

# Central limit theorems for local empirical processes near boundaries of sets

JOHN H.J. EINMAHL<sup>1</sup> and ESTÁTE V. KHMALADZE<sup>2</sup>

<sup>1</sup>*Department of Econometrics and OR and CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: j.h.j.einmahl@uvt.nl*

<sup>2</sup>*School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand. E-mail: estate.khmaladze@vuw.ac.nz*

We define the local empirical process, based on  $n$  i.i.d. random vectors in dimension  $d$ , in the neighborhood of the boundary of a fixed set. Under natural conditions on the shrinking neighborhood, we show that, for these local empirical processes, indexed by classes of sets that vary with  $n$  and satisfy certain conditions, an appropriately defined uniform central limit theorem holds. The concept of differentiation of sets in measure is very convenient for developing the results. Some examples and statistical applications are also presented.

*Keywords:* convex body; differentiation of sets; Gaussian behavior; local empirical process; set boundary; weak convergence

## 1. Introduction

Let  $X_1, \dots, X_n$  be independent and identically distributed (i.i.d.) random vectors in  $\mathbb{R}^d$  ( $d \in \mathbb{N}$ ), distributed according to an absolutely continuous probability measure  $P$ . Denote the corresponding density by  $p$ . For a Borel measurable subset  $D$  of  $\mathbb{R}^d$ , write

$$\Psi_n(D) = \sum_{i=1}^n 1_D(X_i).$$

The process  $\Psi_n(D)$ ,  $D \in \mathcal{D}$  ( $\mathcal{D}$  being the class of Borel sets), is, by definition, a binomial process on  $\mathbb{R}^d$ ;  $\Psi_n/n$  is the empirical measure corresponding to  $X_1, \dots, X_n$ . Clearly,  $\mathbb{E}\Psi_n(D) = nP(D)$ .

Let  $K$  be a convex body in  $\mathbb{R}^d$ . The set  $K$  will be *fixed* throughout. Denote its boundary by  $\partial K$ . It is the aim of this paper to study the behavior of  $\Psi_n$  in the neighborhood of  $\partial K$ . Write  $\|z - \partial K\| = \min_{x \in \partial K} \|z - x\|$  and let

$$\mathcal{V}_\varepsilon(\partial K) = \{z \in \mathbb{R}^d : \|z - \partial K\| \leq \varepsilon\}, \quad \varepsilon > 0,$$

denote this neighborhood. Set  $a = P(\mathcal{V}_\varepsilon(\partial K))$ . For a Borel set  $A \subset \mathcal{V}_\varepsilon(\partial K)$ , define

$$z_n(A) = \frac{1}{\sqrt{na}} [\Psi_n(A) - nP(A)].$$

If  $\varepsilon \rightarrow 0$ , then all sets  $A$  will “shrink toward”  $\partial K$ . If, however,  $n \rightarrow \infty$  at the same time, the random variables  $z_n(A)$  do not have to converge to 0 and, if  $n\varepsilon \rightarrow \infty$ , they should typically

converge to Gaussian random variables. However, where would these Gaussian random variables “live”? Would they form some set-parametric process? These are the questions we seek to address in this paper. We will do this using the concept of differentiability of set-valued functions, as was recently developed in [24].

One cannot prove a sufficiently interesting Gaussian limit theorem for  $z_n$  which is defined on all Borel subsets of  $\mathcal{V}_\varepsilon(\partial K)$ . Instead, one needs to consider smaller classes of sets. Let  $\varepsilon = \varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$  and let  $\mathcal{A}_{\varepsilon_n}$  be a class of measurable subsets of  $\mathcal{V}_{\varepsilon_n}(\partial K)$ . The *canonical example* of  $\mathcal{A}_{\varepsilon_n}$  is constructed as follows. Let  $\mathcal{K}$  be a fixed class of Borel sets in  $\mathbb{R}^d$  and define  $\mathcal{A} = \{K' \Delta K : K' \in \mathcal{K}\}$ , where  $\Delta$  denotes “symmetric difference”. Now, take  $\mathcal{A}_{\varepsilon_n} = \{A \in \mathcal{A} : A \subset \mathcal{V}_{\varepsilon_n}(\partial K)\}$ . Particular cases can be found in Examples 1 and 2 below.

Our main result is the central limit theorem for the *local empirical process near  $\partial K$  and indexed by  $\mathcal{A}_{\varepsilon_n}$* ,

$$\{z_n(A), A \in \mathcal{A}_{\varepsilon_n}\}. \tag{1}$$

Denoting the conditional probability distribution on  $\mathcal{V}_\varepsilon(\partial K)$  by  $P_\varepsilon(A) = P(A)/a$ , we can also write

$$z_n(A) = \frac{1}{\sqrt{na}}[\Psi_n(A) - naP_\varepsilon(A)].$$

This reflects the fact that, on average, the effective sample size is equal to  $na$ , not  $n$ . We therefore assume, in addition to  $\varepsilon_n \rightarrow 0$ , that

$$n\varepsilon_n \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

This will imply that  $na \rightarrow \infty$  and ensure that the sets in  $\mathcal{A}_{\varepsilon_n}$  contain enough observations to obtain Gaussian limit behavior.

Although very natural here, it is, in general, unusual that an empirical process is defined on a class of sets that depends on  $n$ . We will show that its limiting process should be defined on a class of subsets not of the “same”  $\mathbb{R}^d$ , but of the cylinder  $\partial K \times [-1, 1]$ . The subsets in this class are properly defined derivatives of sequences of sets, with the  $n$ th set an element of  $\mathcal{A}_{\varepsilon_n}$ .

Poisson limit behavior of  $\Psi_n$  on  $\mathcal{V}_{\varepsilon_n}(\partial K)$  has been studied in [26]. The main limit result there had a somewhat unusual property: it contained a functional limit theorem, but not a one-dimensional limit theorem for  $\Psi_n$ . Indeed, although it showed weak convergence and, moreover, convergence in total variation, for the process  $\Psi_n$  given on *all* Borel subsets of  $\mathcal{V}_{\varepsilon_n}(\partial K)$ , for a particular sequence of subsets  $\mathcal{A}_{\varepsilon_n}$ , it remained unspecified which random variable from the limiting process the sequence  $\Psi_n(A_{\varepsilon_n})$  would converge to. This happened because the notion of derivative sets had not been developed at the time Khmaladze and Weil [26] was accepted for publication. In this paper, the situation is different – extracting the one-dimensional limit theorem from Theorem 1 gives the following statement: *if the set-valued function  $A_\varepsilon$  is differentiable in  $\varepsilon$  at  $\varepsilon = 0$  and  $dA_\varepsilon/d\varepsilon$  is its derivative (see Section 3 or [24]), then*

$$z_n(A_{\varepsilon_n}) \xrightarrow{d} W(dA_\varepsilon/d\varepsilon),$$

where  $W$  is the set-parametric Brownian motion defined just before Lemma 1, Section 4.

The local empirical process for one-dimensional  $X_i$ , that is, the empirical process in the neighborhood of a point  $c \in \mathbb{R} \cup \{-\infty, \infty\}$ , is a classical object in probability theory, one which has proven to be very valuable in statistics; see, for example, [5,9,10,12,14,23,29], the book by Csörgő and Horváth [5] and Khmaladze [23]. The one-dimensional local empirical process has been extended to the multivariate set-up, but, typically, only the neighborhood of a point  $c \in \mathbb{R}^d$  or the region outside a large sphere are considered; see, for example, [8,11,13,30,34]. Perhaps the closest to the present paper are [18] and [15]. For a local empirical process for function-valued random elements, see [16].

