

## Cube Root Weak Convergence of Empirical Estimators of a Density Level Set

Berthet, Philippe; Einmahl, John

*Document version:*

Early version, also known as pre-print

*Publication date:*

2020

[Link to publication](#)

*Citation for published version (APA):*

Berthet, P., & Einmahl, J. (2020). *Cube Root Weak Convergence of Empirical Estimators of a Density Level Set*. (CentER Discussion Paper; Vol. 2020-015). CentER, Center for Economic Research.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2020-015

**CUBE ROOT WEAK CONVERGENCE OF EMPIRICAL  
ESTIMATORS OF A DENSITY LEVEL SET**

By

Philippe Berthet, John H.J. Einmahl

3 June 2020

ISSN 0924-7815  
ISSN 2213-9532

# 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.

**MSC 2010 subject classifications:** Primary 62G05, 62G20; secondary 60F05, 60F17

**JEL codes:** C13, C14.

**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 estimated finite-dimensional parameters.

**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 et al. (2011), and Chen et al. (2017). Statistical/machine learning approaches to the aforementioned criteria, include Cléménçon et al. (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. It will be shown that the choice does not matter since they 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  $n^{1/3}$ , 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 one only, where the sets are intervals which can be represented by two numbers, like in the estimation of the shorth. Hence those central limit theorems can be stated in the usual way, see, e.g., [Kim and Pollard \(1990\)](#).

**A local empirical process approach.** 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. Then 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  $L_\lambda$ . The relevant sets of  $\mathcal{A}$  have to be close to  $L_\lambda$  in Hausdorff distance at scale  $n^{-1/3}$ . Interestingly, the local nature of the convergence makes both the rate dimension-free and the Wiener process, appearing in the limit, distribution-free.

**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 and condition $H_1$

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  $\tau_\varepsilon$  at magnitude  $\varepsilon > 0$  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  $s(x) = \text{sgn}(x - \Pi(x))\|x - \Pi(x)\|$  the signed distance between  $x$  and  $\Pi(x)$ .

**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

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

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  $L_\lambda$ ,

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

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  $\tau_\varepsilon^{-1}(B) \triangle (L_\lambda \setminus (L_\lambda^* \cup S_\lambda^*))$ . For  $A \in \mathcal{A}$ , we then have  $\varphi_\varepsilon(\tau_\varepsilon(A \triangle L_\lambda)) = A$ .

**The limiting class  $\mathcal{B}$ .** We need to magnify with  $\varepsilon = n^{-1/3}$ . 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.) Let us further 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.** Define  $d_n(A, A') = (n^{1/3}P(A \Delta A'))^{1/2}$ . For  $c > 0$ , let  $[\mathcal{A}]_{c,n}$  and  $[\mathcal{B}]_{c,n}$  be the usual bracketing numbers w.r.t.  $d_n$  of  $\mathcal{C}^{cn^{-1/3}}$  and  $\{\tau_{n^{-1/3}}^{-1}(B) : B \in \mathcal{B}_c\}$ , respectively; see Einmahl and Khmaladze (2011). We assume either that for any  $c > 0$  we have

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

$$n^{1/2} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| = O_{\mathbb{P}}(1), \quad n \rightarrow \infty, \quad (2.8)$$

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \int_0^\delta \sqrt{\log [\mathcal{B}]_{c,n}(\varepsilon)} d\varepsilon = 0, \quad (2.9)$$

or that

$$\mathcal{A} \text{ and } \mathcal{B} \text{ are Vapnik-Chervonenkis (VC) classes.} \quad (2.10)$$

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

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

$$A \subset A_r, \quad \mu(A_r) = \mu(A) + r. \quad (2.11)$$

Let us denote by  $H_1$  the conditions in Subsection 2.1, in particular compactness of  $(\mathcal{B}_c, d)$ , (2.6), “Donsker classes”, and (2.11).

## 2.2. Condition $H_2$

We will need conditions on the behavior of  $f$  near  $S_\lambda$ . Let  $H_2$  denote the conditions (2.12)–(2.16) below.

