F_{a\mathbf{X}} < c + \varepsilon_n\}) \xrightarrow{n \rightarrow \infty} 0.$$

The second part of this demonstration is equal to proof of Theorem 3.1. Then we take ε_n such that

$$\varepsilon_n = \left(\frac{p_{n,a}}{v_n}\right)^{\frac{1}{p}}. \quad (4)$$

Then, from Assumption **A1**, in Section 3, we obtain $p_{n,a} \lambda(\{x \in [0, aT_n]^d : |F_{a\mathbf{X}} - F_{a,n}| > \varepsilon_n\}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$. As $p_{n,a} = o\left(v_n^{\frac{1}{p+1}} / \left(a^{\frac{dp}{p+1}} T_n^{\frac{(d-1)p}{p+1}}\right)\right)$ we can choose ε_n that satisfies (3) and (4). Hence the result. \square

Acknowledgements: This work has been partially supported by the French research national agency (ANR) under the reference ANR-08BLAN-0314-01. The authors thank Yannick Baraud, Christine Tuleau-Malot and Patricia Reynaud-Bouret for fruitful discussions.

References:

References

- [1] A. Baíllo. Total error in a plug-in estimator of level sets. *Statist. Probab. Lett.*, 65(4):411–417, 2003.
- [2] A. Baíllo, J. Cuesta-Albertos, and A. Cuevas. Convergence rates in nonparametric estimation of level sets. *Statistics and Probability Letters*, 53:27–35, 2001.
- [3] G. Biau, B. Cadre, and B. Pelletier. A graph-based estimator of the number of clusters. *ESAIM. Probability and Statistics*, 11:272–280, 2007.
- [4] L. Cavalier. Nonparametric estimation of regression level sets. *Statistics. A Journal of Theoretical and Applied Statistics*, 29(2):131–160, 1997.
- [5] A. Cuevas and R. Fraiman. A plug-in approach to support estimation. *The Annals of Statistics*, 25(6):2300–2312, 1997.
- [6] A. Cuevas, W. González-Manteiga, and A. Rodríguez-Casal. Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19, 2006.
- [7] E. Di Bernardino, T. Laloë, V. Maume-Deschamps, and C. Prieur. Plug-in estimation of level sets in a non-compact setting with applications in multivariable risk theory. *ESAIM: Probability and Statistics*, <http://dx.doi.org/10.1051/ps/2011161>, 2011.
- [8] L. Imlahi, M. Ezzerg, and A. Chakak. Estimación de la curva mediana de una cópula $C(x_1, \dots, x_n)$. *Rev. R. Acad. Cien. Exact. Fis. Nat*, 93(2):241–250, 1999.
- [9] T. Laloë. *Sur Quelques Problèmes d'Apprentissage Supervisé et Non Supervisé*. PhD thesis, University Montpellier II, 2009.
- [10] P. Rigollet and R. Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 15(4):1154–1178, 2009.
- [11] A. Rodríguez-Casal. *Estimación de conjuntos y sus fronteras. Un enfoque geométrico*. PhD thesis, University of Santiago de Compostela, 2003.
- [12] L. Tibiletti. On a new notion of multidimensional quantile. *Metron. International Journal of Statistics*, 51(3-4):77–83, 1993.
- [13] A. B. Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 25(3):948–969, 1997.