



**HAL**  
open science

# Cube root weak convergence of empirical estimators of a density level set

Philippe Berthet, John H.J. Einmahl

► **To cite this version:**

Philippe Berthet, John H.J. Einmahl. Cube root weak convergence of empirical estimators of a density level set. *Annals of Statistics*, 2022, 50 (3), pp.1423-1446. hal-02732982v2

**HAL Id: hal-02732982**

**<https://hal.science/hal-02732982v2>**

Submitted on 28 Oct 2023

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

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

# Cube root weak convergence of empirical estimators of a density level set

Philippe Berthet and John H.J. Einmahl\*

*Institut de Mathématiques de Toulouse; UMR5219  
Université de Toulouse; CNRS  
UPS IMT, F-31062 Toulouse Cedex 9  
France*

*e-mail: [philippe.berthet@math.univ-toulouse.fr](mailto:philippe.berthet@math.univ-toulouse.fr)*

*Dept. of Econometrics and OR and CentER  
Tilburg University  
PO Box 90153, 5000 LE Tilburg  
The Netherlands  
e-mail: [j.h.j.einmahl@uvt.nl](mailto:j.h.j.einmahl@uvt.nl)*

## Abstract:

Given  $n$  independent random vectors with common density  $f$  on  $\mathbb{R}^d$ , we study the weak convergence of three empirical-measure based estimators of the convex  $\lambda$ -level set  $L_\lambda$  of  $f$ , namely the excess mass set, the minimum volume set and the maximum probability set, all selected from a class of convex sets  $\mathcal{A}$  that contains  $L_\lambda$ . Since these set-valued estimators approach  $L_\lambda$ , even the formulation of their weak convergence is non-standard. We identify the joint limiting distribution of the symmetric difference of  $L_\lambda$  and each of the three estimators, at rate  $n^{-1/3}$ . It turns out that the minimum volume set and the maximum probability set estimators are asymptotically indistinguishable, whereas the excess mass set estimator exhibits “richer” limit behavior. Arguments rely on the boundary local empirical process, its cylinder representation, dimension-free concentration around the boundary of  $L_\lambda$ , and the set-valued argmax of a drifted Wiener process.

**MSC2020 subject classifications:** Primary 62G05, 62G20; secondary 60F05, 60F17.

**Keywords and phrases:** Argmax drifted Wiener process, cube root asymptotics, density level set, excess mass, local empirical process, minimum volume set, set-valued estimator.

## 1. Introduction

### 1.1. Three level set estimators

Let  $X_1, \dots, X_n$ ,  $n \in \mathbb{N}$ , be independent and identically distributed random variables taking values in  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ , endowed with Lebesgue measure  $\mu$  and Borel sets  $\mathcal{B}(\mathbb{R}^d)$ . Assume that the law  $P$  of  $X_1$  is absolutely continuous with respect

---

\*John Einmahl holds the Arie Kapteyn Chair 2019–2022 and gratefully acknowledges the corresponding research support.

to  $\mu$  with continuous density  $f$ . We intend to establish novel, non-standard weak limit theorems for three set-valued estimators of a convex level set of  $f$ , treated as random sets rather than random vectors estimating the parameters of a set in a parametric class.

**Motivation.** Several classical problems in multivariate statistics involve set-valued estimators based on  $X_1, \dots, X_n$ . For instance, in order to detect areas having high probability  $P$ , to localize modes or clusters, to test for multimodality, to find outliers, or to test for goodness-of-fit to a family of distributions. In particular, many approaches and procedures rely on  $\lambda$ -level sets  $L_\lambda$  of the density  $f$  ( $\lambda > 0$ ). The plug-in method consists of using the corresponding level set of some density estimator. Alternatively, estimators of  $L_\lambda$  can be obtained by selecting a set in a class  $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^d)$  according to some optimization criterion applied directly to the empirical measure of  $X_1, \dots, X_n$ . Here we avoid density estimation and follow the latter approach. Note that maybe the most natural class of sets  $\mathcal{A}$  is the class of all closed ellipsoids. We will consider the classical nonparametric M-estimators of  $L_\lambda$  based on the following three criteria:

- excess mass,
- minimum volume, and
- maximum probability.

In particular, the first two criteria have been studied in the literature extensively. The third one is also very natural, since it is a kind of inverse of the minimum volume approach.

Seminal papers on the excess mass approach are Müller and Sawitzki (1991), Nolan (1991), Müller (1992), and Polonik (1995), and pioneering work on the minimum volume approach can be found in Silverman and Titterton (1980), Rousseeuw (1985), Davies (1992), and Polonik (1997). For the maximum probability approach we refer to Polonik (1998). For different, early approaches to the estimation of density level sets see Hartigan (1987) and Tsybakov (1997), and for more recent work, see, e.g., Cadre (2006), Cai, Einmahl and de Haan (2011), Chen, Genovese and Wasserman (2017), Brunel (2018), Rodríguez-Casal and Saavedra-Nieves (2019), and Xu and Samworth (2019). Statistical/machine learning approaches to the aforementioned criteria, include Cléménçon, Goix and Sabourin (2015) and Scott and Novak (2006). As far as asymptotic theory is concerned, the results in the literature regarding empirical estimators of the level sets study rates of convergence towards the true level set for appropriately defined distances. Other types of results consider weak convergence for estimators of the parameters of a parametrically defined level set.

The main goal of this paper is to deal with the weak convergence of the three classical, competing *set-valued estimators* of the level set  $L_\lambda$  *themselves* and look for their differences or similarities, jointly. Since these estimators approach  $L_\lambda$ , even the formulation of weak convergence is non-standard. Our main results are novel central limit theorems for the aforementioned three empirical-measure based estimators of  $L_\lambda$ , which reveal their interesting asymptotic behavior as random sets and provide the distribution of their limiting sets, obtained after

cube-root- $n$  magnification. The proofs raised various challenges as indicated in Subsection 1.2 below.

**Target level set.** Fix  $\lambda > 0$  throughout and assume that the level set

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

is a convex body, that is, it is convex, compact, and has non-empty interior, and that  $S_\lambda = \{x \in \mathbb{R}^d : f(x) = \lambda\}$  coincides with its boundary:  $S_\lambda = \partial L_\lambda$ . Note that  $f > \lambda$  on  $L_\lambda \setminus S_\lambda$  and  $f < \lambda$  on  $\mathbb{R}^d \setminus L_\lambda$ . Hence,

$$e_\lambda = p_\lambda - \lambda v_\lambda > 0, \text{ with } p_\lambda = P(L_\lambda) \in (0, 1) \text{ and } v_\lambda = \mu(L_\lambda) \in (0, 1/\lambda).$$

We denote the Hausdorff surface measure of  $S_\lambda$  by  $s_\lambda$  and have  $s_\lambda \geq c_d v_\lambda^{1-1/d} > 0$  by the isoperimetric inequality, with  $c_d > 0$ . Let  $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^d)$  be a class of closed, convex sets with  $L_\lambda \in \mathcal{A}$ . Then we have

$$\begin{aligned} L_\lambda &= \arg \max_{A \in \mathcal{A}} \{P(A) - \lambda \mu(A)\} \\ &= \arg \min_{A \in \mathcal{A}} \{\mu(A) : P(A) \geq p_\lambda\} \\ &= \arg \max_{A \in \mathcal{A}} \{P(A) : \mu(A) \leq v_\lambda\}, \end{aligned}$$

and the maximizing/minimizing set is unique. In other words, if  $\lambda$  is known then  $L_\lambda$  maximizes on  $\mathcal{A}$  the excess mass function  $A \mapsto e_\lambda(A) = P(A) - \lambda \mu(A)$ , if  $p_\lambda$  is known then  $L_\lambda$  minimizes on  $\{A \in \mathcal{A} : P(A) \geq p_\lambda\}$  the volume function  $A \mapsto \mu(A)$  and if  $v_\lambda$  is known then  $L_\lambda$  maximizes on  $\{A \in \mathcal{A} : \mu(A) \leq v_\lambda\}$  the probability mass function  $A \mapsto P(A)$ .

**Empirical level sets.** Let  $\delta_x$  denote the Dirac measure at  $x$ . From the non-parametric viewpoint it is natural to estimate  $P$  with the empirical measure  $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$  in the above argmax and argmin. To motivate a joint study, imagine that three statisticians want to estimate the level set  $L_\lambda$  by using the same sample  $X_1, \dots, X_n$ . Assume that they all know  $\mathcal{A}$  and that  $L_\lambda \in \mathcal{A}$ , but that they have their own private, auxiliary information. The first statistician knows the level  $\lambda$  and therefore makes use of the set-valued excess mass estimator

$$L_{1,n} \in \arg \max_{A \in \mathcal{A}} \{P_n(A) - \lambda \mu(A)\}. \quad (1.1)$$

The second one knows  $p_\lambda$  and then makes use of the minimum volume estimator

$$L_{2,n} \in \arg \min_{A \in \mathcal{A}} \{\mu(A) : P_n(A) \geq p_\lambda\}. \quad (1.2)$$

The third statistician knows  $v_\lambda$  and thus makes use of the maximum probability estimator

$$L_{3,n} \in \arg \max_{A \in \mathcal{A}} \{P_n(A) : \mu(A) \leq v_\lambda\}. \quad (1.3)$$

We assume that  $P$  and  $\mathcal{A}$  are such that almost surely an  $L_{1,n}$  and an  $L_{2,n}$  exist and that  $P_n(L_{2,n}) = \lceil np_\lambda \rceil / n$ . Since  $P_n$  takes at most  $n+1$  values, an  $L_{3,n}$  always

exists. If  $L_{j,n}$ ,  $j = 1, 2, 3$ , are not unique, just choose any maximizer/minimizer. This choice does not matter since it will be shown that all versions of  $L_{j,n}$  are indistinguishable asymptotically.

## 1.2. Overview of the results

What can be put forward before introducing more precisely our geometrical and probabilistic framework is as follows.

**Convergence of random sets.** In order to compare the performance of the empirical sets  $L_{j,n}$  we study the joint limiting behavior of  $L_{j,n} \triangle L_\lambda$ ,  $j = 1, 2, 3$ , where  $L \triangle L' = (L \cup L') \setminus (L \cap L')$  denotes the symmetric difference. The ensuing non-classical asymptotics for these set-valued estimators goes beyond the usual statistical risk approach which only provides rates for the random variables  $P(L_{j,n} \triangle L_\lambda)$  or  $\mu(L_{j,n} \triangle L_\lambda)$ , for  $j = 1, 2, 3$ . Instead we address the question of the weak convergence of the random sets  $L_{j,n} \triangle L_\lambda$  themselves. We then have to design an appropriate setting allowing to state central limit theorems for random sets, that is, for sets properly centered and then magnified at a diverging scale. Our joint limit results reveal, when magnifying with  $n^{1/3}$  as explained below, how the three empirical sets  $L_{j,n}$  asymptotically differ or coincide. In particular we find that  $L_{2,n}$  and  $L_{3,n}$  are asymptotically indistinguishable. Note that in the literature these limit theorems have been considered for dimension  $d = 1$  only, where the sets are intervals which can be represented by two numbers, like in the estimation of the shorth. Hence those limit theorems can be stated for real-valued random variables in the usual way. Our method for sets when  $d > 1$  is discussed next.

**A local empirical process approach and rate of convergence.** In order to analyze how the estimators  $L_{j,n}$  oscillate around  $L_\lambda$  we first show that they concentrate at rate  $n^{-1/3}$  under regularity conditions that are satisfied in most of the natural settings. To see this, consider an order  $\varepsilon_n$ -neighborhood of the level set  $L_\lambda$ . If the relevant indexing class of sets on this neighborhood is such that there *is* weak convergence of the local empirical process, then by looking at one or finitely many sets  $A$  the rate is determined by the classical central limit theorem for the binomial distribution. Therefore, if  $\mu(A \triangle L_\lambda)$  and  $P(A \triangle L_\lambda)$  are of exact order  $\varepsilon_n$  the rate of  $P_n(A) - P(A) - P_n(L_\lambda) + P(L_\lambda)$  is  $\sqrt{\varepsilon_n/n}$ ; it cannot be something else. Next we exhibit natural smoothness assumptions, see “local excess mass” in particular (2.16) below, under which the local empirical process has a “drift” of order  $\varepsilon_n^2$ . Balancing these two components by equating the rate  $\sqrt{\varepsilon_n/n}$  and the drift  $\varepsilon_n^2$  yields the cube root rate  $\varepsilon_n = n^{-1/3}$ ; details can be found in the proof of Lemma 3.1. This rate for vectors instead of sets can be found in, e.g., Kim and Pollard (1990). For Vapnik-Chervonenkis classes of sets it is indicated in Polonik (1995) although in there an additional  $(\log n)^{1/3}$  term is still present and hence no convergence can be considered.

As just indicated, we use an appropriate boundary empirical process, see Khmaladze (2007), Khmaladze and Weil (2008), and Einmahl and Khmaladze

(2011) and study its weak convergence on a “cylinder space” associated with the boundary  $S_\lambda$  of  $L_\lambda$ . Interestingly, the local nature of the convergence makes both the rate dimension-free, and the limiting Wiener process distribution-free (it depends on  $\mathcal{A}$  and  $L_\lambda$  only). Note that the convexity of  $L_\lambda$  and the elements of  $\mathcal{A}$  is not relevant for the rate. Instead, geometrical conditions on the class  $\mathcal{A}$  have to be worked out to achieve the concentration rate  $n^{-1/3}$  and determine the drift process depending on  $P$ .

**Organization.** Section 2 is devoted to the setup of the paper, including the relevant definitions, notation, and assumptions. In Section 3 we present and discuss the main results and provide a few explicit, illuminating examples. The proofs are deferred to Section 4.

## 2. Setup, notation and assumptions

### 2.1. The geometrical framework

In order to define the appropriate limit setting the following notation and definitions are needed.