The paper is organized as follows. In the next section, we present statistical applications. In Section 3, we introduce the necessary geometry and the appropriate concept of differentiation of sets. In Section 4, the main results, central limit theorems for  $z_n$ , and some examples will be presented. Proofs are collected in Section 5.

## 2. Statistical motivation

Although the local empirical process near the boundary of a set is an interesting probabilistic object in its own right, the study of this type of process was mostly motivated by problems in spatial statistics. Consider a family of distributions, indexed by some parameter  $\theta$ , and denote by  $L_n(\theta, \theta')$  the log-likelihood ratio. If the parameter were a vector, as in parametric problems (see, e.g., [22]), the local analysis of  $L_n(\theta, \theta')$  (or any other process which the inference is based upon), in  $\theta'$  from the neighborhood of the true value  $\theta$ , is a crucial step in asymptotic statistical theory. It forms, for example, the basis of contiguity theory. The situation is similar when the parameter is a function (see, e.g., [4,40]). However, it has thus far not been known how to carry out such a local analysis when the parameter is a set.

Examples of set-parametric problems are provided by the class of spatial change point problems or *change set problems* (see, e.g., [25]). In these problems, the observation is usually a (marked) point process in  $\mathbb{R}^d$  and the model assumption is that there is a set, or an image,  $K$ , such that outside  $K$ , the distribution of the point process (e.g., the distribution of the marks) sharply changes. One can think of  $K$  as, for example, an ore deposit site, a pollution site or a site with different magnetic properties. The literature on this problem is very broad; see, for example, [6,19,28,32].

In most of the particular formulations of the change set problem, the log-likelihood ratio  $L_n(K, K')$  is some form or another of the local empirical process (1), where  $K$  plays the role of the true value of the change set, while the sets  $K'$  are small deviations from it. To be more precise, let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be independent random vectors, with  $X_i$  being (as before) a  $d$ -dimensional location and  $Y_i$  being a “mark”, not necessarily one-dimensional. Write  $K(\varepsilon)$  instead of  $K'$  in order to explicitly express the dependence on  $\varepsilon > 0$ . Let  $P_1$  and  $P_2$  be the distributions of  $Y_i$  outside  $K$  and on  $K$ , respectively. The log-likelihood ratio then has the form

$$L_n(K, K(\varepsilon)) = \sum_{i=1}^n [1_{K(\varepsilon) \setminus K}(X_i) - 1_{K \setminus K(\varepsilon)}(X_i)] \xi(Y_i),$$

where  $\xi(Y_i) = \log \frac{dP_2}{dP_1}(Y_i)$ . We focus on

$$\sum_{i=1}^n 1_{K(\varepsilon) \setminus K}(X_i) \quad \text{and} \quad \sum_{i=1}^n 1_{K \setminus K(\varepsilon)}(X_i); \tag{2}$$

a discussion of the behavior of the  $\xi(Y_i)$  is of secondary importance here. Let  $\tilde{\mathcal{K}}$  be a class of set-valued functions  $K(\cdot)$  all converging to the same  $K$ , that is,  $K(\varepsilon) \Delta K$  shrinks toward the boundary  $\partial K$  when  $\varepsilon \rightarrow 0$ . These sets describe the deviations from the hypothetical change set  $K$ . Let us consider the processes in (2) given on these deviations and investigate their joint limit in distribution when  $n \rightarrow \infty$  and  $\varepsilon = \varepsilon_n \rightarrow 0$ : the larger the number of observations, the smaller, or narrower, the sets we consider. In the appropriate formulation of “local alternatives”, where not only  $K(\varepsilon)$  tends to  $K$ , but also  $P_2$  tends to  $P_1$ , when  $n \rightarrow \infty$ , the convergence to a Gaussian limiting process – as shown in this paper – is of fundamental importance. Indeed, although there is a rich literature on statistical estimation of sets, we know very few results on testing hypotheses about sets and no results for testing against local alternatives.

To illustrate another class of statistical problems where the parameter is a set, consider two prominent examples: the excess mass approach (cf. [31,33]), and the shorth [2,20]) and its generalization [17]. Let  $\mathcal{K}$  be a fixed subset of  $\mathcal{D}$ , as in the canonical example. One could choose, for instance,  $\mathcal{K}$  to be the class of all ellipsoids. Define the excess mass set  $K$  for level  $\lambda > 0$  by

$$K = \arg \max_{K' \in \mathcal{K}} \{P(K') - \lambda \mu_d(K')\},$$

where  $\mu_d$  denotes  $d$ -dimensional Lebesgue measure. Similarly, the generalized shorth or minimum volume set  $K$  for probability  $\alpha \in (0, 1)$  is given by

$$K = \arg \min_{K' \in \mathcal{K}} \{\mu_d(K') : P(K') \geq \alpha\};$$

see [7,35] when  $\mathcal{K}$  is the class of all ellipsoids. It turns out that both of these sets  $K$  and their M-estimators can be analyzed somewhat similarly. Therefore, we confine ourselves to the excess mass set.

The obvious non-parametric estimator for this set is obtained by replacing  $P$  by  $\Psi_n/n$ :

$$\begin{aligned} K_n &= \arg \max_{K' \in \mathcal{K}} \{\Psi_n(K')/n - \lambda \mu_d(K')\} \\ &= \arg \max_{K' \in \mathcal{K}} n^{2/3} \{ \Psi_n(K')/n - \lambda \mu_d(K') - (\Psi_n(K)/n - \lambda \mu_d(K)) \}. \end{aligned}$$

Let  $\varepsilon_n$  be such that  $a = n^{-1/3}$  (cf. the “cube root asymptotics” of [27]). Under certain conditions, it can be shown that for large  $T > 0$ , with high probability,  $K_n \Delta K \in \mathcal{A}_{T\varepsilon_n}$ . Observe that in that case,

$$K_n = \arg \max_{K' : K' \Delta K \in \mathcal{A}_{T\varepsilon_n}} \{z_n(K') - z_n(K) + n^{2/3} [P(K') - P(K) - \lambda(\mu_d(K') - \mu_d(K))]\}.$$

Now, a central limit theorem for  $z_n$  makes it possible to show that, asymptotically,  $K_n \Delta K$  can be described as a deterministic function depending on  $n$  (actually, on  $\tau_{\varepsilon_n}^{-1}$ ; see Section 4) evaluated at a random variable that does not depend on  $n$ . This random variable is the arg max of some Brownian motion with drift. Such a result is very useful for a refined analysis of  $K_n$ . See [3] for a study of the behavior of such  $K_n$ 's along these lines.

### 3. Some geometry and differentiability of sets

In this section, we first briefly review some relevant notation and facts from geometry. We then recall the concept of “differentiation of sets in measure”, as given in [24]. In that paper and the references therein (in particular [36]), more details about the required geometry can be found. We also refer to the recent monograph [37].