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 risk of excess mass.** Consider the excess risk of 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  $\delta > 0$ ,

$$\inf_{A \in \mathcal{A}: d_H(L_\lambda, A) \geq \delta} \int_{L_\lambda \Delta A} |f(x) - \lambda| d\mu(x) > 0. \quad (2.12)$$

**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

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon^2} \int_{S_\lambda^\varepsilon \setminus L_\lambda} |f(x) - \lambda + s(x)f'_+(\Pi(x))| d\mu(x) = 0, \quad (2.13)$$

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

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 risk of 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

$$\int_{L_\lambda \Delta A} |s(x)|g(\Pi(x))d\mu(x) \geq \eta_0 d_H^2(L_\lambda, A). \quad (2.15)$$

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

$$D(B) \geq \eta_0 c^2(B), \quad (2.16)$$

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

### 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

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

The intrinsic, standard deviation metric of  $W$  on  $\mathcal{B}$  is defined to be  $(Var(W(B) - W(B')))^{1/2} = d(B, B')$ . 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  $\eta_1 > 0$ ,

$$\mathbb{E} \left( \max_{B \in \mathcal{B}_c} W^2(B) \right) < \eta_1 c, \quad \text{for all } c > 0. \quad (3.2)$$

**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  $n \rightarrow \infty$ ,*

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

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).

### 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  $\mathcal{B}_{c,n}^v, \mathcal{B}_{c,n}^p$ . We assume

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

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

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” subset

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

with the properties that  $A_s \subset A_{s'}$  for  $s < s'$ ,  $A_0 = L_\lambda$ , 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  $n \rightarrow \infty$ ,

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

**Theorem 3.2.** *Assume that  $H_1$ ,  $H_2$ , (3.2), (3.4) - (3.8) hold. Then on some probability space there exists a triangular array  $X_{n,1}, \dots, X_{n,n}$ ,  $n \in \mathbb{N}$ , of row-wise 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 argmin  $L_{2,n}$  of (1.2) and every argmax  $L_{3,n}$  of (1.3) satisfy, for  $j = 2, 3$ , as  $n \rightarrow \infty$ ,*

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

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  $W_n$ . It is beyond the scope of this paper to study  $Z(\mathcal{B})$  and  $Z(\mathcal{B}^*)$  in more detail, to see which argmax is “closer” to, say,  $\Sigma_0$  (corresponding to  $L_\lambda$ ), that is, which estimator performs better. However, a small simulation study for one-dimensional data shows that in that case  $L_{2,n}$  and  $L_{3,n}$  asymptotically outperform  $L_{1,n}$ .

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. Discussion and examples

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 (2.13) and (2.14) 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.

**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. More, relevant details about this class of sets for the bivariate case when  $L_\lambda$  is the unit disc are given in Example 1a in Einmahl and Khmaladze (2011). In particular  $\mathcal{B}_c$  is determined therein. A more restricted class is the class of all closed balls.

**Convex polytopes.** Another natural choice for  $\mathcal{A}$  is the class of all closed, convex polytopes. 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, we can let  $\mathcal{A}$  be the class of all closed, convex sets. Since this class is much larger than those in the previous examples, the restriction on  $f$  that  $L_\lambda$  is a convex body is much weaker now. For this and the previous example, see again Einmahl and Khmaladze (2011), Example 2, for more details; in particular  $\mathcal{B}_c$  is determined therein in case  $L_\lambda$  is the unit square.

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  $Nor(L_\lambda)$ , see [Schneider \(1993\)](#) and [Schneider and Weil \(2008\)](#). These finite measures carry the geometrical information about  $L_\lambda$ . 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 (2.13)-(2.14) 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 (2.13) - (2.14) hold, then

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

*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

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

Thus by (2.2),  $\Theta_{d-1}(g^+) = M(B_n^+)$  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 (2.13) and (2.14), 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

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

**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 (2.13) - (2.14) 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 (2.13) - (2.14) 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 (2.13)-(2.14) 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  $A \triangle L_\lambda \subset S_\lambda^\varepsilon$ .*

*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  $x \in S_\lambda^\varepsilon$ .  $\square$

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

$$L_{4,n} \in \arg \max_{A \in \mathcal{A}} \{P_n(A) : \mu(A) = v_\lambda\}. \quad (4.14)$$

**Lemma 4.6.** *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  $j = 1$  and for  $j = 2, 4$ , and  $n \geq n_0$ ,*

$$\mathbb{P}(d_H(L_{j,n}, L_\lambda) \geq c\varepsilon_n) \leq \delta.$$

*Proof.* Consider

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

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

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

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

We obtain from (2.12) 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  $c > 1$ ,

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

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 (2.15) we obtain for large  $n$ ,

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

Hence for large  $n$

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

This yields

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

where

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

Now, very similar as in the proof of Theorem 2 in Einmahl and Khmaladze (2011), we obtain, using Lemma 4.5, that the latter probability is bounded by

$$c_1 \exp(-c_2 \eta_0^2 c^2),$$

for some constants  $c_1, c_2 > 0$ .