**The magnification map  $\tau_\varepsilon$ .** Let  $\|x\|$  denote the Euclidean norm of  $x \in \mathbb{R}^d$  and  $U = \{u : \|u\| = 1\}$  the unit sphere. Since  $L_\lambda$  is a convex body, the metric projection  $\Pi(x) \in S_\lambda$  of  $x \in \mathbb{R}^d$  on  $S_\lambda = \partial L_\lambda$  is unique except for so-called skeleton points  $x \in L_\lambda^* \subset L_\lambda$  with  $\mu(L_\lambda^*) = 0$ . A unit vector  $u \in U$  is called an outer normal of  $L_\lambda$  at  $\pi \in S_\lambda$  if there is some  $x \in \mathbb{R}^d \setminus L_\lambda$  such that  $\pi = \Pi(x)$  and  $u = (x - \Pi(x))/\|x - \Pi(x)\|$ . At each  $\pi \in S_\lambda$ , we denote the non-empty set of outer normals by  $N(\pi)$  and write  $S_\lambda^* = \{\pi \in S_\lambda : \text{card}(N(\pi)) > 1\}$ . Note that  $\mu(S_\lambda) = 0$  and hence  $\mu(S_\lambda^*) = 0$ . The normal bundle of  $L_\lambda$  is

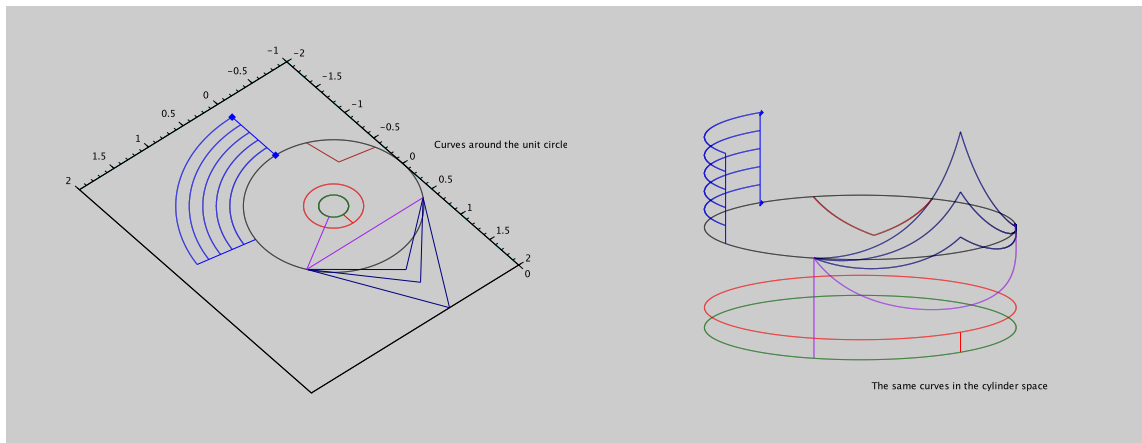
$$\text{Nor}(L_\lambda) = \{(\pi, u) : \pi \in S_\lambda, u \in N(\pi)\}.$$

As in Khmaladze (2007) and Einmahl and Khmaladze (2011) define the magnification map  $\tau_\varepsilon$ , with  $\varepsilon > 0$ , to be

$$\tau_\varepsilon(x) = \left( \Pi(x), u(x), \frac{s(x)}{\varepsilon} \right) \in \text{Nor}(L_\lambda) \times \mathbb{R}, \quad \text{for } x \in \mathbb{R}^d \setminus (L_\lambda^* \cup S_\lambda^*), \quad (2.1)$$

where  $x = \Pi(x) + s(x)u(x)$ , with  $s(x) = \text{sgn}(x)\|x - \Pi(x)\|$  the signed distance between  $x$  and  $\Pi(x)$ ; here  $\text{sgn}(x) = -1$  if  $x \in L_\lambda$  and 1 otherwise.

**Example.** Here we illustrate the magnification map  $\tau_\varepsilon$  in case  $u$  is unique on  $S_\lambda$  and hence can be omitted. The two figures concern the case  $d = 2$  and  $\varepsilon = 1$  with the unit disk  $L_\lambda$ .



The curves around the unit circle  $S = S_\lambda$  in the left figure are seen (in the same color) in the right figure on the cylinder  $S \times \mathbb{R}$  by assimilating  $(\Pi(x), u(x)) \in \text{Nor}(L_\lambda)$  to  $\Pi(x) \in S$  and moving away  $s(x)/\varepsilon$  from this point in vertical direction.

**The cylinder space.** Define  $\Sigma = \text{Nor}(L_\lambda) \times \mathbb{R}$ . Let  $\nu_{d-1}$  denote both the Hausdorff surface measure on  $S_\lambda$  (putting no mass at  $S_\lambda^*$ ) and its canonical extension to  $\text{Nor}(L_\lambda)$  supported by the product Borel  $\sigma$ -algebra  $\mathcal{G}_{d-1}$  on  $S_\lambda \times U$ . Thus,  $\nu_{d-1}$  on  $\text{Nor}(L_\lambda)$  is the so-called first support measure, and we have  $0 < s_\lambda = \nu_{d-1}(S_\lambda) = \nu_{d-1}(\text{Nor}(L_\lambda)) < \infty$ . Let  $\mu_1$  be Lebesgue measure on  $\mathbb{R}$ . The cylinder space  $(\Sigma, \mathcal{F}, M, d)$  is defined to be  $\Sigma$  endowed with the product Borel  $\sigma$ -algebra  $\mathcal{F} = \mathcal{G}_{d-1} \times \mathcal{B}(\mathbb{R})$ , the  $\sigma$ -finite product measure  $M$  and the semi-metric  $d$  given by

$$M = \nu_{d-1} \times \mu_1, \quad d(B, B') = (M(B \Delta B'))^{1/2}, \quad \text{for } B, B' \in \mathcal{F}. \quad (2.2)$$

Note that  $M$  plays an important role in convex analysis, see [Schneider \(1993\)](#) and in particular (4.2) below. Near  $S_\lambda$ , the measure  $\mu/\varepsilon$  can be approximated by  $M$ , after the transformation  $\tau_\varepsilon$ . This is made precise in (4.10). For  $c > 0$  denote  $\Sigma_c = \text{Nor}(L_\lambda) \times [-c, c]$  and  $\mathcal{F}_c = \{B \in \mathcal{F} : B \subset \Sigma_c\}$ .

**The sufficiently parallel sets  $\mathcal{A}^\varepsilon$ .** Given  $\varepsilon > 0$  the  $\varepsilon$ -parallel set of  $S_\lambda$  is defined by  $S_\lambda^\varepsilon = \{x : \|x - \Pi(x)\| \leq \varepsilon\}$  and we consider the sets in  $\mathcal{A}$  that are “sufficiently parallel” to  $L_\lambda$ ,

$$\mathcal{A}^\varepsilon = \{A \in \mathcal{A} : A \Delta L_\lambda \subset S_\lambda^\varepsilon\}, \quad \mathcal{C}^\varepsilon = \{A \Delta L_\lambda : A \in \mathcal{A}^\varepsilon\}. \quad (2.3)$$

Define the set-to-set mapping

$$\tau_\varepsilon(C) = \{\tau_\varepsilon(x) : x \in C \setminus (L_\lambda^* \cup S_\lambda^*)\}, \quad C \in \mathcal{B}(\mathbb{R}^d),$$

and the inverse  $\tau_\varepsilon^{-1}(B) = \{x \in \mathbb{R}^d : \tau_\varepsilon(x) \in B\}$ , for  $B \in \mathcal{F}$ . Note that  $\tau_\varepsilon^{-1}(\tau_\varepsilon(C)) = C \setminus (L_\lambda^* \cup S_\lambda^*)$ . For  $B \in \mathcal{F}$ , define  $\varphi_\varepsilon(B)$  to be the closure of  $\tau_\varepsilon^{-1}(B) \Delta (L_\lambda \setminus (L_\lambda^* \cup S_\lambda^*))$ . For  $A \in \mathcal{A}$ , we then have  $\varphi_\varepsilon(\tau_\varepsilon(A \Delta L_\lambda)) = A$ .

**The limiting class  $\mathcal{B}$ .** As explained in Subsection 1.2, we need to magnify with  $1/\varepsilon = n^{1/3}$ . Define for  $c > 0$

$$\mathcal{B}_{c,n} = \tau_{n^{-1/3}}(\mathcal{C}^{cn^{-1/3}}) = \{\tau_{n^{-1/3}}(A \triangle L_\lambda) : A \in \mathcal{A}^{cn^{-1/3}}\} \quad (2.4)$$

and  $\mathcal{B} = \bigcup_{c>0} \mathcal{B}_c$  where

$$\mathcal{B}_c = \left\{ B \in \mathcal{F}_c : \text{for some } B_n \in \mathcal{B}_{c,n}, \lim_{n \rightarrow \infty} d(B, B_n) = 0 \right\}. \quad (2.5)$$

Since  $L_\lambda \in \mathcal{A}^{cn^{-1/3}}$  we have  $\mathcal{B} \neq \emptyset$ . In the language of Khmaladze (2007) each  $B \in \mathcal{B}_c$  is a derivative at 0 of the set-valued function  $\varepsilon \mapsto \tau_\varepsilon(\mathcal{A}^{c\varepsilon})$  along the sequence  $\varepsilon = n^{-1/3}$ . Such limits are not uniquely determined. Actually the limit “set”  $B$  is an equivalence class of sets having  $d$ -distance equal to 0. Out of every equivalence class, we choose (only) one limit set  $B \in \mathcal{F}_c$ . This makes  $d$  a metric on  $\mathcal{B}_c$  and on  $\mathcal{B}$ . (The choices of the limit set matter. In applications we choose  $B$ ’s such that the assumptions of our theorems are satisfied.)

## 2.2. Condition $H_1$

In order to exploit the weak convergence of the local empirical process in a  $cn^{-1/3}$ -neighborhood of  $S_\lambda$ ,  $c > 0$ , at standard rate  $n^{-2/3}$  we borrow assumptions (2.6)–(2.11) from Einmahl and Khmaladze (2011). Let us first assume that, for any  $c > 0$ ,  $(\mathcal{B}_c, d)$  is compact and

$$\lim_{n \rightarrow \infty} \sup_{B_n \in \mathcal{B}_{c,n}} \inf_{B \in \mathcal{B}_c} d(B_n, B) = 0. \quad (2.6)$$

**Donsker classes.** We require standard conditions on the size of the class  $\mathcal{A}$ . Let  $c > 0$ . Define  $d_n(C, C') = (n^{1/3}P(C \triangle C'))^{1/2}$ . Assume for every  $\kappa > 0$  there exists a finite collection of brackets  $[\underline{C}(\kappa), \overline{C}(\kappa)]$  of Borel sets in  $S_\lambda^{cn^{-1/3}}$  with  $d_n(\underline{C}(\kappa), \overline{C}(\kappa)) \leq \kappa$  such that for every set  $C \in \mathcal{C}^{cn^{-1/3}}$  there is a bracket such that  $\underline{C}(\kappa) \subset C \subset \overline{C}(\kappa)$ . Denote the cardinality (the bracketing number) of such a class of brackets with minimal cardinality by  $[\mathcal{A}]_{c,n}(\kappa)$ . We assume the same for  $\{\tau_{n^{-1/3}}^{-1}(B) : B \in \mathcal{B}_c\}$  (instead of  $\mathcal{C}^{cn^{-1/3}}$ ) and denote the corresponding bracketing number with  $[\mathcal{B}]_{c,n}(\kappa)$ . We require either that

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \int_0^\delta \sqrt{\log [\mathcal{A}]_{c,n}(\kappa)} d\kappa = 0, \quad \text{for any } c > 0, \quad (2.7)$$

$$\limsup_{n \rightarrow \infty} \sup_{1 \leq c \leq \delta n^{1/3}} \int_0^1 \sqrt{\log [\mathcal{A}]_{c,n}(c^{1/2}\kappa)} d\kappa < \infty, \quad \text{for every small } \delta \succ \mathfrak{A}\delta$$

$$n^{1/2} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| = O_{\mathbb{P}}(1), \quad n \rightarrow \infty, \quad (2.9)$$

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \int_0^\delta \sqrt{\log [\mathcal{B}]_{c,n}(\kappa)} d\kappa = 0, \quad \text{for any } c > 0, \quad (2.10)$$



or that

$$\mathcal{A} \text{ and } \mathcal{B} \text{ are Vapnik-Chervonenkis (VC) classes.} \quad (2.11)$$

We also assume that  $\mathcal{A}$  and  $\mathcal{B}$  are pointwise measurable.

As discussed in more detail below the main results, these conditions are satisfied for the classes of ellipsoids, convex polytopes with a bound on the number of faces, convex sets in dimension 2 (in case  $L_\lambda$  is a square) and many other often-used parametric classes of sets. When the class is too massive like for instance the class of convex sets in dimension 3 or higher, they do not hold. (For this non-VC class example, condition (2.7) is not satisfied.)

**Nested class.** Assume that for all  $r > 0$ , all  $A \in \mathcal{A}$  there exists  $A_r \in \mathcal{A}$  such that

$$A \subset A_r, \quad \mu(A_r) = \mu(A) + r. \quad (2.12)$$

Let us denote by  $H_1$  the conditions in Subsection 2.2, that is compactness of  $(\mathcal{B}_c, d)$  and (2.6), ‘‘Donsker class’’ (with either conditions (2.7)–(2.10) or (2.11)), and (2.12).

### 2.3. Condition $H_2$

Given the weak convergence of local empirical processes, two difficulties remain for the convergence of the set-valued estimators. First, we need to establish the dimension-free concentration rate in  $\mathbb{R}^d$ . Second, we need to determine the ‘‘statistical’’ drift that competes with the convergence rate of the local empirical processes. For this the behavior of  $f$  near  $S_\lambda$  has to be specified. The formulation and in-depth analysis of this drift in relation with the local empirical processes convergence is one of the main challenges of this paper.