Let  $K \in \mathcal{D}$  be our convex body, that is, a closed, bounded convex set that has interior points. Denote by  $\Pi(z)$  the metric projection of  $z \in \mathbb{R}^d$  on  $\partial K$ , that is,  $\Pi(z)$  is a nearest point to  $z$  on  $\partial K$ . The set of  $z$ -values for which such a nearest point is not unique is a subset  $S_K$  of  $K$  called the *skeleton* of  $K$ . Let  $\mu_d$  denote  $d$ -dimensional Lebesgue measure. It is then known that  $\mu_d(S_K) = 0$ . A unit vector  $u$  is called an *outer normal* of  $K$  at  $x \in \partial K$  if there is some  $z \in \mathbb{R}^d \setminus K$  such that  $x = \Pi(z)$  and  $u = (z - \Pi(z)) / \|z - \Pi(z)\|$ . Let  $B_r(z)$  denote the closed ball with center  $z$  and radius  $r$ . For  $x \in \partial K$ , we define the local interior reach

$$r(x) = \max\{r : x \in B_r(z) \subset K\}.$$

If  $r(x) > 0$ , then the outer normal  $u$  at  $x \in \partial K$  is unique. In this case, the unit vector  $-u$  is the unique inner normal. In general, at each  $x \in \partial K$ , we denote the set of outer normals by  $N(x)$  and the normal bundle of  $K$  is defined as

$$Nor(K) = \{(x, u) : x \in \partial K, u \in N(x)\}.$$

The cylinder  $\Sigma = Nor(K) \times [-1, 1]$  will be important for describing our limiting processes. Note, however, that it will eventually be possible to work with the cylinder  $\partial K \times [-1, 1]$ , which is much easier to visualize.

We also need the so-called local magnification map  $\tau_\varepsilon$ ; see [24]. Any point  $z \in \mathbb{R}^d \setminus S_K$  can be written as  $z = \Pi(z) + d_s(z)u$ , where  $d_s(z)$  is the signed (“+” outside) distance between  $z$  and  $\Pi(z)$  and  $u$  an outer normal at  $\Pi(z)$  that satisfies the equality. Now, define

$$\tau_\varepsilon(z) = \left( \Pi(z), u, \frac{d_s(z)}{\varepsilon} \right), \quad z \in \mathbb{R}^d \setminus S_K, \varepsilon > 0.$$

Observe that  $\tau_\varepsilon$  maps  $\mathcal{V}_\varepsilon(\partial K) \setminus S_K$  into  $\Sigma$ .

We are now prepared to introduce the aforementioned differentiation of sets. Consider the first support measure  $\vartheta_{d-1}$  on  $Nor(K)$ ; see [36]. It attributes measure 0 to the set of all points  $(x, u)$ , where, at  $x$ , there is more than one outer normal  $u$ . Hence, we can map it to  $\partial K$  in a one-to-one way. On  $\partial K$ , this map coincides with Hausdorff measure  $\nu$  and if, for a Borel set  $H \subset Nor(K)$ , we write

$$H_0 = \{x \in \partial K : (x, u) \in H\},$$

then

$$\vartheta_{d-1}(H) = \nu(H_0). \tag{3}$$

On  $\Sigma$ , define the measure  $M = \vartheta_{d-1} \times \mu$  ( $\mu$  being one-dimensional Lebesgue measure). Consider a (Borel) set-valued function  $K(\varepsilon)$ ,  $\varepsilon \in [0, 1]$ , such that  $K(0) = K$ , with  $K$  as before; write  $A(\varepsilon) = K(\varepsilon) \Delta K$  and assume that  $A(\varepsilon) \subset \mathcal{V}_\varepsilon(\partial K)$ . The set-valued function  $A(\varepsilon)$ ,  $\varepsilon \in [0, 1]$ , is called *differentiable* at  $\partial K$  and  $\varepsilon = 0$  if there exists a Borel set  $B \subset \Sigma$  such that  $M(\tau_\varepsilon A(\varepsilon) \Delta B) \rightarrow 0$  as  $\varepsilon \rightarrow 0$  (where  $\tau_\varepsilon A = \{\tau_\varepsilon(z) : z \in A\}$ ). The set  $B$  is called the *derivative* of  $A(\varepsilon)$  at  $\partial K$ . In this case, we also say that  $K(\varepsilon)$ ,  $\varepsilon \in [0, 1]$ , is differentiable with the same derivative and write

$$\left. \frac{d}{d\varepsilon} K(\varepsilon) \right|_{\varepsilon=0} = \left. \frac{d}{d\varepsilon} A(\varepsilon) \right|_{\varepsilon=0} = B.$$

Note that  $B$  is not unique, but can be changed on a set of  $M$ -measure 0.

Let  $P$  now be as in Section 1. We require that the density  $p$  can be approximated in the neighborhood of  $\partial K$  by a function depending only on  $\Pi(z)$  and on whether or not  $z \in K$ . This latter possibility is easy to imagine in the change set problems: the limit of  $p(z)$  from inside  $K$  can indeed be different from that from outside if  $K$  is the change set. More formally, we require the existence of two functions,  $p_+$  and  $p_-$ , on  $\partial K$  such that, as  $\varepsilon \rightarrow 0$ ,

$$\frac{1}{\varepsilon} \int_{\mathcal{V}_\varepsilon(\partial K) \setminus K} |p(z) - p_+(\Pi(z))| d\mu_d(z) \rightarrow 0, \tag{4}$$

$$\frac{1}{\varepsilon} \int_{\mathcal{V}_\varepsilon(\partial K) \cap K} |p(z) - p_-(\Pi(z))| d\mu_d(z) \rightarrow 0. \tag{5}$$

Now, define a measure  $M_p$  on  $\Sigma$  as follows:

$$\begin{aligned} dM_p(x, u, s) &= p_+(x) d\vartheta_{d-1}(x, u) \times ds && \text{for } s > 0, \\ dM_p(x, u, s) &= p_-(x) d\vartheta_{d-1}(x, u) \times ds && \text{for } s \leq 0. \end{aligned}$$

For convenience, assume that  $p_+$  and  $p_-$  are bounded (although a weaker, integrability, condition would suffice). An easy, but practically interesting, situation occurs when  $p_+(x) = c_+$  and  $p_-(x) = c_-$  for all  $x \in \partial K$ , where  $c_+, c_- \geq 0$  are two constants.

The following key result from [24] shows the “differentiability of sets in measure”: if  $A(\varepsilon)$  is differentiable at  $\partial K$ , then

$$\left. \frac{d}{d\varepsilon} P(A(\varepsilon)) \right|_{\varepsilon=0} = M_p \left( \left. \frac{d}{d\varepsilon} A(\varepsilon) \right|_{\varepsilon=0} \right). \tag{6}$$

### 4. Main results

Let  $\mathcal{A}_{\varepsilon_n}$  be as in Section 1 and assume  $M_p(\Sigma) > 0$ . Writing  $a_n = P(\mathcal{V}_{\varepsilon_n}(\partial K))$ , it easily follows, using (6), that  $a_n/\varepsilon_n \rightarrow M_p(\Sigma)$ . Hence, we have, just as for  $\varepsilon_n$ ,

$$a_n \rightarrow 0 \quad \text{and} \quad na_n \rightarrow \infty.$$

Denote by  $\mathcal{B}$  the class of all possible derivatives at  $\varepsilon = 0$  corresponding to  $\mathcal{A}_{\varepsilon_n}$ , which, by definition, means that  $B \in \mathcal{B}$  if and only if there exists a sequence of sets  $(A_n)_{n=1}^\infty$  with  $A_n \in \mathcal{A}_{\varepsilon_n}$  and  $M(\tau_{\varepsilon_n} A_n \Delta B) \rightarrow 0$ . (Observe that for a thus converging sequence of Borel subsets of  $\Sigma$ , the limit set is not well defined. This limit “set” is actually an equivalence class of sets, defined by the property that for any two sets  $B_1, B_2$  in the class,  $M(B_1 \Delta B_2) = 0$ . Out of every such an equivalence class, we choose one limit (Borel) set  $B$ , say. Whether or not the conditions of our results are satisfied will depend on the choices of these  $B$ ’s. In applications/examples, we should choose natural or appropriate  $B$ ’s to make the theorems work.)