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

$$\begin{aligned} &\mathbb{P}(d_H(L_{1,n}, L_\lambda) \geq c\varepsilon_n) \\ &\leq \mathbb{P}(d_H(L_{1,n}, L_\lambda) \geq \delta) + \sum_{m=0}^{\infty} \mathbb{P}(c^{2^m} \varepsilon_n \leq d_H(L_{1,n}, L_\lambda) \leq \min(c^{2^{m+1}} \varepsilon_n, \delta)) \\ &\leq \frac{1}{2}\delta + c_1 \sum_{m=0}^{\infty} \exp(-c_2 \eta_0^2 c^{2^{m+1}}) \leq \delta, \end{aligned}$$

if  $c$  is large enough.

In case (2.7) and (2.8) hold, the proof for  $L_{1,n}$  follows the same lines, but now the arguments in the proof of Theorem 1 in Einmahl and Khmaladze (2011)

should be used, in particular the application of Lemma 19.34 in [van der Vaart \(1998\)](#).

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 4.7.** *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  $n \rightarrow \infty$ ,*

$$\sup_{B_n \in \mathcal{B}_{c,n}, B \in \mathcal{B}_c, d(B_n, B) \leq \gamma_{c,n}} |w_n(B_n) - w_n(B)| \xrightarrow{\mathbb{P}} 0, \quad (4.15)$$

and, with probability 1,

$$\sup_{B \in \mathcal{B}_c} |w_n(B) - \sqrt{\lambda}W(B)| \rightarrow 0. \quad (4.16)$$

*Proof.* Note that the assumptions of Theorems 1 and 2 in [Einmahl and Khmaladze \(2011\)](#) are satisfied. Hence, using these theorems, including a Skorohod construction as on page 554 therein, yields (4.15) and (4.16). 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  $\check{\mathcal{B}}$  of  $\mathcal{B}$ , define

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

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 4.8.** *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 (2.16),

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

which is by (3.2), bounded from above by

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

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  $\mathbb{P}(\cup_{c>0} \Omega_c) = 1$ .  $\square$

#### 4.4. Proof of Theorem 3.1

We work in the setting of Lemma 4.7. 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 (4.16) in Lemma 4.7. Clearly, Lemmas 4.5, 4.6 and 4.8 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

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

and observe that Lemmas 4.7, 4.3 and 4.4 imply

$$m_{c,n} \xrightarrow{\mathbb{P}} m_c, \text{ as } n \rightarrow \infty. \quad (4.17)$$

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 (4.17)

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

which by (4.15) 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 (4.16) and then by Lemma 2.6 in Kim and Pollard (1990)

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

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 (4.14) 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 (2.11), 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

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

Again, Lemmas 4.6 and 4.8 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 4.7 given in Einmahl and Khmaladze (2011)), and Lemma 4.3, for  $\varepsilon > 0$ ,

$$\mathbb{P}(m_{c,n}^p \leq \hat{m}_{c,n} + \varepsilon) \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (4.18)$$

and that by (3.5), (4.12) and Lemma 4.7 (possibly with a larger  $\gamma_{c,n} \rightarrow 0$ ),

$$m_{c,n}^p \xrightarrow{\mathbb{P}} m_c^*, \quad \text{as } n \rightarrow \infty. \quad (4.19)$$

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  $n$

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

which is by (4.18)

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

which is by (4.19)

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

which by (4.15), 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 (4.16) and then by again Lemma 2.6 in Kim and Pollard (1990)

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

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

## References

- BERTHET, P. and EL-NOUTY, C. (2006). Almost sure asymptotic behaviour of the shorth estimators. *Math. Methods Statist.* **14** 379–403.
- 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.
- 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.
- 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.
- 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.