Define the Hausdorff distance for the Euclidean norm on  $\mathbb{R}^d$  as

$$d_H(A, A') = \max \left( \sup_{x \in A} \inf_{x' \in A'} \|x - x'\|, \sup_{x' \in A'} \inf_{x \in A} \|x - x'\| \right), \quad \text{for } A, A' \subset \mathbb{R}^d.$$

**Excess mass.** Consider the excess mass

$$e_\lambda - e_\lambda(A) = P(L_\lambda) - P(A) - \lambda(\mu(L_\lambda) - \mu(A)) = \int_{L_\lambda \Delta A} |f(x) - \lambda| d\mu(x).$$

We require that for all  $\delta > 0$ ,

$$\inf_{A \in \mathcal{A}: d_H(L_\lambda, A) \geq \delta} \int_{L_\lambda \Delta A} |f(x) - \lambda| d\mu(x) > 0. \quad (2.13)$$

**The quadratic drift measure  $D$ .** We assume that for some second-order derivatives  $f'_+ \geq 0$  and  $f'_- \geq 0$  defined on  $S_\lambda$  we have

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon^2} \int_{S_\lambda^\varepsilon \setminus L_\lambda} |f(x) - \lambda + s(x)f'_+(\Pi(x))| d\mu(x) = 0, \quad (2.14)$$

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon^2} \int_{S_\lambda^\varepsilon \cap L_\lambda} |f(x) - \lambda + s(x)f'_-(\Pi(x))| d\mu(x) = 0. \quad (2.15)$$

If  $f$  is differentiable at  $\pi \in S_\lambda \setminus S_\lambda^*$  then  $f'_+(\pi) = f'_-(\pi)$ . Let us define on  $(\Sigma, M)$  the quadratic drift measure  $D$  having density with respect to  $M$  given by

$$\frac{dD}{dM}(\pi, u, s) = sf'_+(\pi)1_{s>0} - sf'_-(\pi)1_{s\leq 0}.$$

**Local excess mass.** Write  $g(\pi) = \min(f'_+(\pi), f'_-(\pi))$  for  $\pi \in S_\lambda$ . Let assume that for some  $\varepsilon_0 > 0$ ,  $\eta_0 > 0$  and all  $A \in \mathcal{A}$  such that  $d_H(L_\lambda, A) \leq \varepsilon_0$ , we have

$$\int_{L_\lambda \triangle A} |s(x)|g(\Pi(x))d\mu(x) \geq \eta_0 d_H^2(L_\lambda, A). \quad (2.16)$$

Similarly we require that for all  $B \in \mathcal{B}$

$$D(B) \geq \eta_0 c^2(B), \quad (2.17)$$

where  $c(B) = \inf\{c > 0 : B \in \mathcal{B}_c\}$ .

Let  $H_2$  denote the conditions (2.13)–(2.17) above. Conditions (2.13) and (2.16) do not allow that sets that are not close to  $L_\lambda$  in Hausdorff distance have a small excess mass, and (2.17) is a version of (2.16) in the limiting setting. These conditions exclude classes of sets that allow spiky deviations from  $L_\lambda$ . Conditions (2.14) and (2.15) are needed to express the drift in the inward and outward derivatives of  $f$  along the boundary  $S_\lambda$ .

To see that  $H_2$  is rather weak, consider the second order differentiability assumptions H1-H2 imposed on  $f$  in Cadre (2006) to control  $\mu(L_\lambda \triangle \widehat{L})$  where  $\widehat{L} = \{x : \widehat{f}(x) > \lambda\}$  is the plug-in level set derived from a kernel estimator  $\widehat{f}$ . Clearly, H1 implies (2.14) and (2.15) whereas H2 implies that  $g = f'_+ = f'_-$  is bounded away from 0 on  $S_\lambda$ . Thus (2.13) and (2.16) hold if  $\inf_{A \in \mathcal{A} \setminus \{L_\lambda\}} \mu(L_\lambda \triangle A)/d_H(L_\lambda, A) > 0$ . Note that the Hausdorff and (here second order) Nikodym distances connected by (2.16) are known to play a different, sometimes opposite, role when approximating a convex support, see Brunel (2018) for smooth versus polyhedral supports.

### 3. Main results

#### 3.1. Convergence of the excess mass set estimator

Since  $(\Sigma, M)$  is  $\sigma$ -finite and  $(\mathcal{B}, d)$  is  $\sigma$ -compact we can define a Wiener process  $W$  indexed by  $\mathcal{B}$ , that is a centered Gaussian process with covariance

$$\text{Cov}(W(B), W(B')) = M(B \cap B'), \quad \text{for } B, B' \in \mathcal{B}.$$

The intrinsic, standard deviation metric of  $W$  on  $\mathcal{B}$  is defined to be  $(\text{Var}(W(B) - W(B')))^{1/2} = d(B, B')$ . The relevant limiting random set is

$$Z(\mathcal{B}) = \arg \max_{B \in \mathcal{B}} \left\{ \sqrt{\lambda} W(B) - D(B) \right\}. \quad (3.1)$$

This quantity has been studied in the univariate case where the sets reduce to numbers, see Groeneboom (1985), Dykstra and Carolan (1999), and Berthet and El-Nouty (2006). Observe that  $\mathbb{E}W^2(B) \leq 2s_\lambda c$  for  $B \in \mathcal{B}_c$ . We assume that for some  $\eta_1 > 0$ ,

$$\mathbb{E} \sup_{B \in \mathcal{B}_c} W(B) < \eta_1 \sqrt{c}, \quad \text{for all } c > 0. \quad (3.2)$$

For many classes  $\mathcal{A}$ , including those mentioned in  $H_1$ , it holds that  $\mathcal{B}_c = \{cB : B \in \mathcal{B}_1\}$ , where  $cB = \{(\pi, u, cs) : (\pi, u, s) \in B\}$ . This readily yields the “scaling” with  $c$  in (3.2).

**Proposition 3.1.** *Assume that  $H_1$ ,  $H_2$ , and (3.2) hold. With probability one, the random set  $Z(\mathcal{B})$  of (3.1) exists and is unique.*

We are now ready to state our non-standard weak convergence result for the sequence of random sets  $L_{1,n}$  in (1.1).

**Theorem 3.1.** *Assume that  $H_1$ ,  $H_2$ , and (3.2) hold. Then on some probability space there exists a triangular array  $X_{n,1}, \dots, X_{n,n}$ ,  $n \in \mathbb{N}$ , of rowwise independent random vectors with law  $P$  on  $\mathbb{R}^d$  together with a sequence  $Z_n(\mathcal{B})$  of versions of  $Z(\mathcal{B})$  such that for every argmax  $L_{1,n}$  of (1.1), as  $n \rightarrow \infty$ ,*

$$\begin{aligned} M(\tau_{n^{-1/3}}(L_{1,n} \Delta L_\lambda) \Delta Z_n(\mathcal{B})) &\xrightarrow{\mathbb{P}} 0, \\ n^{1/3} \mu(L_{1,n} \Delta \varphi_{n^{-1/3}}(Z_n(\mathcal{B}))) &\xrightarrow{\mathbb{P}} 0, \\ n^{1/3} P(L_{1,n} \Delta \varphi_{n^{-1/3}}(Z_n(\mathcal{B}))) &\xrightarrow{\mathbb{P}} 0. \end{aligned}$$

To connect the first statement to the other two, recall that by the definition of  $\varphi_{n^{-1/3}}$  we have  $L_{1,n} = \varphi_{n^{-1/3}}(\tau_{n^{-1/3}}(L_{1,n} \Delta L_\lambda))$ . Theorem 3.1 states that, at the scale  $n^{-1/3}$ , the symmetric difference between the empirical excess mass set and  $L_\lambda$  has as a limiting distribution that of the argmax of a drifted Wiener process, as defined in (3.1).

In Theorem 3.1  $f$ ,  $L_\lambda$  and  $\mathcal{A}$  are fixed. Even if  $f$  is regular enough to be estimated by the usual nonparametric density estimators  $f_n$  the shape of the plug-in estimator  $L_\lambda(f_n)$  of the level set is not easy to describe geometrically and strongly depends on the smoothing parameter (instead of on  $\mathcal{A}$ ). Compared to the more classical CLT for the centered measure  $\mu(L_\lambda(f_n) \Delta L_\lambda) - \mathbb{E}(\mu(L_\lambda(f_n) \Delta L_\lambda))$  (see Mason and Polonik (2009) for  $f_n$  being the kernel density estimator) we obtain the limiting distribution of the set  $L_{1,n} \Delta L_\lambda$  itself. Moreover, smooth transforms of  $L_{1,n} \Delta L_\lambda$  can be controlled by those of  $L_\lambda \Delta \varphi_{n^{-1/3}}(Z(\mathcal{B}))$ , which provides a new access to risks such as  $\mathbb{E}(\mu(L_{1,n} \Delta L_\lambda))$ .

To control uniformly a risk of  $L_{1,n}$ , Theorem 3.1 should be extended to a class of densities instead of a single density  $f$ . Now, in order to keep the  $n^{-1/3}$  rate,  $\mathcal{A}$  should contain *all* the associated level sets, and also the entropy conditions in  $H_1$  as well as the conditions in  $H_2$  have to hold for each density in the class. In Tsybakov (1997) minimax rates for the expected volume of the symmetric difference of level sets estimators are obtained which are sometimes (for non-VC

classes) slower than the  $n^{-1/3}$  rate obtained here (especially in high dimensions). The slower rates are due to the more general setup in that paper in comparison with our conditions in  $H_1$  and  $H_2$ , for the large  $\mathcal{A}$ , that includes all the level sets of the class of densities, too consider. In particular,  $L_{1,n}$  is shown to be minimax for radially regular densities with convex level sets containing a small ball at the origin. Notice that if  $f$  is differentiable on  $S_\lambda$  then (2.14)–(2.15) hold and if the derivative is bounded away from zero then  $f$  is radially 1-regular in Tsybakov (1997), thus the main topic when considering  $L_{1,n}$  is studying the geometrical features of the sets in  $\mathcal{A}$  with regard to each allowed  $L_\lambda$  in view of the intended limiting result. Clearly for the present results, some sequences of sets may be too spiky (hence  $H_2$  does not hold), or have too large bracketing entropy numbers ( $H_1$  does not hold).

For similar reasons, the rates obtained for  $P(L_{j,n}\Delta L_\lambda)$  for non-VC classes in Polonik (1995), for  $j = 1$ , and in Polonik (1997), for  $j = 2$ , are also slower than  $n^{-1/3}$ .

### 3.2. Convergence of the minimum volume set and the maximum probability set estimators

For the second main result about  $L_{2,n}$  and  $L_{3,n}$  we need some more notation and assumptions.

**The limiting class  $\mathcal{B}^*$ .** Write  $B^+ = B \cap (Nor(L_\lambda) \times \mathbb{R}^+)$  and  $B_- = B \setminus B^+$ , for  $B \in \mathcal{F}$ . Now define

$$\mathcal{B}^* = \{B \in \mathcal{B} : M(B^+) = M(B_-)\}, \quad \mathcal{B}_c^* = \mathcal{B}^* \cap \mathcal{B}_c.$$

Note that  $(\mathcal{B}_c^*, d)$  is also compact. By replacing  $\mathcal{A}$  in (2.3)–(2.4) with

$$\mathcal{A}_v = \{A \in \mathcal{A} : \mu(A) = v_\lambda\}, \quad \mathcal{A}_p = \{A \in \mathcal{A} : P(A) = p_\lambda\}, \quad (3.3)$$

respectively, we define in the same way the classes  $\mathcal{A}_v^\varepsilon, \mathcal{C}_v^\varepsilon, \mathcal{A}_p^\varepsilon, \mathcal{C}_p^\varepsilon$  and  $\mathcal{B}_{c,n}^v, \mathcal{B}_{c,n}^p$ .

For the excess mass estimator  $L_{1,n}$  the sets  $A \in \mathcal{A}$  are not restricted in terms of  $\mu$  or  $P$ . For the maximum probability estimator  $L_{3,n}$  there is a restriction in terms of  $\mu$  and for the minimum volume estimator  $L_{2,n}$  there is a restriction in terms of  $P_n$ . It turns out that the sets considered in the definition of  $L_{3,n}$  can be further restricted to being in  $\mathcal{A}_v$ . Somewhat similar, but more complicated, it will be shown that the sets considered in the definition of  $L_{2,n}$  can be replaced by those being very close to members of  $\mathcal{A}_p$ . Since  $P \approx \lambda\mu$  near  $S_\lambda$ , both  $\mathcal{A}_v$  and  $\mathcal{A}_p$  lead to the same limiting class  $\mathcal{B}^*$  under the conditions (3.4) and (3.5) below.

*Condition  $H_3$ .* We assume

$$\lim_{n \rightarrow \infty} \sup_{B \in \mathcal{B}_c^*} \inf_{B_n \in \mathcal{B}_{c,n}^v} d(B, B_n) = 0, \quad (3.4)$$

$$\lim_{n \rightarrow \infty} \sup_{B \in \mathcal{B}_c^*} \inf_{B_n \in \mathcal{B}_{c,n}^p} d(B, B_n) = 0. \quad (3.5)$$

Note that (3.4) or (3.5) typically do not hold if, e.g., for all sets  $A \in \mathcal{A} \setminus \{L_\lambda\}$ ,  $\mu(A) > v_\lambda$  or  $P(A) > p_\lambda$ , respectively. In those cases  $\mathcal{B}_c^*$  can be large, whereas  $\mathcal{B}_{c,n}^v$  or  $\mathcal{B}_{c,n}^p$  contain only the one set corresponding to  $L_\lambda$ .

Consider the Wiener process  $W$  indexed by  $\mathcal{B}^*$  and define

$$Z(\mathcal{B}^*) = \arg \max_{B \in \mathcal{B}^*} \left\{ \sqrt{\lambda} W(B) - D(B) \right\}.$$

As in Proposition 3.1, under  $H_1$ ,  $H_2$  and (3.2), with probability one  $Z(\mathcal{B}^*)$  exists and is unique.