Consider the local empirical process (1) from Section 1. To establish its limit in distribution, we need the following steps. Write  $\tau_\varepsilon^{-1}C = \{z \in \mathcal{V}_\varepsilon(\partial K) : \tau_\varepsilon(z) \in C\}$  for a Borel set  $C \subset \Sigma$ . First, using the local magnification map, induce the point process  $\Phi_n$  and the distribution  $Q_n$  on  $\Sigma : \Phi_n(C) = \Psi_n(\tau_{\varepsilon_n}^{-1}C)$  and  $Q_n(C) = P_{\varepsilon_n}(\tau_{\varepsilon_n}^{-1}C)$ . Thus, for any Borel set  $C \subset \Sigma$ , we can define

$$\begin{aligned} v_n(C) &:= \frac{1}{\sqrt{na_n}}[\Phi_n(C) - na_n Q_n(C)] \\ &= \frac{1}{\sqrt{na_n}}[\Psi_n(\tau_{\varepsilon_n}^{-1}C) - na_n P_{\varepsilon_n}(\tau_{\varepsilon_n}^{-1}C)] = z_n(\tau_{\varepsilon_n}^{-1}C). \end{aligned} \tag{7}$$

Hence, we can define the local empirical processes on two classes of sets: on  $\mathcal{B}_n := \{\tau_{\varepsilon_n} A : A \in \mathcal{A}_{\varepsilon_n}\}$ , which changes with  $n$ , and on  $\mathcal{B}$ , the class of its limits, or derivative sets, which is fixed. We denote these processes by

$$v_{n, \mathcal{B}_n} := \{v_n(B) : B \in \mathcal{B}_n\} \quad \text{and} \quad v_{n, \mathcal{B}} := \{v_n(B) : B \in \mathcal{B}\}.$$

(Local empirical processes of the latter type – i.e. for a fixed  $\mathcal{B}$  – have been studied in, e.g., [15,18]. Here, however, our main object is  $v_{n, \mathcal{B}_n}$ ;  $v_{n, \mathcal{B}}$  is an auxiliary process, a bridge between  $v_{n, \mathcal{B}_n}$  and its limiting process.) Second, we show that the distribution  $Q_n$ , which “governs” these processes, converges to the distribution  $Q(C) = M_p(C)/M_p(\Sigma)$  and the processes  $v_{n, \mathcal{B}_n}$  can be approximated by the processes  $v_{n, \mathcal{B}}$ . Next, we verify that  $v_{n, \mathcal{B}}$  converges in distribution to a set-parametric Brownian motion  $W_{\mathcal{B}}$  and, finally, we note that one can switch from  $W_{\mathcal{B}}$ , given on subsets of  $\Sigma$ , to its isometric image, given on the “easier” cylinder  $\partial K \times [-1, 1]$ .

Below, we write  $C_- = \{(x, u, s) \in C : (x, u) \in \text{Nor}(K), s \leq 0\}$  and  $C_+ = C \setminus C_-$ . Note that in the case  $p_+(x) = c_+$  and  $p_-(x) = c_-$  for all  $x \in \partial K$ , we have

$$Q(C) = \frac{c_+ M(C_+) + c_- M(C_-)}{(c_+ + c_-)v(\partial K)}.$$

When, for example,  $0 = c_- < c_+$ , we obtain  $Q(C) = Q(C_+) = M(C_+)/v(\partial K)$ .

For Borel sets  $C, C' \subset \Sigma$ , define  $d(C, C') = (Q(C \Delta C'))^{1/2}$ . Throughout, we will assume that  $(\mathcal{B}, d)$  is totally bounded and that

$$\sup_{A \in \mathcal{A}_{\varepsilon_n}} \inf_{B \in \mathcal{B}} d(\tau_{\varepsilon_n} A, B) \rightarrow 0. \tag{8}$$

In particular, every sequence  $(A_n)_{n=1}^\infty$  with  $A_n \in \mathcal{A}_{\varepsilon_n}$  has a subsequence  $(A_{n_k})_{k=1}^\infty$  such that for some  $B \in \mathcal{B}$ ,  $d(\tau_{\varepsilon_{n_k}} A_{n_k}, B) \rightarrow 0$ . Assumption (8) can be written as

$$\sup_{B_n \in \mathcal{B}_n} \inf_{B \in \mathcal{B}} d(B_n, B) \rightarrow 0.$$

From the definition of  $\mathcal{B}$  and the assumption that  $(\mathcal{B}, d)$  is totally bounded, it follows that

$$\sup_{B \in \mathcal{B}} \inf_{B_n \in \mathcal{B}_n} d(B_n, B) \rightarrow 0.$$

Thus, the Hausdorff distance between the classes  $\mathcal{B}_n$  and  $\mathcal{B}$  tends to 0:

$$\gamma_n := \max \left( \sup_{B_n \in \mathcal{B}_n} \inf_{B \in \mathcal{B}} d(B_n, B), \sup_{B \in \mathcal{B}} \inf_{B_n \in \mathcal{B}_n} d(B_n, B) \right) \rightarrow 0. \tag{9}$$

Recall that it is the aim of this paper to present a central limit theorem for  $z_{n, \mathcal{A}_{\varepsilon_n}}$ , or, equivalently,  $v_{n, \mathcal{B}_n}$ . By ‘‘central limit theorem for  $z_{n, \mathcal{A}_{\varepsilon_n}}$ ’’ we mean:

$$(a) \quad \sup_{B_n \in \mathcal{B}_n, B \in \mathcal{B}: d(B_n, B) \leq \gamma_n} |v_n(B_n) - v_n(B)| \xrightarrow{P} 0;$$

and

$$(b) \quad v_{n, \mathcal{B}} \xrightarrow{d} W_{\mathcal{B}} := \{W(B), B \in \mathcal{B}\}.$$

Here,  $W_{\mathcal{B}}$  is set-parametric Brownian motion: a bounded, uniformly  $d$ -continuous Gaussian process with mean 0 and covariance structure  $\mathbb{E}W(B)W(B') = Q(B \cap B')$ . We view  $v_n$  and  $W$  as processes taking values in  $\ell^\infty(\mathcal{B})$  endowed with the uniform distance and understand weak convergence in the sense of van der Vaart and Wellner [39]. (We assume, for convenience, that our classes of sets are such that the various ‘‘suprema’’ are measurable, i.e., that they are random variables.) The following fact is very useful for proving this central limit theorem.

**Lemma 1.** *From (4) and (5), it follows that  $Q_n$  converges to  $Q$  in total variation:*

$$\sup |Q_n(C) - Q(C)| \rightarrow 0,$$

with the sup taken over all Borel sets  $C \subset \Sigma$ .

Define  $d_n(A, A') := (P_{\varepsilon_n}(A \Delta A'))^{1/2} = (P(A \Delta A')/a_n)^{1/2}$ ; observe that  $d_n(A, A') = (Q_n(\tau_{\varepsilon_n} A \Delta \tau_{\varepsilon_n} A'))^{1/2}$ . Assume, for any  $\delta > 0$ , that there exists a finite collection of pairs (brackets)  $[\underline{A}(\delta), \overline{A}(\delta)]$  of Borel sets in  $\mathcal{V}_{\varepsilon_n}(\partial K)$  with  $d_n(\underline{A}(\delta), \overline{A}(\delta)) \leq \delta$ , such that any set  $A \in \mathcal{A}_{\varepsilon_n}$  can be placed in a bracket from this collection:  $\underline{A}(\delta) \subset A \subset \overline{A}(\delta)$ . Consider such a class of brackets with minimal cardinality; denote this cardinality (the bracketing number) by  $N_{[\cdot], n}(\delta)$  and let  $\mathcal{N}_{[\cdot], n}(\delta)$  be the set of  $\underline{A}(\delta)$ ’s in this class. We assume the same for  $\tau_{\varepsilon_n}^{-1} \mathcal{B} := \{\tau_{\varepsilon_n}^{-1} B : B \in \mathcal{B}\}$  and

use the notation  $\tilde{N}_{[\cdot],n}(\delta)$  and  $\tilde{N}_{[\cdot],n}(\delta)$ . We will require

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \int_0^\delta \sqrt{\log N_{[\cdot],n}(x)} \, dx = 0, \tag{10}$$

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \int_0^\delta \sqrt{\log \tilde{N}_{[\cdot],n}(x)} \, dx = 0. \tag{11}$$

**Theorem 1.** *Under the aforementioned assumptions, in particular, the growth conditions on  $\varepsilon_n$ , the approximation of  $p$  by  $p_+$  or  $p_-$  in (4) and (5), the relation between  $\mathcal{A}_{\varepsilon_n}$  and  $\mathcal{B}$  specified in (8) and the entropy conditions (10) and (11), the central limit theorem for  $z_{n,\mathcal{A}_{\varepsilon_n}}$  holds, that is, statements (a) and (b) hold true.*

We also present a version of Theorem 1 without assuming bracketing conditions. To be more precise, we will assume that our classes of sets near  $\partial K$  are Vapnik–Chervonenkis (VC) classes (see, e.g., [39], Section 2.6, for definition and properties).

**Theorem 2.** *Let  $\mathcal{A}_{\varepsilon_n}$  be a VC class with index  $t_n \leq t$  for some  $t \in \mathbb{N}$ ; also, assume that  $\mathcal{B}$  is a VC class. If we assume that  $\varepsilon_n \rightarrow 0$ ,  $n\varepsilon_n \rightarrow \infty$  and (4), (5) and (8), then the central limit theorem for  $z_{n,\mathcal{A}_{\varepsilon_n}}$  holds, that is, statements (a) and (b) hold true.*

**Remark 1.** Consider the canonical example of Section 1 and let  $\mathcal{K}$  be a VC class. Then  $\mathcal{A}$  is also a VC class, with index  $t$ , say. Since  $\mathcal{A}_{\varepsilon_n} \subset \mathcal{A}$ , the index  $t_n$  of  $\mathcal{A}_{\varepsilon_n}$  indeed satisfies  $t_n \leq t$ .

**Remark 2.** Similar to the discussions in [24,26], we note that Theorems 1 and 2, as well as the whole construction, can be carried over to the case where  $K$  is a finite union of convex bodies and, even more easily, to the case where  $K$  is closed and bounded and has a boundary of positive reach (intuitively,  $K$  has a “smooth” boundary). Indeed, the key objects, such as the local magnification map  $\tau_\varepsilon$  (uniquely defined almost everywhere on  $\mathbb{R}^d$ ), the local Steiner formula, the notion of derivative sets and Lemma 1, are all valid for such a  $K$ . Moreover, the existence of the local Steiner formula for a very general  $K$  has been demonstrated in [21]. This offers perspectives for considering such a general  $K$  in the statements of our results.

The limiting process  $W_{\mathcal{B}}$  is defined on subsets of the cylinder  $\Sigma = \text{Nor}(K) \times [-1, 1]$ . This cylinder is not easy to visualize. However, since the support measure  $\vartheta_{d-1}$  depends on  $H$  only through  $H_0$  (cf. (3)), we have a similar result for the measure  $Q$ . That is, if we write, for a Borel set  $C \subset \Sigma$ ,

$$C_0 = \{(x, s) \in \partial K \times [-1, 1] : (x, u, s) \in C\}$$

and we use the same letter  $Q$  for the measure

$$dQ(x, s) = \frac{p_{\pm}(x) \, d\nu(x) \times ds}{\int_{\partial K} (p_+(x) + p_-(x)) \, d\nu(x)} \quad \text{for } s \geq 0,$$

which lives on  $\partial K \times [-1, 1]$ , then

$$Q(C_0) = Q(C).$$

Therefore, if convenient, we will replace  $\Sigma$  by  $\Gamma = \partial K \times [-1, 1]$  and replace  $W_{\mathcal{B}}$  with the process  $W_{\mathcal{B}_0}$  defined on  $\mathcal{B}_0 = \{B_0 : B \in \mathcal{B}\}$ , a class of subsets of  $\Gamma$ . However, we could not do this with  $v_{n, \mathcal{B}_n}$ .

Weak convergence in function spaces is important because of its statistical application, the continuous mapping theorem, which states that continuous functionals, or statistics, of the random elements involved inherit the weak convergence. We now formulate a continuous mapping theorem in our unusual setting, where the domain of the functions depends on  $n$ . Let  $\ell^\infty(\mathcal{B}_n)$  and  $\ell^\infty(\mathcal{B})$  be the spaces of bounded functions on  $\mathcal{B}_n$  and  $\mathcal{B}$ , respectively; let  $x_n \in \ell^\infty(\mathcal{B}_n)$ ,  $x \in \ell^\infty(\mathcal{B})$  and assume that  $x$  is  $d$ -continuous. Also, assume the functionals  $\varphi_n : \ell^\infty(\mathcal{B}_n) \rightarrow \mathbb{R}$  and  $\varphi : \ell^\infty(\mathcal{B}) \rightarrow \mathbb{R}$  are such that (with  $\gamma_n$  as in (9))

$$\sup_{B_n \in \mathcal{B}_n, B \in \mathcal{B}; d(B_n, B) \leq \gamma_n} |x_n(B_n) - x(B)| \rightarrow 0 \tag{12}$$

implies

$$\varphi_n(x_n) \rightarrow \varphi(x).$$

We then have

$$\varphi_n(v_{n, \mathcal{B}_n}) \xrightarrow{d} \varphi(W_{\mathcal{B}}). \tag{13}$$

As an example, we see that

$$\sup_{B_n \in \mathcal{B}_n} |v_n(B_n)| \xrightarrow{d} \sup_{B \in \mathcal{B}} |W(B)|.$$

For the proof of (13), we only mention that a Skorokhod almost sure representation theorem yields the existence of  $\tilde{v}_{n, \mathcal{B}} \stackrel{d}{=} v_{n, \mathcal{B}}$  and  $\tilde{W}_{\mathcal{B}} \stackrel{d}{=} W_{\mathcal{B}}$  such that

$$\sup_{B \in \mathcal{B}} |\tilde{v}_n(B) - \tilde{W}(B)| \rightarrow 0 \quad \text{a.s.}$$

If we extend  $\tilde{v}_{n, \mathcal{B}}$  to  $\mathcal{B}_n$ , we obtain, from (a),

$$\begin{aligned} & \sup_{B_n \in \mathcal{B}_n, B \in \mathcal{B}; d(B_n, B) \leq \gamma_n} |\tilde{v}_n(B_n) - \tilde{W}(B)| \\ & \leq \sup_{B_n \in \mathcal{B}_n, B \in \mathcal{B}; d(B_n, B) \leq \gamma_n} |\tilde{v}_n(B_n) - \tilde{v}_n(B)| \\ & \quad + \sup_{B \in \mathcal{B}} |\tilde{v}_n(B) - \tilde{W}(B)| \xrightarrow{P} 0. \end{aligned}$$

Now, compare this with (12). The rest of the proof is elementary.