In order to control the minimum volume set estimator we need the following two conditions. The class  $\mathcal{A}$  contains a “univariate” subclass

$$\mathcal{A}_l = \{A_s \in \mathcal{A} : s \in (-p_\lambda, 1 - p_\lambda), P(A_s) = p_\lambda + s\} \quad (3.6)$$

of nested sets with the properties that  $A_s \subset A_{s'}$  if  $s < s'$ ,  $A_0 = L_\lambda$ , and for some  $s_0 > 0, \zeta > 0$  and for all  $-s_0 \leq s \leq s_0$ :

$$e_\lambda - e_\lambda(A_s) \leq \zeta s^2. \quad (3.7)$$

For every  $c > 0$ , we have as  $n \rightarrow \infty$ ,

$$\sup_{A \in \mathcal{A}^{cn^{-1/3}}, P(A)=p_\lambda} \inf_{\tilde{A} \in \mathcal{A}^{cn^{-1/3}}, P_n(\tilde{A})=\lceil np_\lambda \rceil/n} d(\tau_\varepsilon(A), \tau_\varepsilon(\tilde{A})) \xrightarrow{\mathbb{P}} 0. \quad (3.8)$$

Let  $H_3$  denote the conditions (3.4)–(3.8) above. These conditions, although technical, are rather mild and they are satisfied by many natural classes  $\mathcal{A}$  like the ones mentioned in  $H_1$ . This mainly follows from the fact that a given ellipsoid, say, can be continuously inflated or deflated, even when e.g. its major diameter has to stay fixed.

**Theorem 3.2.** *Assume that  $H_1$ ,  $H_2$ ,  $H_3$  and (3.2) hold. Then on some probability space there exists a triangular array  $X_{n,1}, \dots, X_{n,n}$ ,  $n \in \mathbb{N}$ , of rowwise independent random vectors with law  $P$  on  $\mathbb{R}^d$  together with a sequence  $Z_n(\mathcal{B}^*)$  of versions of  $Z(\mathcal{B}^*)$  such that every  $\text{argmin}$   $L_{2,n}$  of (1.2) and every  $\text{argmax}$   $L_{3,n}$  of (1.3) satisfy, for  $j = 2, 3$ , as  $n \rightarrow \infty$ ,*

$$\begin{aligned} M(\tau_{n^{-1/3}}(L_{j,n} \triangle L_\lambda) \triangle Z_n(\mathcal{B}^*)) &\xrightarrow{\mathbb{P}} 0, \\ n^{1/3} \mu(L_{j,n} \triangle \varphi_{n^{-1/3}}(Z_n(\mathcal{B}^*))) &\xrightarrow{\mathbb{P}} 0, \\ n^{1/3} P(L_{j,n} \triangle \varphi_{n^{-1/3}}(Z_n(\mathcal{B}^*))) &\xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Note that by definition the distributions of  $Z_n(\mathcal{B})$  and  $Z_n(\mathcal{B}^*)$  do not depend on  $n$  and they are in general not degenerate (cf. the univariate case). When going back to  $\mathbb{R}^d$ , we obtain  $\varphi_{n^{-1/3}}(Z_n(\mathcal{B}))$  and  $\varphi_{n^{-1/3}}(Z_n(\mathcal{B}^*))$  which are the asymptotic approximations of  $L_{1,n}$  and  $L_{2,n}$ ,  $L_{3,n}$  respectively.

Comparing Theorems 3.1 and 3.2 we see that the limiting behavior of  $L_{2,n}$  and  $L_{3,n}$  is substantially “less rich” than that of  $L_{1,n}$ . The symmetry of the

sets in  $\mathcal{B}^*$  shows that for  $j = 2, 3$  the inner and outer differences  $L_{j,n} \setminus L_\lambda$  and  $L_\lambda \setminus L_{j,n}$  tend to compensate. Theorems 3.1 and 3.2 could be stated jointly since they can indeed be proved with the same sequence of underlying Wiener processes  $W_n$ .

From a statistical point of view it would be interesting to investigate which estimator performs better:  $L_{1,n}$  on the one hand or  $L_{2,n}$  and  $L_{3,n}$  on the other hand. Theorems 3.1 and 3.2 provide the asymptotic theory for such a comparison. Since all three estimators converge at the same rate, to answer such a question we would need to study  $Z(\mathcal{B})$  and  $Z(\mathcal{B}^*)$  in order to see which argmax is “closer” to, say,  $\Sigma_0$  (corresponding to  $L_\lambda$ ). For the univariate case, simulations indicate that  $L_{2,n}$  and  $L_{3,n}$  asymptotically outperform  $L_{1,n}$ . However, a thorough comparison in particular in the multivariate case, even when considering a simple indexing class like the class of all closed ellipsoids and a simple density like some multivariate normal, is theoretically and computationally challenging and beyond the scope of the present paper.

In Di Bucchianico, Einmahl and Mushkudiani (2001) guaranteed coverage tolerance regions and mean coverage tolerance (or prediction) regions are derived with the aid of minimum volume estimators. Although strictly speaking the guaranteed coverage tolerance regions are not contained in the setup of the present paper since  $p_\lambda$  in (1.2) depends on  $n$  (but it converges), the mean coverage tolerance regions do fit the present setup. It is shown therein that (in our notation)  $\mathbb{E}P(L_{2,n}) = p_\lambda + o(n^{-1/2})$ , whereas Theorem 3.2 now yields the slower  $n^{-1/3}$  convergence rate as well as the fine limiting behavior of the mean coverage tolerance regions themselves (compared with the oracle tolerance region  $L_\lambda$ ).

Theorems 3.1 and 3.2 suggest new tools to derive confidence regions for  $L_\lambda$  in case the class of densities considered is not too large. First the relevant unknown quantities, in particular  $D$ , used in the limiting set  $Z(\mathcal{B})$  or  $Z(\mathcal{B}^*)$  should be estimated and then in principle this approximated limiting set, can be simulated many times. Now determine  $c_\alpha$  such that the proportion of these simulated sets that has, e.g., the last coordinate on the approximated cylinder included in  $[-c_\alpha, c_\alpha]$ . Then inverting this part of the cylinder to  $\mathbb{R}^d$  yields a  $(1 - \alpha)$ -confidence band for  $L_\lambda$ . Alternatively we could try to take into account the behavior of  $f'_-$  and  $f'_+$  when taking the proportion  $1 - \alpha$  of the simulated sets, which would naturally lead to bands which are narrower where  $f'_-$  or  $f'_+$  is larger.

Many further applications of level sets estimation can be found in Mammen and Polonik (2013) and Qiao and Polonik (2019) and the references therein.

From the proof of Theorem 3.2 it follows that the sequence of versions  $Z_n(\mathcal{B}^*)$  can be chosen the same for  $L_{2,n}$  and  $L_{3,n}$ . Hence, we obtain, as stated in the next result, that  $L_{2,n}$  and  $L_{3,n}$  are asymptotically equivalent.

**Corollary 3.1.** *Under the assumption of Theorem 3.2, as  $n \rightarrow \infty$ ,*

$$\begin{aligned} M(\tau_{n^{-1/3}}(L_{2,n} \triangle L_{3,n})) &\xrightarrow{\mathbb{P}} 0, \\ n^{1/3} \mu(L_{2,n} \triangle L_{3,n}) &\xrightarrow{\mathbb{P}} 0, \\ n^{1/3} P(L_{2,n} \triangle L_{3,n}) &\xrightarrow{\mathbb{P}} 0. \end{aligned}$$

### 3.3. The main concentration lemma

After all the conditions for its formulation have been introduced, we would like to highlight in this “Main results” section the following lemma which is the key to the convergence rate  $n^{-1/3}$ .

Consider the following variant of  $L_{3,n}$ :

$$L_{4,n} \in \arg \max_{A \in \mathcal{A}} \{P_n(A) : \mu(A) = v_\lambda\}. \quad (3.9)$$

**Lemma 3.1.** *Under the assumptions of Theorems 3.1 or 3.2, respectively, for every  $\delta > 0$ , there exists a  $c > 1$ , and an  $n_0$ , such that, for  $j = 1, 2, 4$ , and  $n \geq n_0$ ,*

$$\mathbb{P}(d_H(L_{j,n}, L_\lambda) \geq cn^{-1/3}) \leq \delta.$$

### 3.4. Discussion and examples

This paper uses the probabilistic setting, with magnification map, cylinder space, and the weak convergence (in Lemma 4.6 below) provided by Einmahl and Khmaladze (2011). That paper is a convenient starting point for the statistical inference here. The goal of the present paper as well as our main results and various lemmas are novel and very different, however: Einmahl and Khmaladze (2011) is about uniform weak convergence, with arbitrary intermediate rate, of a local empirical process indexed by a class of sets, whereas the present paper is about set estimation, about convergence of the random sets themselves. Here a drift competing with the local empirical processes appears which leads to the cube root rate, lacking and irrelevant in Einmahl and Khmaladze (2011). Technically, the main novelties of our work are the statements and proofs of Theorems 3.1 and 3.2, including the assumptions in  $H_2$  and  $H_3$  required to control the drift, as well as the statement of the key Lemma 3.1 below.

The conditions on the class  $\mathcal{A}$  are such that natural classes, like in particular the class of all closed ellipsoids, are included. If the class is “small”, e.g., by allowing not all or only a few positive values for  $\mu(A)$  or for  $P(A)$  we can obtain pathological and/or degenerate behavior of the set-valued estimators. E.g., if  $\mathcal{A}$  contains  $L_\lambda$  and further only sets with  $\mu(A) > v_\lambda$ , then  $L_{3,n} = L_\lambda$ .

The assumptions in (2.14) and (2.15) consider the “most regular” behavior of the density  $f$  near  $S_\lambda$ . They lead to the cube root asymptotics in this paper. Faster or slower convergence rates are also possible, see, e.g., Polonik (1995). This would lead to  $W$  drifted by a non-quadratic measure on the cylinder space,

generalizing  $W$  drifted by a convex power function used in [Berthet and El-Nouty \(2006\)](#) to control the estimation of the shorth, the minimum volume convex set on the real line. It is the goal of the present paper, however, to reveal the asymptotic theory in the most regular setup, and not to present the most general results under the weakest assumptions.

We now present some specific examples of classes of sets and probability distributions where the three level set estimators can be used, see for more relevant details, including expressions for  $\mathcal{B}_c$  in some two-dimensional cases, Examples 1a, 2a and 2b in [Einmahl and Khmaladze \(2011\)](#).

**Ellipsoids.** The natural and most studied example is the case where  $\mathcal{A}$  is the class of all closed ellipsoids with non-empty interior and  $P$  is an elliptical probability distribution. A more restricted class is, of course, the class of all closed balls.

**Convex polytopes.** Another natural choice for  $\mathcal{A}$  is the class of all closed, convex polytopes with a bound on the number of faces. In particular in dimension 2, the class of all closed, convex quadrangles can be considered. In this case we could take a density  $f$  such that  $L_\lambda$  is a rectangle. An interesting difference with the previous example is that  $L_\lambda$  is non-smooth here, resulting, e.g., in a non-empty skeleton  $L_\lambda^*$ .

**Planar convex sets.** For dimension two, in case  $L_\lambda$  is a square (or rectangle), we can let  $\mathcal{A}$  be the large class of all closed, convex sets. Although this interesting case is also considered in Example 2b in [Einmahl and Khmaladze \(2011\)](#), we will provide at the end of Section 4 details about why conditions (2.7) and (2.8) are satisfied. Note that the bracketing numbers bound used there also indicates that our results do not hold true for the class of closed, convex sets in dimensions exceeding 2.

It might be difficult to determine  $L_{1,n}$ ,  $L_{2,n}$  and  $L_{3,n}$  and therefore some more flexibility in their definitions could be convenient. Consider for instance the following “relaxed” maximizers/minimizers: given any sequence  $\delta_n$  of positive numbers converging to 0, choose random sets  $R_{1,n}$ ,  $R_{2,n}$ , and  $R_{3,n}$  in  $\mathcal{A}$  such that  $P_n(R_{2,n}) \geq p_\lambda$ ,  $\mu(R_{3,n}) \leq v_\lambda$ , and

$$\begin{aligned} P_n(R_{1,n}) - \lambda\mu(R_{1,n}) &\geq \sup\{P_n(A) - \lambda\mu(A) : A \in \mathcal{A}\} - \delta_n n^{-2/3}, \\ \mu(R_{2,n}) &\leq \inf\{\mu(A) : A \in \mathcal{A}, P_n(A) \geq p_\lambda\} + \delta_n n^{-2/3}, \\ P_n(R_{3,n}) &\geq \sup\{P_n(A) : A \in \mathcal{A}, \mu(A) \leq v_\lambda\} - \delta_n n^{-2/3}. \end{aligned}$$

Our approach and convergence results naturally extend to  $R_{j,n}$ ,  $j = 1, 2, 3$ , but their detailed analysis is beyond the scope of this paper. Whenever  $\delta_n$  is chosen not too small (i.e.,  $\delta_n n^{1/3} \rightarrow \infty$ ) more flexible algorithms for the computation of  $R_{j,n}$  could be used.



#### 4. Proofs

We first collect various lemmas for the proof of the theorems. From now on we write  $\varepsilon_n = n^{-1/3}$ .

##### 4.1. Distances, measures and drift

For  $j = 1, \dots, d$ , let  $\nu_{d-j}(\cdot)$  denote the  $j$ -th support measure of  $L_\lambda$  on  $Nor(L_\lambda)$ . These finite measures carry the geometrical information about  $L_\lambda$  and are a common generalization of the curvature measures and the area measures, see [Schneider \(1993\)](#) and [Schneider and Weil \(2008\)](#). The local inner reach at  $\pi \in S_\lambda$  is the largest radius  $r(\pi)$  of a ball included in  $L_\lambda$  that has  $\pi$  as a boundary point. Theorem 1 in [Khmaladze and Weil \(2008\)](#) states a general Steiner formula for convex bodies: for any  $g \in L_1(\mu)$ ,

$$\int_{\mathbb{R}^d} g(x) d\mu(x) = \sum_{j=1}^d \binom{d-1}{j-1} \Theta_{d-j}(g), \text{ where} \quad (4.1)$$

$$\Theta_{d-j}(g) = \int_{Nor(L_\lambda)} \int_{-r(\pi)}^\infty s^{j-1} g(\pi + su) d\mu_1(s) d\nu_{d-j}(\pi, u). \quad (4.2)$$

It follows from (2.14)–(2.15) and this Steiner formula with  $g = f1_{S_\lambda^\varepsilon}$ , for small  $\varepsilon > 0$ , that

$$\int_{Nor(L_\lambda)} f'_\pm(\pi) \nu_{d-j}(\pi, u) < \infty, \text{ for } j = 1, \dots, d. \quad (4.3)$$

Define

$$\mathcal{B}_{c,n}^{p,+} = \{\tau_{\varepsilon_n}(A \triangle L_\lambda) : A \in \mathcal{A}^{c\varepsilon_n}, |P(A) - p_\lambda| \leq n^{-2/5}\}. \quad (4.4)$$

**Lemma 4.1.** *Let  $c > 0$ . We have, as  $n \rightarrow \infty$ ,*

$$\sup_{B_n \in \mathcal{B}_{c,n}^v} |M(B_n^+) - M(B_n^-)| = O(\varepsilon_n) \quad (4.5)$$

and, if (2.14)–(2.15) hold, then

$$\sup_{B_n \in \mathcal{B}_{c,n}^{p,+}} |M(B_n^+) - M(B_n^-)| \rightarrow 0. \quad (4.6)$$

*Proof.* For  $A_n \in \mathcal{A}^{c\varepsilon_n}$  and  $B_n = \tau_{\varepsilon_n}(A_n \triangle L_\lambda) \in \mathcal{B}_{c,n}$ , we have  $B_n^+ = \tau_{\varepsilon_n}(A_n \setminus L_\lambda)$  and  $B_n^- = \tau_{\varepsilon_n}(L_\lambda \setminus A_n)$ . Consider

$$g^+(x) = \varepsilon_n^{-1} 1_{A_n \setminus L_\lambda}(x) = \varepsilon_n^{-1} 1_{B_n^+}(\Pi(x), u(x), s(x)/\varepsilon_n)$$

in (4.1). Then

$$\Theta_{d-j}(g^+) = \varepsilon_n^{j-1} \int_{Nor(S_\lambda)} \int_0^c s^{j-1} 1_{B_n^+}(\pi, u, s) d\mu_1(s) d\nu_{d-j}(\pi, u).$$

Thus  $\Theta_{d-1}(g^+) = M(B_n^+)$ , by (2.2), and  $\Theta_{d-j}(g^+) = O(\varepsilon_n^{j-1})$  uniformly over  $\mathcal{A}^{c\varepsilon_n}$ , for  $j = 2, \dots, d$ . Since  $\varepsilon_n^{-1}\mu(A_n \setminus L_\lambda) = \int_{\mathbb{R}^d} g^+(x)d\mu(x)$  we see that (4.1) implies

$$\sup_{A \in \mathcal{A}^{c\varepsilon_n}} \left| \frac{1}{\varepsilon_n} \mu(A \setminus L_\lambda) - M(\tau_{\varepsilon_n}(A \setminus L_\lambda)) \right| = O(\varepsilon_n). \quad (4.7)$$

Similarly we obtain

$$\sup_{A \in \mathcal{A}^{c\varepsilon_n}} \left| \frac{1}{\varepsilon_n} \mu(L_\lambda \setminus A) - M(\tau_{\varepsilon_n}(L_\lambda \setminus A)) \right| = O(\varepsilon_n). \quad (4.8)$$

For  $A \in \mathcal{A}_v^{c\varepsilon_n}$ ,  $\mu(A) = v_\lambda$  and hence  $\mu(L_\lambda \setminus A) = \mu(A \setminus L_\lambda)$ . By (4.7) and (4.8), we obtain (4.5) by definition of  $\mathcal{B}_{c,n}^v$ .

Define

$$\mathcal{A}_p^{c\varepsilon_n,+} = \{A \in \mathcal{A}^{c\varepsilon_n} : |P(A) - p_\lambda| \leq n^{-2/5}\}.$$

For  $A_n \in \mathcal{A}_p^{c\varepsilon_n,+}$  we thus have

$$\left| \int_{A_n \setminus L_\lambda} f(x)d\mu(x) - \int_{L_\lambda \setminus A_n} f(x)d\mu(x) \right| \leq n^{-2/5},$$

and, by (2.14) and (2.15), uniformly over  $\mathcal{A}_p^{c\varepsilon_n,+}$ ,

$$\begin{aligned} & \lambda\mu(A_n \setminus L_\lambda) - \int_{A_n \setminus L_\lambda} s(x)f'_+(\Pi(x))d\mu(x) \\ &= \lambda\mu(L_\lambda \setminus A_n) - \int_{L_\lambda \setminus A_n} s(x)f'_-(\Pi(x))d\mu(x) + o(\varepsilon_n^2) + O(n^{-2/5}). \end{aligned} \quad (4.9)$$

Now consider  $\tilde{g}^+(x) = \varepsilon_n^{-1}1_{A_n \setminus L_\lambda}(x)s(x)f'_+(\Pi(x))$ . Then by (4.1)–(4.3) we find that uniformly over  $\mathcal{A}_p^{c\varepsilon_n,+}$ ,  $\varepsilon_n^{-1} \int_{A_n \setminus L_\lambda} s(x)f'_+(\Pi(x))d\mu(x) = O(\varepsilon_n)$ . We can deal similarly with the integral on  $L_\lambda \setminus A_n$ . Using this in (4.9) in combination with (4.7) and (4.8) yields (4.6).  $\square$

Observe that (4.7) and (4.8) immediately yield

$$\sup_{A \in \mathcal{A}^{c\varepsilon_n}} \left| \varepsilon_n^{-1} \mu(A \triangle L_\lambda) - M((\tau_{\varepsilon_n}(A \triangle L_\lambda))) \right| = O(\varepsilon_n). \quad (4.10)$$

**Lemma 4.2.** *If  $H_1$  and  $H_2$  hold, then*

$$\lim_{n \rightarrow \infty} \sup_{B_n \in \mathcal{B}_{c,n}^v} \inf_{B \in \mathcal{B}_c^*} d(B_n, B) = 0, \quad (4.11)$$

$$\lim_{n \rightarrow \infty} \sup_{B_n \in \mathcal{B}_{c,n}^p} \inf_{B \in \mathcal{B}_c^*} d(B_n, B) = 0, \quad (4.12)$$

$$\lim_{n \rightarrow \infty} \sup_{B_n \in \mathcal{B}_{c,n}^{p,+}} \inf_{B \in \mathcal{B}_c^*} d(B_n, B) = 0. \quad (4.13)$$

*Proof.* If (4.11) is false, then for some  $\delta > 0$  and some subsequence  $n_k$  we can find sets  $\tilde{B}_{n_k} \in \mathcal{B}_{c,n_k}^v$  such that  $\inf_{\tilde{B} \in \mathcal{B}_c^*} d(\tilde{B}_{n_k}, \tilde{B}) > \delta$ . But because of (2.6) and the compactness of  $(\mathcal{B}_c, d)$  one can extract a further subsequence  $n_{k_j}$  and sets  $B_j$  converging w.r.t.  $d$  to some  $B \in \mathcal{B}_c$ . Lemma 4.1 yields that  $|M(B_j^+) - M(B_j^-)| \rightarrow 0$  which implies  $B \in \mathcal{B}_c^*$  and hence the contradictory fact that  $d(B_j, B) \rightarrow 0$ .

The proof of (4.13) follows similarly. Clearly (4.13) implies (4.12).  $\square$

For  $c > 0$  consider  $C \in \mathcal{C}^{c\varepsilon_n}$ . Write  $C^+ = C \setminus L_\lambda$  and  $C^- = C \cap L_\lambda$ . Define

$$D_n(\tau_{\varepsilon_n}(C)) = n^{2/3} (e_\lambda(C^-) - e_\lambda(C^+))$$

and observe that  $e_\lambda(C^-) \geq 0$  and  $e_\lambda(C^+) \leq 0$ .

**Lemma 4.3.** *If (2.14)–(2.15) hold, then, as  $n \rightarrow \infty$ ,*

$$\sup_{C \in \mathcal{C}^{c\varepsilon_n}} |D(\tau_{\varepsilon_n}(C)) - D_n(\tau_{\varepsilon_n}(C))| \rightarrow 0.$$

*Proof.* Write  $f'(\pi, s) = 1_{s>0}f'_+(\pi) + 1_{s\leq 0}f'_-(\pi)$ . From the Steiner formula (4.1)–(4.2) and from (2.14)–(2.15) we obtain by a straightforward calculation that, uniformly for  $C \in \mathcal{C}^{c\varepsilon_n}$ ,

$$\begin{aligned} & D_n(\tau_{\varepsilon_n}(C)) \\ &= \frac{1}{\varepsilon_n^2} \int_{\text{Nor}(L_\lambda)} \int_{-(r(\pi) \wedge c\varepsilon_n)}^{c\varepsilon_n} s f'(\pi, s) (1_{C^+}(\pi + su) - 1_{C^-}(\pi + su)) ds d\nu_{d-1}(\pi, u) \\ &+ \sum_{j=2}^d \binom{d-1}{j-1} \frac{1}{\varepsilon_n^2} \int_{\text{Nor}(L_\lambda)} \int_{-(r(\pi) \wedge c\varepsilon_n)}^{c\varepsilon_n} s^j f'(\pi, s) (1_{C^+}(\pi + su) - 1_{C^-}(\pi + su)) ds d\nu_{d-j}(\pi, u) \\ &+ o(1) \\ &=: T_{1,n}(C) + \sum_{j=2}^d \binom{d-1}{j-1} T_{j,n}(C) + o(1). \end{aligned}$$

Now by a change of variables it follows that  $T_{1,n}(C) = D(\tau_{\varepsilon_n}(C))$ . Hence it remains to show that  $\sup_{C \in \mathcal{C}^{c\varepsilon_n}} \sum_{j=2}^d \binom{d-1}{j-1} |T_{j,n}(C)| \rightarrow 0$ , but this follows from  $\sup_{C \in \mathcal{C}^{c\varepsilon_n}} |T_{j,n}(C)| = O(\varepsilon_n^{j-1})$ , which we obtain from (4.3).  $\square$

The following lemma is immediate from basic measure theory, more precisely the fact that an  $M$ -small set has a small integral.

**Lemma 4.4.** *Assuming (2.14)–(2.15) we have, as  $n \rightarrow \infty$ ,*

$$\sup_{B_n \in \mathcal{B}_{c,n}, B \in \mathcal{B}_c, d(B_n, B) \leq \gamma_{c,n}} |D(B_n) - D(B)| \rightarrow 0.$$

## 4.2. Concentration lemmas

**Lemma 4.5.** *Let  $\varepsilon > 0$  fixed and  $A \in \mathcal{A}$  with  $d_H(A, L_\lambda) \leq \varepsilon$ , then  $A \Delta L_\lambda \subset S_\lambda^\varepsilon$ .*

*Proof.* Assume  $d_H(A, L_\lambda) \leq \varepsilon$  and  $x \in A \setminus L_\lambda$ . Then  $\|x - \Pi(x)\| \leq \varepsilon$ . Hence  $x \in S_\lambda^\varepsilon$ . Now assume  $d_H(A, L_\lambda) \leq \varepsilon$  and  $x \in L_\lambda \setminus A$ . Assume  $x \notin S_\lambda^\varepsilon$ . Let  $\Pi_A(x)$  be the orthogonal projection of  $x$  on  $\partial A$ , that is unique since  $x \notin A$  and  $A$  is convex. There exists an  $y \in S_\lambda$  such that  $\Pi_A(y) = \Pi_A(x)$ . To see this consider the tangent space of  $A$  at  $\Pi_A(x)$  that is orthogonal to the outer normal of  $\partial A$  at  $\Pi_A(x)$  driven by  $(x - \Pi_A(x))$  and take  $y$  as the intersection of that line with  $S_\lambda$ . Then  $\|y - \Pi_A(x)\| > \|y - x\| \geq \|\Pi(x) - x\| > \varepsilon$ . This implies  $d_H(\{y\}, A) > \varepsilon$  and hence  $d_H(A, L_\lambda) > \varepsilon$ . Contradiction. Hence we have  $x \in S_\lambda^\varepsilon$ .  $\square$

PROOF OF LEMMA 3.1. Consider

$$\begin{aligned} L_{1,n} &\in \arg \max_{A \in \mathcal{A}} \{P_n(A) - \lambda \mu(A)\} \\ &= \arg \max_{A \in \mathcal{A}} \{P_n(A) - P(A) - P_n(L_\lambda) + P(L_\lambda) + e_\lambda(A) - e_\lambda\}. \end{aligned} \quad (4.14)$$

Observe that the expression of which the latter argmax is taken is equal to 0 in case  $A = L_\lambda$ .