**Example 1.** Let  $K = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$  be the unit disc, so  $\partial K = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$  is the unit circle. We have  $S_K = \{(0, 0)\}$  and  $r(x) = 1$  for all  $x \in \partial K$ . Also,  $\mathcal{V}_\varepsilon(\partial K) = \{(x, y) \in \mathbb{R}^2 : (1 - \varepsilon)^2 \leq x^2 + y^2 \leq (1 + \varepsilon)^2\}$ .

(a) Let  $\mathcal{E}$  be the VC class of all closed ellipses (with interior) in  $\mathbb{R}^2$ . This  $\mathcal{E}$  is an example of the general  $\mathcal{K}$  in the canonical example in Section 1. Thus,  $\mathcal{A} = \{E \Delta K : E \in \mathcal{E}\}$  and  $\mathcal{A}_{\varepsilon_n} = \{A \in \mathcal{A} : A \subset \mathcal{V}_{\varepsilon_n}(\partial K)\}$ . By Remark 2,  $\mathcal{A}_{\varepsilon_n}$  is a VC class with uniformly bounded index.

We parametrize  $\partial K$  with the angle  $\theta \in [0, 2\pi)$  and re-express the cylinder  $\Gamma = \partial K \times [-1, 1]$  as  $[0, 2\pi) \times [-1, 1]$ . Consider the functions  $f_{\alpha,a,b,c,d} : [0, 2\pi) \rightarrow [-1, 1]$ , defined by

$$f_{\alpha,a,b,c,d}(\theta) = f(\theta) = a + b \sin^2(\theta - \alpha) + c \sin(\theta - \alpha) + d \cos(\theta - \alpha),$$

with  $\alpha \in [0, \pi/2)$  and  $a, b, c, d \in \mathbb{R}$  such that  $\sup_{0 \leq \theta < 2\pi} |f_{0,a,b,c,d}(\theta)| \leq 1$ . Denote the class of all such functions by  $\mathcal{F}_{\mathcal{E}}$ . A tedious calculation shows that

$$\mathcal{B}_0 = \{(\theta, y) \in [0, 2\pi) \times [-1, 1] : 0 < y \leq f(\theta) \text{ or } f(\theta) < y \leq 0\} : f \in \mathcal{F}_{\mathcal{E}}\}.$$

Since  $\mathcal{B}_0$  is a limit class, it can be shown, directly using the definition of a VC class, that  $\mathcal{B}_0$  is also a VC class. For  $B \in \mathcal{B}_0$ , note that for every  $\theta \in [0, 2\pi)$ , the intersection of  $B$  with  $\{(\theta, y) : y \in [-1, 1]\}$  is convex (an interval). Part (b) shows that this need not be the case in general.

(b) Consider, for the same  $K$ , the very simple class

$$\mathcal{A}_{\varepsilon_n} = \{z \in \mathbb{R}^2 : \|z - \partial K\|/\varepsilon_n \in [a, b] \cup [c, d] : -1 \leq a \leq b \leq c \leq d \leq 1\}.$$

Now,

$$\mathcal{B}_0 = \{(\theta, y) \in [0, 2\pi) \times [-1, 1] : y \in [a, b] \cup [c, d] : -1 \leq a \leq b \leq c \leq d \leq 1\}.$$

Here,  $\mathcal{B}_n = \mathcal{B}$ .

**Example 2.** Let  $K = \{(x, y) \in \mathbb{R}^2 : 0 \leq x, y \leq 1\}$  be the unit square with boundary  $\partial K$ . We obtain  $S_K = \{(x, x) : 0 < x < 1\} \cup \{(x, 1 - x) : 0 < x < 1\}$  and for, for example,  $\{(x, 0) : 0 \leq x \leq 1\} \subset \partial K$ , we see that  $r((x, 0)) = \min(x, 1 - x)$ . It is notationally somewhat cumbersome to describe  $\mathcal{V}_\varepsilon(\partial K)$  explicitly, but it is trivial to see that it is the difference of a set which is a “square with circular corners” and a smaller square.

(a) Let  $\mathcal{Q}$  be the VC class of all closed quadrangles in  $\mathbb{R}^2$ . Set  $\mathcal{A} = \{Q \Delta K : Q \in \mathcal{Q}\}$  and  $\mathcal{A}_{\varepsilon_n} = \{A \in \mathcal{A} : A \subset \mathcal{V}_{\varepsilon_n}(\partial K)\}$ . Again by Remark 2,  $\mathcal{A}_{\varepsilon_n}$  is a VC class with uniformly bounded index. The present example is somewhat similar to Example 1, but there is a substantial difference since a square is less smooth than a disc.

We parametrize  $\partial K$  with  $\theta \in [0, 4)$ , the counterclockwise “distance” from the origin, and re-express the cylinder  $\Gamma$  as  $[0, 4) \times [-1, 1]$ . Consider the functions  $f_{a,b} : [0, 4) \rightarrow [-1, 1]$ , with  $a = (a_0, a_1, a_2, a_3)$  and  $b = (b_0, b_1, b_2, b_3)$ , defined by

$$f_{a,b}(\theta) = f(\theta) = a_m(\theta - m) + b_m \quad \text{for } m \leq \theta < m + 1, m = 0, 1, 2, 3,$$

with  $a, b$  such that  $a_m \in [-2, 2]$  and  $\sup_{0 \leq \theta < 4} |f_{a,b}(\theta)| \leq 1$ . Denote the class of all such functions by  $\mathcal{F}_Q$ . Note that  $f \in \mathcal{F}_Q$  is typically discontinuous, in contrast to an  $f \in \mathcal{F}_E$  of Example 1. It can be shown that

$$\mathcal{B}_0 = \{(\theta, y) \in [0, 4) \times [-1, 1] : 0 < y \leq f(\theta) \text{ or } f(\theta) < y \leq 0\} : f \in \mathcal{F}_Q\}.$$

It readily follows that  $\mathcal{B}_0$  is a VC class.

(b) Consider (for the same  $K$ ) a larger class than  $\mathcal{Q}$ , namely  $\mathcal{C}$ , the class of all convex bodies in  $\mathbb{R}^2$ . For convenience, let  $P$  be the uniform distribution on  $[-1, 2]^2$ . The class  $\mathcal{C}$  is again an example of the general  $\mathcal{K}$  in the canonical example in Section 1, but it is not a VC class. We have  $\mathcal{A} = \{C \Delta K : C \in \mathcal{C}\}$  and  $\mathcal{A}_{\varepsilon_n} = \{A \in \mathcal{A} : A \subset \mathcal{V}_{\varepsilon_n}(\partial K)\}$ .