**VC class.** First assume that  $\mathcal{A}$  is a VC class. We begin with showing that for  $n$  large enough

$$\mathbb{P}(d_H(L_{1,n}, L_\lambda) \geq \delta) \leq \frac{1}{2}\delta. \quad (4.15)$$

We obtain from (2.13) that there exists an  $\eta > 0$ , such that  $d_H(L_\lambda, A) \geq \delta$  implies  $e_\lambda - e_\lambda(A) \geq 2\eta$ . The Glivenko-Cantelli theorem on  $\mathcal{A}$  yields that for the above  $\eta$  for large  $n$

$$\mathbb{P}(\sup_{A \in \mathcal{A}} |P_n(A) - P(A) - P_n(L_\lambda) + P(L_\lambda)| \leq \eta) \geq 1 - \frac{1}{2}\delta.$$

Hence

$$\mathbb{P}(d_H(L_{1,n}, L_\lambda) < \delta) \geq 1 - \frac{1}{2}\delta.$$

Define, for  $c > 1$  (recall  $\varepsilon_n = n^{-1/3}$ ),

$$p_c = \mathbb{P}(c\varepsilon_n \leq d_H(L_{1,n}, L_\lambda) \leq \min(c^2\varepsilon_n, \delta)). \quad (4.16)$$

and  $\mathcal{A}_1 = \{A \in \mathcal{A} : c\varepsilon_n \leq d_H(L_\lambda, A) \leq \min(c^2\varepsilon_n, \delta)\}$ . From (2.16) we obtain

$$e_\lambda - e_\lambda(A) > \frac{1}{2}\eta_0 d_H^2(L_\lambda, A), \quad \text{for small } d_H(L_\lambda, A).$$

Hence for small  $\delta > 0$

$$\inf_{A \in \mathcal{A}_1} e_\lambda - e_\lambda(A) \geq \frac{1}{2}\eta_0 \inf_{A \in \mathcal{A}_1} d_H^2(L_\lambda, A) \geq \frac{1}{2}\eta_0 c^2 \varepsilon_n^2.$$

This yields, see (4.14) and the observation directly below it,

$$\begin{aligned}
 p_c &= \mathbb{P}(L_{1,n} \in \mathcal{A}_1) \\
 &\leq \mathbb{P}\left(\sup_{A \in \mathcal{A}_1} P_n(A) - P(A) - P_n(L_\lambda) + P(L_\lambda) + e_\lambda(A) - e_\lambda \geq 0\right) \\
 &\leq \mathbb{P}\left(\sup_{A \in \mathcal{A}_1} P_n(A) - P(A) - P_n(L_\lambda) + P(L_\lambda) \geq \inf_{A \in \mathcal{A}_1} e_\lambda - e_\lambda(A)\right) \\
 &\leq \mathbb{P}\left(\sup_{A \in \mathcal{A}_1} P_n(A) - P(A) - P_n(L_\lambda) + P(L_\lambda) \geq \frac{1}{2}\eta_0 c^2 \varepsilon_n^2\right) \\
 &\leq \mathbb{P}\left(2 \sup_{D \in \mathcal{D}_n} |P_n(D) - P(D)| \geq \frac{1}{2}\eta_0 c^2 \varepsilon_n^2\right) =: \tilde{p}_c, \tag{4.17}
 \end{aligned}$$

where

$$\mathcal{D}_n = \{A \setminus L_\lambda : A \in \mathcal{A}_1\} \cup \{L_\lambda \setminus A : A \in \mathcal{A}_1\}.$$

Denote  $N = nP_n(S_\lambda^{c^2 \varepsilon_n \wedge \delta})$ , the number of observations in  $S_\lambda^{c^2 \varepsilon_n \wedge \delta}$ ,  $a_n = P(S_\lambda^{c^2 \varepsilon_n \wedge \delta})$ , and  $k = na_n$ , the expected number of observations in  $S_\lambda^{c^2 \varepsilon_n \wedge \delta}$ . Then

$$\begin{aligned}
 \tilde{p}_c &= \mathbb{P}\left(\frac{n^{1/2}}{a_n^{1/2}} \sup_{D \in \mathcal{D}_n} |P_n(D) - P(D)| \geq \frac{1}{4}\eta_0 c^2 \varepsilon_n^2 \frac{n^{1/2}}{a_n^{1/2}}\right) \\
 &\leq \mathbb{P}\left(\frac{n^{1/2}}{a_n^{1/2}} \sup_{D \in \mathcal{D}_n} |P_n(D) - P(D)| \geq \frac{1}{5}(2\lambda s_\lambda)^{-1/2} \eta_0 c\right),
 \end{aligned}$$

since  $c^2 \varepsilon_n / a_n \geq (\frac{4}{5})^2 (2\lambda s_\lambda)^{-1}$ , for small  $\delta > 0$ . (Observe that here by the choice  $\varepsilon_n = n^{-1/3}$  the threshold in the first probability is bounded from below by a positive number not depending on  $n$ .) Hence, writing  $\eta_2 = 1 \wedge \frac{1}{5}(2\lambda s_\lambda)^{-1/2} \eta_0$ , we have

$$\begin{aligned}
 \tilde{p}_c &\leq \mathbb{P}\left(\frac{n^{1/2}}{a_n^{1/2}} \sup_{D \in \mathcal{D}_n} |P_n(D) - P(D)| \geq \eta_2 c\right) \\
 &= \sum_{m=0}^n \mathbb{P}\left(\frac{n}{k^{1/2}} \sup_{D \in \mathcal{D}_n} |P_n(D) - P(D)| \geq \eta_2 c \mid N = m\right) \mathbb{P}(N = m) \\
 &\leq \sum_{m=\lceil k - (\eta_2/3)c\sqrt{k} \rceil}^{m=\lfloor k + (\eta_2/3)c\sqrt{k} \rfloor} \mathbb{P}\left(\sup_{D \in \mathcal{D}_n} \left| \frac{1}{\sqrt{k}}(nP_n(D) - nP(D)) \right| \geq \eta_2 c \mid N = m\right) \mathbb{P}(N = m) \\
 &\quad + \mathbb{P}(|N - k| \geq \frac{\eta_2}{3} c\sqrt{k}) =: T + T_3.
 \end{aligned}$$

Let  $Y_j, j = 1, \dots, n$ , be i.i.d. random vectors taking values in  $S_\lambda^{c^2 \varepsilon_n \wedge \delta}$  distributed according to  $\tilde{P} := P/a_n$ . Then, using Lemma 4.5, for  $n \geq n_1$  (for some  $n_1$  not

depending on  $c$ ) and large  $c$ ,

$$\begin{aligned} T \leq & \sum_{m=\lceil k-(\eta_2/3)c\sqrt{k} \rceil}^{m=\lfloor k+(\eta_2/3)c\sqrt{k} \rfloor} \mathbb{P} \left( \sup_{D \in \mathcal{D}_n} \left| \frac{1}{\sqrt{m}} \left( \sum_{j=1}^m 1_D(Y_j) - m\tilde{P}(D) \right) \right| \geq \frac{\sqrt{\eta_2 c}}{2\sqrt{2}} \right) \mathbb{P}(N = m) \\ & + \sum_{m=\lceil k-(\eta_2/3)c\sqrt{k} \rceil}^{m=\lfloor k+(\eta_2/3)c\sqrt{k} \rfloor} \mathbb{P} \left( \sup_{D \in \mathcal{D}_n} \frac{1}{\sqrt{k}} |m - k| \tilde{P}(D) \geq \frac{\eta_2}{2} c \right) \mathbb{P}(N = m) =: T_1 + T_2. \end{aligned}$$

Note that for each  $m$  in the latter sum

$$\mathbb{P} \left( \sup_{D \in \mathcal{D}_n} \frac{1}{\sqrt{k}} |m - k| \tilde{P}(D) \geq \frac{\eta_2}{2} c \right) \leq \mathbb{P} \left( \frac{1}{\sqrt{k}} |m - k| \geq \frac{\eta_2}{2} c \right) = 0.$$

Hence  $T_2 = 0$ . From Bennett's inequality we have

$$T_3 \leq 2 \exp \left( -\frac{\eta_2^2 c}{19} \right).$$

Using Corollary 2.9 in [Alexander \(1984\)](#) we obtain for large enough  $c$

$$T_1 \leq \sum_{m=\lceil k-(\eta_2/3)c\sqrt{k} \rceil}^{m=\lfloor k+(\eta_2/3)c\sqrt{k} \rfloor} 16 \exp \left( -\frac{\eta_2 c}{64} \right) \mathbb{P}(N = m) \leq 16 \exp \left( -\frac{\eta_2 c}{64} \right).$$

Hence

$$p_c \leq 18 \exp \left( -\frac{\eta_2^2 c}{64} \right).$$

Using this bound on  $p_c$  with  $c$  replaced by  $c^{2^m}$ ,  $m = 0, 1, 2, \dots$ , we obtain that for large  $n$

$$\begin{aligned} & \mathbb{P}(d_H(L_{1,n}, L_\lambda) \geq c\varepsilon_n) \\ & \leq \mathbb{P}(d_H(L_{1,n}, L_\lambda) \geq \delta) + \sum_{m=0}^{\infty} \mathbb{P}(c^{2^m} \varepsilon_n \leq d_H(L_{1,n}, L_\lambda) \leq \min(c^{2^{m+1}} \varepsilon_n, \delta)) \\ & \leq \frac{1}{2} \delta + 18 \sum_{m=0}^{\infty} \exp \left( -\frac{\eta_2^2 c^{2^m}}{64} \right) \leq \delta, \end{aligned}$$

for  $c$  large enough.

**Conditions (2.7)–(2.10).** Now we assume that (2.8) and (2.9) hold. Then the proof for  $L_{1,n}$  follows similar lines. (Here (2.7) and (2.10) are not required, but they are crucial for the convergence statements in Lemma 4.6 below.) As above, we show that (4.15) holds, because of

$$\sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty,$$

which now directly follows from (2.9). For the derivation of the upper bound for  $p_c$  in (4.16) we first follow the same lines as below that formula, until (4.17). Now observe that

$$[\mathcal{A}]_{c^2 \wedge \delta / \varepsilon_n, n} ((P(S_\lambda^{c^2 \varepsilon_n \wedge \delta}) / \varepsilon_n)^{1/2}) = 1.$$

Using this and Lemma 4.5, we now bound the final probability in (4.17) directly by Lemma 19.34 in van der Vaart (1998) and the Markov inequality. This yields for some constants  $c_1, c_2 > 1$ ,

$$\begin{aligned} p_c &\leq \frac{4c_1}{\eta_0 c^2 \varepsilon_n^2 n^{1/2}} \int_0^{c_2 (c \varepsilon_n^{1/2} \wedge \delta^{1/2})} \sqrt{\log [\mathcal{A}]_{c^2 \wedge \delta / \varepsilon_n, n} (t / \varepsilon_n^{1/2})} dt \\ &= \frac{4c_1}{\eta_0 c^2} \int_0^{c_2 (c \wedge (\delta / \varepsilon_n))^{1/2}} \sqrt{\log [\mathcal{A}]_{c^2 \wedge \delta / \varepsilon_n, n} (\kappa)} d\kappa \leq \frac{4c_1 c_2}{\eta_0 c^2} \int_0^{c \wedge (\delta / \varepsilon_n)^{1/2}} \sqrt{\log [\mathcal{A}]_{c^2 \wedge \delta / \varepsilon_n, n} (\kappa)} d\kappa. \end{aligned}$$

Using this bound on  $p_c$  with  $c$  replaced by  $c^{2^m}$ ,  $m = 0, 1, 2, \dots$ , we obtain as above

$$\mathbb{P}(d_H(L_{1,n}, L_\lambda) \geq c \varepsilon_n) \leq \frac{1}{2} \delta + \sum_{m=0}^{\infty} \frac{4c_1 c_2}{\eta_0 c^{2^{m+1}}} \int_0^{c^{2^m} \wedge (\delta / \varepsilon_n)^{1/2}} \sqrt{\log [\mathcal{A}]_{c^{2^{m+1}} \wedge \delta / \varepsilon_n, n} (\kappa)} d\kappa.$$

Now using (2.8) and a change of variables we find that the expression on the right is, for some  $c_3 > 0$ , bounded by

$$\frac{1}{2} \delta + \sum_{m=0}^{\infty} \frac{4c_1 c_2 c_3}{\eta_0 c^{2^{m+1}}} c^{2^m} \wedge (\delta / \varepsilon_n)^{1/2} \leq \frac{1}{2} \delta + \frac{4c_1 c_2 c_3}{\eta_0} \sum_{m=0}^{\infty} \frac{1}{c^{2^m}} \leq \delta,$$

for large  $n$  and  $c$  large enough.

Next we consider  $L_{4,n}$ . We have

$$L_{4,n} \in \arg \max_{A \in \mathcal{A}, \mu(A) = v_\lambda} \{P_n(A) - P(A) - P_n(L_\lambda) + P(L_\lambda) + e_\lambda(A) - e_\lambda\}.$$

This expression is very similar to the one for  $L_{1,n}$ . The only difference is that  $\mathcal{A}$  there is replaced by its subset  $\{A \in \mathcal{A} : \mu(A) = v_\lambda\}$ . Since the arguments above – dealing with suprema and infima – hold for the entire class  $\mathcal{A}$ , they remain to hold for this subset.

Finally consider  $L_{2,n}$ . We have, almost surely,

$$\begin{aligned} L_{2,n} &\in \arg \min_{A \in \mathcal{A}, nP_n(A) = \lceil np_\lambda \rceil} \{\mu(A)\} \\ &= \arg \max_{A \in \mathcal{A}, nP_n(A) = \lceil np_\lambda \rceil} \{P_n(A) - \lambda \mu(A)\} \\ &= \arg \max_{A \in \mathcal{A}, nP_n(A) = \lceil np_\lambda \rceil} \{P_n(A) - P(A) - P_n(L_\lambda) + P(L_\lambda) + e_\lambda(A) - e_\lambda\}. \end{aligned}$$

This expression looks similar to the ones for  $L_{1,n}$  and  $L_{4,n}$ , but the difference is that the supremum of the expression of which the latter argmax is taken is not

guaranteed to be non-negative since the choice  $A = L_\lambda$ , as before, is not allowed. However, it follows from (3.6) that, almost surely, there exists an  $A_{\hat{s}} \in \mathcal{A}_l$  such that  $nP_n(A_{\hat{s}}) = \lceil np_\lambda \rceil$ . Then, using  $P(A_{\hat{s}}) = p_\lambda + O_{\mathbb{P}}(1/\sqrt{n})$ , we obtain from (3.7) and the behavior of the oscillation modulus of the univariate, uniform empirical process, that with arbitrarily high probability for large  $n$  that the just mentioned supremum is larger than  $-n^{-17/24}$  (instead of being non-negative). Since  $n^{-17/24}/\varepsilon_n^2 \rightarrow 0$  as  $n \rightarrow \infty$ , the proof for  $L_{1,n}$  can be easily adapted, replacing  $\mathcal{A}$  by its (random) subset  $\{A \in \mathcal{A} : nP_n(A) = \lceil np_\lambda \rceil\}$ .  $\square$

### 4.3. Processes on the cylinder space

Here we describe more precisely the local objects, magnified into the cylinder space, namely the empirical process, the drift induced by the local variation of the density, and then the limiting drifted Gaussian process.