Consider the functions  $f : [0, 4) \rightarrow [-1, 1]$  defined by

$$f(\theta) = f_m(\theta - m) \quad \text{for } m \leq \theta < m + 1, m = 0, 1, 2, 3,$$

with  $f_m : [0, 1) \rightarrow [-1, 1]$  a concave function. Denote the class of all such functions by  $\mathcal{F}_C$ . It can be shown that

$$\mathcal{B}_0 = \{(\theta, y) \in [0, 4) \times [-1, 1] : 0 < y \leq f(\theta) \text{ or } f(\theta) < y \leq 0\} : f \in \mathcal{F}_C\}.$$

The conditions of Theorem 1 are satisfied. In particular, using [39], Corollary 2.7.9, it can be deduced that (10) and (11) hold true.

### 5. Proofs

**Proof of Lemma 1.** Based on the local Steiner formula, in the proof of Theorem 2 of [26], it is shown that the measure  $P(\tau_{\varepsilon_n}^{-1} \cdot) / \varepsilon_n$  converges in total variation to the measure  $M_p$ . This implies that  $P(\mathcal{V}_{\varepsilon_n}(\partial K)) / \varepsilon_n \rightarrow M_p(\Sigma)$  and hence that  $Q_n = P(\tau_{\varepsilon_n}^{-1} \cdot) / P(\mathcal{V}_{\varepsilon_n}(\partial K))$  converges in total variation to  $Q = M_p / M_p(\Sigma)$ . □

**Proof of Theorem 1.** First, we prove statement (a):

$$\sup_{B_n \in \mathcal{B}_n, B \in \mathcal{B}; d(B_n, B) \leq \gamma_n} |v_n(B_n) - v_n(B)| \xrightarrow{P} 0.$$

From relation (7), Lemma 1 and the Markov inequality, it follows that it is sufficient to show that

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{E} \sup_{\substack{A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1} \mathcal{B} \\ d_n(A, \tilde{A}) < \delta}} |z_n(A) - z_n(\tilde{A})| = 0. \tag{14}$$

We use [38], Lemma 19.34, page 286, for the proof of (14); in that lemma, we choose the indexing functions to be  $1_A - 1_{\tilde{A}}$ . We then obtain, taking the  $\delta$  there to be equal to  $\delta \sqrt{a_n}$ , that

for some constant  $c$ ,

$$\begin{aligned} & \mathbb{E} \sup_{\substack{A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1} \mathcal{B} \\ d_n(A, \tilde{A}) < \delta}} |z_n(A) - z_n(\tilde{A})| \\ & \leq c \left( \frac{1}{\sqrt{a_n}} \int_0^{\delta \sqrt{a_n}} \sqrt{\log \left( N_{[\cdot], n} \left( \frac{\varepsilon}{\sqrt{a_n}} \right) \tilde{N}_{[\cdot], n} \left( \frac{\varepsilon}{\sqrt{a_n}} \right) \right)} d\varepsilon \right. \\ & \quad \left. + \frac{\sqrt{n}}{\sqrt{a_n}} \int_{\mathcal{V}_{\varepsilon_n}(\partial K)} \mathbb{1}_{\{\sqrt{\log(N_{[\cdot], n}(\delta) \tilde{N}_{[\cdot], n}(\delta))} > \delta \sqrt{na_n}\}} dP \right). \end{aligned}$$

Using  $na_n \rightarrow \infty$  and (10), (11), we see that the second term on the right is equal to 0 for small  $\delta$  and large  $n$ . The first term is easily seen to be bounded by

$$c \int_0^\delta \sqrt{\log N_{[\cdot], n}(x)} dx + c \int_0^\delta \sqrt{\log \tilde{N}_{[\cdot], n}(x)} dx.$$

Hence, (14) follows using (10) and (11).

For a proof of statement (b), we need weak convergence of the finite-dimensional distributions and tightness of  $v_{n, \mathcal{B}}$ . The weak convergence of the finite-dimensional distributions follows easily from Lemma 1 and an appropriate version of the multivariate central limit theorem.

To prove tightness, we use [39], Theorem 2.11.9, a general bracketing central limit theorem. We will choose  $d$  for the semimetric  $\rho$  on  $\mathcal{B}$  which is required in that theorem. For tightness, three conditions have to be fulfilled. The first one holds trivially since  $\Psi_n$  is a sum of indicators. The third one follows readily since it is essentially our condition (11). It remains to show the second condition:

$$s_n := \sup_{\substack{B, B' \in \mathcal{B} \\ d(B, B') < \delta_n}} \sum_{i=1}^n \mathbb{E} \left( \frac{1}{\sqrt{na_n}} \mathbb{1}_{\tau_{\varepsilon_n}^{-1} B}(X_i) - \frac{1}{\sqrt{na_n}} \mathbb{1}_{\tau_{\varepsilon_n}^{-1} B'}(X_i) \right)^2 \rightarrow 0 \quad \text{for every } \delta_n \downarrow 0.$$

However,

$$\begin{aligned} s_n &= \frac{1}{na_n} \sup_{d(B, B') < \delta_n} \sum_{i=1}^n \mathbb{E} \mathbb{1}_{\tau_{\varepsilon_n}^{-1} B \Delta \tau_{\varepsilon_n}^{-1} B'}(X_i) \\ &= \frac{1}{a_n} \sup_{d(B, B') < \delta_n} P(\tau_{\varepsilon_n}^{-1}(B \Delta B')) = \sup_{Q(B \Delta B') < \delta_n^2} Q_n(B \Delta B'). \end{aligned}$$

Lemma 1 now immediately yields  $s_n \rightarrow 0$ . □

**Proof of Theorem 2.** Again, we first prove statement (a) and note that it suffices to show, for any  $\eta > 0$ , that for  $\delta > 0$  small enough and then for large  $n$ ,

$$\mathbb{P}\left(\sup_{\substack{A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1}\mathcal{B} \\ d_n(A, \tilde{A}) \leq \sqrt{\delta}}} |z_n(A) - z_n(\tilde{A})| > 2\eta\right) \leq 2\eta. \tag{15}$$

We have, for  $n$  large enough,

$$\begin{aligned} & \mathbb{P}\left(\sup_{\substack{A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1}\mathcal{B} \\ d_n(A, \tilde{A}) \leq \sqrt{\delta}}} |z_n(A) - z_n(\tilde{A})| > 2\eta\right) \\ &= \mathbb{P}\left(\sup_{\substack{A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1}\mathcal{B} \\ P(A\Delta\tilde{A}) \leq \delta a_n}} |z_n(A) - z_n(\tilde{A})| > 2\eta\right) \\ &= \mathbb{P}\left(\sup_{\substack{A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1}\mathcal{B} \\ P(A\Delta\tilde{A}) \leq \delta a_n}} |z_n(A \setminus \tilde{A}) - z_n(\tilde{A} \setminus A)| > 2\eta\right) \\ &\leq \mathbb{P}\left(\sup_{\substack{A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1}\mathcal{B} \\ P(A\Delta\tilde{A}) \leq \delta a_n}} |z_n(A \setminus \tilde{A})| > \eta\right) \\ &\quad + \mathbb{P}\left(\sup_{\substack{A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1}\mathcal{B} \\ P(A\Delta\tilde{A}) \leq \delta a_n}} |z_n(\tilde{A} \setminus A)| > \eta\right) \\ &\leq 2\mathbb{P}\left(\sup_{C \in \mathcal{C}_n, P(C) \leq \delta a_n} |z_n(C)| > \eta\right), \end{aligned} \tag{16}$$

where  $\mathcal{C}_n = \{A \setminus \tilde{A} : A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1}\mathcal{B}\} \cup \{\tilde{A} \setminus A : A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_{\varepsilon_n}^{-1}\mathcal{B}\}$ . It can be shown (see, e.g., [39], page 147), using  $A_1 \setminus A_2 = A_1 \cap A_2^c$ , that  $\mathcal{C}_n$  is a VC class. Also, the index  $w_n$  of this VC class is bounded:  $\max_{n \in \mathbb{N}} w_n < \infty$ .