Since for all  $c > 0$ ,  $(\mathcal{B}_c, d)$  is totally bounded, we have

$$\sup_{B \in \mathcal{B}_c} \inf_{B_n \in \mathcal{B}_{c,n}} d(B_n, B) \rightarrow 0.$$

Combining this with (2.6), we have in terms of Hausdorff distance between classes of sets that for any  $c > 0$ , as  $n \rightarrow \infty$ ,

$$\gamma_{c,n} := \max \left( \sup_{B_n \in \mathcal{B}_{c,n}} \inf_{B \in \mathcal{B}_c} d(B_n, B), \sup_{B \in \mathcal{B}_c} \inf_{B_n \in \mathcal{B}_{c,n}} d(B_n, B) \right) \rightarrow 0.$$

Define

$$\Lambda_n(C) = n^{2/3}(P_n(C) - P(C)), \quad C \in \mathcal{B}(\mathbb{R}^d),$$

and

$$w_n(B) = \Lambda_n(\tau_{\varepsilon_n}^{-1}(B^+)) - \Lambda_n(\tau_{\varepsilon_n}^{-1}(B^-)), \quad B \in \mathcal{F}_c.$$

**Lemma 4.6.** *Assume that  $H_1$  and  $H_2$  hold. Let  $c > 0$ . Then on some probability space there exists a triangular array  $X_{n,1}, \dots, X_{n,n}, n \in \mathbb{N}$ , of rowwise independent random vectors with law  $P$  on  $\mathbb{R}^d$  together with a bounded,  $d$ -continuous version of  $W$  on  $\mathcal{B}_c$  such that, as  $n \rightarrow \infty$ ,*

$$\sup_{B_n \in \mathcal{B}_{c,n}, B \in \mathcal{B}_c, d(B_n, B) \leq \gamma_{c,n}} |w_n(B_n) - w_n(B)| \xrightarrow{\mathbb{P}} 0, \quad (4.18)$$

and, with probability 1,

$$\sup_{B \in \mathcal{B}_c} |w_n(B) - \sqrt{\lambda}W(B)| \rightarrow 0. \quad (4.19)$$

*Proof.* Note that the assumptions of Theorems 1 and 2 in Einmahl and Khmaladze (2011) are satisfied. In particular (2.7) and (2.10) or the VC class assumption are crucial, while (2.14) and (2.15) are a bit stronger than needed here, but required for the main proofs. Hence, using these theorems, including a Skorohod construction as on page 554 therein, yields (4.18) and (4.19). Note that the generalization from  $c = 1$  therein to arbitrary  $c > 0$  here, is straightforward. Also the fact that here  $w_n$  is a difference of two terms can be easily dealt with.  $\square$



For a compact subset  $\tilde{\mathcal{B}}$  of  $\mathcal{B}$ , define

$$Z(\tilde{\mathcal{B}}) = \arg \max_{B \in \tilde{\mathcal{B}}} \left\{ \sqrt{\lambda} W(B) - D(B) \right\}.$$

Recall that  $\sqrt{\lambda} W - D$  is  $d$ -continuous on  $\mathcal{B}_c$ , whereas  $\text{Var}(W(B) - W(B')) = 0$  implies  $B' = B$  by our equivalence class convention. Now note that both  $Z(\mathcal{B}_c)$  and  $Z(\mathcal{B}_c^*)$  exist and, by Lemma 2.6 in Kim and Pollard (1990), are almost surely unique on the compact set  $\mathcal{B}_c$ , respectively  $\mathcal{B}_c^*$ . Proposition 3.1 and a similar statement for  $Z(\mathcal{B}^*)$  are consequences of (the above and) the following lemma.

**Lemma 4.7.** *Assume that  $H_1$ ,  $H_2$ , and (3.2) hold. For  $\tilde{\mathcal{B}} = \mathcal{B}, \mathcal{B}^*$  we have*

$$\mathbb{P} \left( \bigcup_{c>0} \{Z(\tilde{\mathcal{B}} \cap \mathcal{B}_c) = Z(\tilde{\mathcal{B}} \cap \mathcal{B}_{\tilde{c}}), \text{ for all } \tilde{c} > c\} \right) = 1.$$

Hence  $Z(\tilde{\mathcal{B}})$  almost surely exists and is unique; it is the “set limit” of  $Z(\tilde{\mathcal{B}} \cap \mathcal{B}_m)$ :

$$Z(\tilde{\mathcal{B}}) = \bigcap_{k=1}^{\infty} \bigcup_{m=k}^{\infty} Z(\tilde{\mathcal{B}} \cap \mathcal{B}_m).$$

*Proof.* We have, using (2.17) and  $c(B)$  defined there,

$$\begin{aligned} & \mathbb{P} \left( \sup_{B \in \tilde{\mathcal{B}}: c(B) \geq c} \sqrt{\lambda} W(B) - D(B) \geq 0 \right) \\ & \leq \sum_{m=0}^{\infty} \mathbb{P} \left( \sup_{B \in \tilde{\mathcal{B}}: c^{2^m} \leq c(B) < c^{2^{m+1}}} \sqrt{\lambda} W(B) - D(B) \geq 0 \right) \\ & \leq \sum_{m=0}^{\infty} \mathbb{P} \left( \sup_{B \in \tilde{\mathcal{B}}: c^{2^m} \leq c(B) < c^{2^{m+1}}} \sqrt{\lambda} W(B) - \eta_0 c^2(B) \geq 0 \right) \\ & \leq \sum_{m=0}^{\infty} \mathbb{P} \left( \sup_{B \in \tilde{\mathcal{B}}: c^{2^m} \leq c(B) < c^{2^{m+1}}} \sqrt{\lambda} W(B) \geq \eta_0 c^{2^{m+1}} \right), \end{aligned}$$

which is by (3.2), bounded from above by

$$\frac{\eta_1 \sqrt{\lambda}}{\eta_0} \sum_{m=0}^{\infty} c^{-2^m},$$

which is, for arbitrary  $\eta > 0$ , bounded by  $\eta$ , for  $c$  large enough.

Hence, since  $L_\lambda \in \mathcal{A}$ , for  $c$  large enough,

$$\mathbb{P}(Z(\tilde{\mathcal{B}} \cap \mathcal{B}_c) = Z(\tilde{\mathcal{B}} \cap \mathcal{B}_{\tilde{c}}), \text{ for all } \tilde{c} > c) \geq 1 - \eta.$$

If this event is denoted by  $\Omega_c$ , then  $\mathbb{P}(\bigcup_{c>0} \Omega_c) = 1$ . □

#### 4.4. Proof of Theorem 3.1

We work in the setting of Lemma 4.6. For  $c > 0$  we have

$$\begin{aligned} & \arg \max_{A \in \mathcal{A}^{c\varepsilon_n}} \{P_n(A) - \lambda\mu(A)\} \\ &= \arg \max_{A \in \mathcal{A}^{c\varepsilon_n}} \left\{ P(A) - \lambda\mu(A) - P(L_\lambda) + \lambda\mu(L_\lambda) + n^{-2/3}(\Lambda_n(A) - \Lambda_n(L_\lambda)) \right\} \\ &= \arg \max_{A \in \mathcal{A}^{c\varepsilon_n}} \left\{ n^{2/3}(e_\lambda(A) - e_\lambda) + \Lambda_n(A) - \Lambda_n(L_\lambda) \right\} \\ &= \varphi_{\varepsilon_n} \left\{ \arg \max_{B \in \mathcal{B}_{c,n}} \{w_n(B) - D_n(B)\} \right\}. \end{aligned}$$

Consider the events

$$\Xi_{c,n}^\Lambda = \{L_{1,n} \triangle L_\lambda \subset S_\lambda^{c\varepsilon_n}\}, \quad \Xi_c^W = \{Z(\mathcal{B}_c) = Z(\mathcal{B})\},$$

where  $Z(\mathcal{B}_c)$  and  $Z(\mathcal{B})$  are defined in terms of a Wiener process  $W$  satisfying (4.19) in Lemma 4.6. Clearly, Lemmas 4.5, 3.1 and 4.7 imply that for any  $\delta > 0$  there exists a  $c = c(\delta) > 0$  such that we have  $\mathbb{P}(\Xi_{c,n}^\Lambda \cap \Xi_c^W) > 1 - \delta$  for all  $n$  large enough. Now define

$$m_{c,n} = \sup_{B \in \mathcal{B}_{c,n}} \{w_n(B) - D_n(B)\}, \quad m_c = \max_{B \in \mathcal{B}_c} \left\{ \sqrt{\lambda}W(B) - D(B) \right\}$$

and observe that Lemmas 4.6, 4.3 and 4.4 imply

$$m_{c,n} \xrightarrow{\mathbb{P}} m_c, \text{ as } n \rightarrow \infty. \quad (4.20)$$

We have for any  $\varepsilon > 0$  fixed, every argmax  $L_{1,n}$ , and all large enough  $n$

$$\begin{aligned} & \mathbb{P}(M(\tau_{\varepsilon_n}(L_{1,n} \triangle L_\lambda) \triangle Z(\mathcal{B})) > \delta^2) \\ & \leq \mathbb{P}(\{M(\tau_{\varepsilon_n}(L_{1,n} \triangle L_\lambda) \triangle Z(\mathcal{B})) > \delta^2\} \cap \Xi_{c,n}^\Lambda \cap \Xi_c^W) + \delta \\ & \leq \mathbb{P}\left(d\left(\arg \max_{B \in \mathcal{B}_{c,n}} \{w_n(B) - D_n(B)\}, Z(\mathcal{B}_c)\right) > \delta\right) + \delta \\ & \leq \mathbb{P}\left(\sup_{B \in \mathcal{B}_{c,n}: d(B, Z(\mathcal{B}_c)) > \delta} \{w_n(B) - D_n(B)\} \geq m_{c,n}\right) + \delta \\ & \leq \mathbb{P}\left(\sup_{B \in \mathcal{B}_{c,n}: d(B, Z(\mathcal{B}_c)) > \delta} \{w_n(B) - D(B)\} \geq m_{c,n} - \varepsilon\right) + \delta \end{aligned}$$

which is by (4.20)

$$\leq \mathbb{P}\left(\sup_{B \in \mathcal{B}_{c,n}: d(B, Z(\mathcal{B}_c)) > \delta} \{w_n(B) - D(B)\} \geq m_c - 2\varepsilon\right) + 2\delta$$

which by (4.18) and Lemma 4.4 is in turn

$$\leq \mathbb{P} \left( \sup_{B \in \mathcal{B}_c: d(B, Z(\mathcal{B}_c)) \geq \delta/2} \{w_n(B) - D(B)\} \geq m_c - 3\varepsilon \right) + 3\delta$$

and this is by (4.19) and then by Lemma 2.6 in Kim and Pollard (1990)

$$\leq \mathbb{P} \left( \max_{B \in \mathcal{B}_c: d(B, Z(\mathcal{B}_c)) \geq \delta/2} \left\{ \sqrt{\lambda} W(B) - D(B) \right\} \geq m_c - 4\varepsilon \right) + 4\delta \leq 5\delta,$$

provided that we choose a small enough  $\varepsilon$  with respect to  $\delta$ .

Note that  $Z(\mathcal{B})$  depends on  $\delta$  through  $c = c(\delta)$ . We can avoid this, but make it instead depend on  $n$  as in the statement of the theorem, by a diagonal selection argument.

The second and third statement in Theorem 3.1 follow directly from the just established first one and the Steiner formula (4.1)–(4.2), since  $M$  can be approximated by  $\varepsilon_n^{-1} \mu$  after transforming back by  $\tau_{\varepsilon_n}^{-1}$  (see (4.10)), and then  $\lambda \mu$  can be approximated by  $P$  near  $S_\lambda$ .  $\square$

#### 4.5. Proof of Theorem 3.2

The proof of Theorem 3.2 with  $L_{3,n}$  replaced by  $L_{4,n}$  from (3.9) is similar to that of Theorem 3.1, only  $\mathcal{A}$  has to be replaced by  $\mathcal{A}_v$  and  $\mathcal{B}$  by  $\mathcal{B}^*$ .

Now take an argmax  $L_{3,n}$  with  $\mu(L_{3,n}) < v_\lambda$ . Then, using (2.12), for some  $L_{4,n}$  we have  $L_{3,n} \subset L_{4,n}$ . Now, since  $n^{1/2}(P_n - P) = O_{\mathbb{P}}(1)$  uniformly on  $\mathcal{A}$ , we have with probability tending to 1,

$$P(L_{3,n}) \geq P_n(L_{3,n}) - \frac{1}{2}n^{-2/5} \geq P_n(L_\lambda) - \frac{1}{2}n^{-2/5} \geq p_\lambda - n^{-2/5}.$$

Since  $P(L_{3,n}) - \lambda \mu(L_{3,n}) \leq p_\lambda - \lambda v_\lambda$  we get  $\mu(L_{3,n}) \geq v_\lambda - \frac{1}{\lambda}n^{-2/5}$  thus

$$\mu(L_{3,n} \triangle L_{4,n}) = \mu(L_{4,n}) - \mu(L_{3,n}) \leq \frac{1}{\lambda}n^{-2/5} = o(\varepsilon_n).$$

This also implies that  $M(\tau_{\varepsilon_n}(L_{3,n} \triangle L_{4,n})) \leq \varepsilon_n(\mu(L_{4,n}) - \mu(L_{3,n})) \xrightarrow{\mathbb{P}} 0$  and the statements of Theorem 3.2 for  $j = 3$  follow from those for  $j = 4$ .

Finally we consider  $L_{2,n}$ . We follow again the line of reasoning and the notation in the proof of Theorem 3.1. Define

$$\hat{\mathcal{B}}_{c,n} = \{\tau_{\varepsilon_n}(A \triangle L_\lambda) : A \in \mathcal{A}^{c\varepsilon_n}, P_n(A) = \lceil np_\lambda \rceil / n\}.$$

We have for  $c > 0$

$$\begin{aligned}
 & \arg \min_{A \in \mathcal{A}_{c\varepsilon_n}, nP_n(A) = \lceil np_\lambda \rceil} \{\mu(A)\} \\
 = & \arg \max_{A \in \mathcal{A}_{c\varepsilon_n}, nP_n(A) = \lceil np_\lambda \rceil} \{P_n(A) - \lambda\mu(A)\} \\
 = & \arg \max_{A \in \mathcal{A}_{c\varepsilon_n}, nP_n(A) = \lceil np_\lambda \rceil} \left\{ P(A) - \lambda\mu(A) - P(L_\lambda) + \lambda\mu(L_\lambda) + n^{-2/3}(\Lambda_n(A) - \Lambda_n(L_\lambda)) \right\} \\
 = & \arg \max_{A \in \mathcal{A}_{c\varepsilon_n}, nP_n(A) = \lceil np_\lambda \rceil} \left\{ n^{2/3}(e_\lambda(A) - e_\lambda) + \Lambda_n(A) - \Lambda_n(L_\lambda) \right\} \\
 = & \varphi_{\varepsilon_n} \left( \arg \max_{B \in \hat{\mathcal{B}}_{c,n}} \{w_n(B) - D_n(B)\} \right).
 \end{aligned}$$

Consider the events

$$\Xi_{c,n}^{\Lambda,*} = \{L_{2,n} \Delta L_\lambda \subset S_\lambda^{c\varepsilon_n}\}, \quad \Xi_c^{W,*} = \{Z(\mathcal{B}_c^*) = Z(\mathcal{B}^*)\}.$$

Again, Lemmas 4.5, 3.1 and 4.7 imply that for any  $\delta > 0$  there exists a  $c = c(\delta) > 0$  such that we have  $\mathbb{P}(\Xi_{c,n}^{\Lambda,*} \cap \Xi_c^{W,*}) > 1 - \delta$  for all  $n$  large enough. Now define

$$\begin{aligned}
 \hat{m}_{c,n} &= \sup_{B \in \hat{\mathcal{B}}_{c,n}} \{w_n(B) - D_n(B)\}, \quad m_{c,n}^p = \sup_{B \in \mathcal{B}_{c,n}^p} \{w_n(B) - D(B)\}, \\
 m_c^* &= \max_{B \in \mathcal{B}_c^*} \left\{ \sqrt{\lambda}W(B) - D(B) \right\}
 \end{aligned}$$

and note that by (3.8), the asymptotic equicontinuity of  $w_n$  (as in the proof of Lemma 4.6 given in Einmahl and Khmaladze (2011)), and Lemma 4.3, for  $\varepsilon > 0$ ,

$$\mathbb{P}(m_{c,n}^p \leq \hat{m}_{c,n} + \varepsilon) \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (4.21)$$

and that by (3.5), (4.12) and Lemma 4.6 (possibly with a larger  $\gamma_{c,n} \rightarrow 0$ ),

$$m_{c,n}^p \xrightarrow{\mathbb{P}} m_c^*, \quad \text{as } n \rightarrow \infty. \quad (4.22)$$

Recall the definition of  $\mathcal{B}_{c,n}^{p,+}$  in (4.4). We have for  $\varepsilon > 0$ , every argmin  $L_{2,n}$ , and all large enough  $n$

$$\begin{aligned}
 & \mathbb{P}(M(\tau_{\varepsilon_n}(L_{2,n} \Delta L_\lambda) \Delta Z(\mathcal{B}^*)) > \delta^2) \\
 & \leq \mathbb{P}(\{M(\tau_{\varepsilon_n}(L_{2,n} \Delta L_\lambda) \Delta Z(\mathcal{B}^*)) > \delta^2\} \cap \Xi_{c,n}^{\Lambda,*} \cap \Xi_c^{W,*}) + \delta \\
 & \leq \mathbb{P}\left(d\left(\arg \max_{B \in \hat{\mathcal{B}}_{c,n}} \{w_n(B) - D_n(B)\}, Z(\mathcal{B}_c^*)\right) > \delta\right) + \delta
 \end{aligned}$$

$$\begin{aligned}
 &\leq \mathbb{P} \left( \sup_{B \in \hat{\mathcal{B}}_{c,n}: d(B, Z(\mathcal{B}_c^*)) > \delta} \{w_n(B) - D_n(B)\} \geq \hat{m}_{c,n} \right) + \delta \\
 &\leq \mathbb{P} \left( \sup_{B \in \hat{\mathcal{B}}_{c,n}: d(B, Z(\mathcal{B}_c^*)) > \delta} \{w_n(B) - D(B)\} \geq \hat{m}_{c,n} - \varepsilon \right) + \delta \\
 &\leq \mathbb{P} \left( \sup_{B \in \mathcal{B}_{c,n}^{\rho,+}: d(B, Z(\mathcal{B}_c^*)) > \delta} \{w_n(B) - D(B)\} \geq \hat{m}_{c,n} - \varepsilon \right) + 2\delta
 \end{aligned}$$

which is by (4.21)

$$\leq \mathbb{P} \left( \sup_{B \in \mathcal{B}_{c,n}^{\rho,+}: d(B, Z(\mathcal{B}_c^*)) > \delta} \{w_n(B) - D(B)\} \geq m_{c,n}^p - 2\varepsilon \right) + 3\delta$$

which is by (4.22)

$$\leq \mathbb{P} \left( \sup_{B \in \mathcal{B}_{c,n}^{\rho,+}: d(B, Z(\mathcal{B}_c^*)) > \delta} \{w_n(B) - D(B)\} \geq m_c^* - 3\varepsilon \right) + 4\delta$$

which by (4.18), Lemma 4.4, and (4.13), is in turn

$$\leq \mathbb{P} \left( \sup_{B \in \mathcal{B}_c^*: d(B, Z(\mathcal{B}_c^*)) > \delta/2} \{w_n(B) - D(B)\} \geq m_c^* - 4\varepsilon \right) + 5\delta$$

and this is by (4.19) and then by again Lemma 2.6 in Kim and Pollard (1990)

$$\leq \mathbb{P} \left( \max_{B \in \mathcal{B}_c^*: d(B, Z(\mathcal{B}_c^*)) > cn^{-1/3}/2} \left\{ \sqrt{\lambda} W(B) - D(B) \right\} \geq m_c^* - 5\varepsilon \right) + 6\delta \leq 7\delta,$$

provided  $\varepsilon$  is chosen small enough. The last two paragraphs of the proof of Theorem 3.1 now yield the stated results.  $\square$

*Details planar convex sets.* In dimension two, let  $\mathcal{A}$  be the class of all closed, convex sets and let  $f$  be such that  $L_\lambda$  is a square, for convenience of presentation the unit square. Let  $\delta > 0$  be fixed and small (so that on  $S_\lambda^\delta$  the density  $f$  is “sufficiently close” to  $\lambda$ ). Now consider, for  $c \in [1, \delta n^{1/3}]$ , the subset  $S_\lambda^{cn^{-1/3}}$ . This set is the difference of a “square with circular corners”, with length of the side  $1 + 2cn^{-1/3}$ , and a smaller square, with length of the side  $1 - 2cn^{-1/3}$ . We cover this set by 8 convex subsets: 4 rectangles  $R_1, \dots, R_4$  and 4 quarter circles  $Q_1, \dots, Q_4$ .  $R_1$  is the rectangle with vertices  $(0, 1 + cn^{-1/3}), (1, 1 + cn^{-1/3}), (1, 1 - cn^{-1/3}),$  and  $(0, 1 - cn^{-1/3})$ ;  $R_2, R_3, R_4$  are defined similarly.  $Q_1 = S_\lambda^{cn^{-1/3}} \cap ((-\infty, 0] \times [1, \infty))$ ;  $Q_2, Q_3, Q_4$  are defined similarly. For these 8 subsets we denote the respective bracketing numbers (similarly as above (2.7)) with  $[\mathcal{A}]_{c,n}^{(j)}(\kappa)$ ,  $j = 1, \dots, 8$ . We have

$$[\mathcal{A}]_{c,n}(\kappa) \leq \prod_{j=1}^8 [\mathcal{A}]_{c,n}^{(j)}(\kappa/\sqrt{8}). \tag{4.23}$$

In Corollary 2.7.9 in van der Vaart and Wellner (1996), it is derived that the log of the usual “global”  $L_2(P)$  bracketing number for closed, convex sets contained in a given subset of  $S_\lambda^\delta$  is bounded by some constant  $K$  times  $\frac{1}{\kappa}$ . Now take, for example,  $j = 1$  and consider the rectangle  $R_5 \subset S_\lambda^\delta$  with vertices  $(0, 1 + \delta)$ ,  $(1, 1 + \delta)$ ,  $(1, 1 - \delta)$  and  $(0, 1 - \delta)$ . This rectangle can be obtained from  $R_1$  by a multiplication in vertical direction (from the line segment from  $(0, 1)$  to  $(1, 1)$ ) with factor  $\delta n^{1/3}/c$ . In this way every set of the form  $(A \triangle L_\lambda) \cap R_1$ ,  $A \in \mathcal{A}^{cn^{-1/3}}$ , has an image  $C'$ , say, contained in  $R_5$ , and  $(C' \triangle L_\lambda) \cap R_5$  is a closed, convex set. Brackets in  $R_1$  can be obtained accordingly via the inverse multiplication. By this scaling property, for  $c \in [1, \delta n^{1/3}]$ , the aforementioned bound for global bracketing numbers can be translated into a bound, not depending on  $n$ , for the local bracketing numbers  $[\mathcal{A}]_{c,n}^{(1)}$ :

$$\log [\mathcal{A}]_{c,n}^{(1)}(c^{1/2}\kappa) \leq 2 \frac{K}{\sqrt{\delta}} \frac{1}{\kappa}, \quad (4.24)$$

and the same bound holds for  $j = 2, 3, 4$ . Now take, for example,  $j = 5$  and define  $Q_5 = S_\lambda^\delta \cap ((-\infty, 0] \times [1, \infty))$ . Again we can obtain  $Q_5$  from  $Q_1$  by a multiplication with factor  $\delta n^{-1/3}/c$ , but now the multiplication has to be applied in both directions (from the point  $(0, 1)$ ). Therefore we obtain by the scaling property, for  $c \in [1, \delta n^{1/3}]$ , the (initially sharper) bound

$$\log [\mathcal{A}]_{c,n}^{(5)}(c^{1/2}\kappa) \leq 2 \frac{K}{\delta} c^{1/2} n^{-1/6} \frac{1}{\kappa} \leq 2 \frac{K}{\sqrt{\delta}} \frac{1}{\kappa}, \quad (4.25)$$

and the same bound holds for  $j = 6, 7, 8$ .

Finally (4.24) and (4.25) in conjunction with (4.23), imply (2.7) and (2.8).

**Acknowledgment.** We thank the Editor, the Associate Editor, and two Referees for many useful comments that greatly improved the paper.

## References

- ALEXANDER, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067.
- BERTHET, P. and EL-NOUTY, C. (2006). Almost sure asymptotic behaviour of the shorth estimators. *Math. Methods Statist.* **14** 379–403.
- BRUNEL, V.-E. (2018). Methods for estimation of convex sets. *Statistical Science* **33** 615–632.
- CADRE, B. (2006). Kernel estimation of density level sets. *J. Multivariate Anal.* **97** 999–1023.
- CAI, J.-J., EINMAHL, J. H. J. and DE HAAN, L. (2011). Estimation of extreme risk regions under multivariate regular variation. *Ann. Statist.* **39** 1803–1826.
- CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2017). Density level sets: asymptotics, inference, and visualization. *J. Amer. Statist. Assoc.* **112** 1684–1696.

- CLÉMENÇON, S., GOIX, N. and SABOURIN, A. (2015). On anomaly ranking and excess-mass curves. In *Proceedings of the AISTATS 2015* 287–295. San Diego.
- DAVIES, L. (1992). The asymptotics of Rousseeuw’s minimum volume ellipsoid estimator. *Ann. Statist.* **20** 1828–1843.
- DI BUCCHIANICO, A., EINMAHL, J. H. J. and MUSHKUDIANI, N. A. (2001). Smallest nonparametric tolerance regions. *Annals of Statistics* 1320–1343.
- DYKSTRA, R. and CAROLAN, C. (1999). The distribution of the argmax of two-sided Brownian motion with quadratic drift. *J. Statist. Comput. and Simul.* **63** 47–58.
- EINMAHL, J. H. J. and KHMALADZE, E. V. (2011). Central limit theorems for local empirical processes near boundaries of sets. *Bernoulli* **17** 545–561.
- GROENEBOOM, P. (1985). Brownian motion with a parabolic drift and airy functions. *Probab. Th. Rel. Fields* **81** 79–109.
- HARTIGAN, J. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267–270.
- KHMALADZE, E. V. (2007). Differentiation of sets in measure. *J. Math. Anal. Appl.* **334** 1055–1072.
- KHMALADZE, E. V. and WEIL, W. (2008). Local empirical processes near convex bodies. *Ann. Inst. Statist. Math.* **60** 813–842.
- KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.
- MAMMEN, E. and POLONIK, W. (2013). Confidence regions for level sets. *Journal of Multivariate Analysis* **122** 202–214.
- MASON, D. M. and POLONIK, W. (2009). Asymptotic normality of plug-in level set estimates. *The Annals of Applied Probability* **19** 1108–1142.
- MÜLLER, D. W. (1992). *The Excess Mass Approach in Statistics*. Beiträge zur Statistik, University of Heidelberg.
- MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746.
- NOLAN, D. (1991). The excess mass ellipsoid. *J. Multivariate Anal.* **39** 348–371.
- POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters - an excess mass approach. *Ann. Statist.* **23** 855–881.
- POLONIK, W. (1997). Minimum volume sets and generalized quantile processes. *Stoch. Processes Appl.* **69** 1–24.
- POLONIK, W. (1998). The silhouette, concentration functions and ML-density estimation under order restrictions. *Ann. Statist.* **26** 1857–1877.
- QIAO, W. and POLONIK, W. (2019). Nonparametric confidence regions for level sets: Statistical properties and geometry. *Electronic Journal of Statistics* **13** 985–1030.
- RODRÍGUEZ-CASAL, A. and SAAVEDRA-NIEVES, P. (2019). Minimax Hausdorff estimation of density level sets. Preprint arXiv:1905.02897.
- ROUSSEEUW, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications* 283–297. Reidel, Dordrecht.
- SCHNEIDER, R. (1993). *Convex Bodies : The Brunn-Minkowski Theory*. Cambridge Univ. Press, Cambridge.

- SCHNEIDER, R. and WEIL, W. (2008). *Stochastic and Integral Geometry*. Springer, New-York.
- SCOTT, C. and NOVAK, R. (2006). Learning minimum volume sets. *J. Machine Learning Res.* **7** 665–704.
- SILVERMAN, B. W. and TITTERINGTON, D. M. (1980). Minimum covering ellipses. *SIAM J. Sci. and Stat. Comput.* **1** 401–409.
- TSYBAKOV, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes, With Applications to Statistics*. Springer, New York.
- XU, M. and SAMWORTH, R. J. (2019). High-dimensional nonparametric density estimation via symmetry and shape constraints. Preprint arXiv:1903.06092.