We have, writing  $N = \Psi_n(\mathcal{V}_{\varepsilon_n}(\partial K))$  and  $k = na_n$ , that

$$\begin{aligned} & \mathbb{P}\left(\sup_{C \in \mathcal{C}_n, P(C) \leq \delta a_n} |z_n(C)| > \eta\right) \\ &= \sum_{m=0}^n \mathbb{P}\left(\sup_{C \in \mathcal{C}_n, P(C) \leq \delta a_n} |z_n(C)| > \eta \mid N = m\right) \mathbb{P}(N = m) \\ &= \sum_{m=0}^n \mathbb{P}\left(\sup_{C \in \mathcal{C}_n, P(C) \leq \delta a_n} \left| \frac{1}{\sqrt{k}} [\Psi_n(C) - nP(C)] \right| > \eta \mid N = m\right) \mathbb{P}(N = m) \end{aligned}$$

$$\begin{aligned} &\leq \sum_{m=\lceil k-C_\eta\sqrt{k} \rceil}^{m=\lfloor k+C_\eta\sqrt{k} \rfloor} \mathbb{P}\left(\sup_{C \in \mathcal{C}_n, P(C) \leq \delta a_n} \left| \frac{1}{\sqrt{k}} [\Psi_n(C) - nP(C)] \right| > \eta \mid N = m\right) \mathbb{P}(N = m) \\ &\quad + \mathbb{P}(|N - k| \geq C_\eta\sqrt{k}), \end{aligned}$$

where  $C_\eta$  is chosen such that the latter probability concerning the binomial( $n, k/n$ ) random variable  $N$  is bounded by  $\eta/2$  for large  $n$ . Hence, for large  $n$ ,

$$\begin{aligned} &\mathbb{P}\left(\sup_{C \in \mathcal{C}_n, P(C) \leq \delta a_n} |z_n(C)| > \eta\right) \\ &\leq \sum_{m=\lceil k-C_\eta\sqrt{k} \rceil}^{m=\lfloor k+C_\eta\sqrt{k} \rfloor} \mathbb{P}\left(\sup_{C \in \mathcal{C}_n, P_{\varepsilon_n}(C) \leq \delta} \left| \frac{1}{\sqrt{m}} \left[ \sum_{j=1}^m 1_C(Y_j) - mP_{\varepsilon_n}(C) \right] \right| > \frac{\eta}{3}\right) \mathbb{P}(N = m) \quad (17) \\ &\quad + \sum_{m=\lceil k-C_\eta\sqrt{k} \rceil}^{m=\lfloor k+C_\eta\sqrt{k} \rfloor} \mathbb{P}\left(\sup_{C \in \mathcal{C}_n, P_{\varepsilon_n}(C) \leq \delta} \frac{1}{\sqrt{k}} |m - k| P_{\varepsilon_n}(C) > \frac{\eta}{2}\right) \mathbb{P}(N = m) + \frac{\eta}{2}, \end{aligned}$$

where the  $Y_j$  are i.i.d. random vectors on  $\mathcal{V}_{\varepsilon_n}(\partial K)$  distributed according to  $P_{\varepsilon_n}$ . Note that in the first probability of the second sum, no randomness is involved and that this sum is equal to 0 for  $\delta$  small enough. For the first sum, we need a good bound for exceedance probabilities for the supremum of the empirical process on a VC class. We will use [1], Corollary 2.9. Using  $\max_{n \in \mathbb{N}} w_n < \infty$ , this leads to the following upper bound for the left-hand side of (17):

$$\sum_{m=\lceil k-C_\eta\sqrt{k} \rceil}^{m=\lfloor k+C_\eta\sqrt{k} \rfloor} 16 \exp(-\eta^2/(36\delta)) \mathbb{P}(N = m) + \frac{\eta}{2} \leq 16 \exp(-\eta^2/(36\delta)) + \frac{\eta}{2} \leq \eta$$

for small enough  $\delta$ . So, because of (16), we have proven (15) and hence (a).

For a proof of (b), we only need to show tightness of  $v_{n,\mathcal{B}}$  since the weak convergence of the finite-dimensional distributions follows as in the proof of Theorem 1.

For proving tightness, we need that, for any  $\eta > 0$ ,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sup_{\substack{B, B' \in \mathcal{B} \\ d(B, B') \leq \delta}} |v_n(B) - v_n(B')| > \eta\right) = 0$$

(see, e.g., [39], Theorem 1.5.7). Again, from (7) and Lemma 1, it suffices to show that

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sup_{\substack{A, A' \in \varepsilon_n^{-1}\mathcal{B} \\ d_n(A, A') \leq 2\delta}} |z_n(A) - z_n(A')| > \eta\right) = 0. \quad (18)$$

The proof of (18) can be given along the same lines as the proof of (15). □

## Acknowledgements

We are grateful to two referees for thoughtful comments that led to improvements of this paper.

## References

- [1] Alexander, K.S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067 (correction: **15** (1987) 428–430). [MR0757769](#)
- [2] Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972). *Robust Estimation of Location: Survey and Advances*. Princeton, NJ: Princeton Univ. Press. [MR0331595](#)
- [3] Berthet, P. and Einmahl, J.H.J. (2011). Central limit theorems for level set estimators. Work in progress.
- [4] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation in Semiparametric Models*. Baltimore: Johns Hopkins Univ. Press.
- [5] Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*. New York: Wiley. [MR1215046](#)
- [6] Cuevas, A., Fraiman, R. and Rodríguez-Casal, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.* **35** 1031–1051. [MR2341697](#)
- [7] Davies, L. (1992). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist.* **20** 1828–1843. [MR1193314](#)
- [8] Davydov, Y. and Zitikis, R. (2008). On weak convergence of random fields. *Ann. Inst. Statist. Math.* **60** 345–365. [MR2403523](#)
- [9] Deheuvels, P. and Mason, D.M. (1990). Nonstandard functional laws of the iterated logarithm for tail empirical and quantile processes. *Ann. Probab.* **18** 1693–1722. [MR1071819](#)
- [10] Deheuvels, P. and Mason, D.M. (1991). A tail empirical processes approach to some nonstandard laws of the iterated logarithm. *J. Theoret. Probab.* **4** 53–85. [MR1088393](#)
- [11] Deheuvels, P. and Mason, D.M. (1994). Functional laws of the iterated logarithm for local empirical processes indexed by sets. *Ann. Probab.* **22** 1619–1661. [MR1303659](#)
- [12] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** 1833–1855. [MR1026315](#)
- [13] Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64** 25–47. [MR1619974](#)
- [14] Einmahl, J.H.J. (1992). The almost sure behavior of the weighted empirical process and the law of the iterated logarithm for the weighted tail empirical process. *Ann. Probab.* **20** 681–695. [MR1159568](#)
- [15] Einmahl, J.H.J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stochastic Process. Appl.* **70** 31–58. [MR1472958](#)
- [16] Einmahl, J.H.J. and Lin, T. (2006). Asymptotic normality of extreme value estimators on  $C[0, 1]$ . *Ann. Statist.* **34** 469–492. [MR2275250](#)
- [17] Einmahl, J.H.J. and Mason, D.M. (1992). Generalized quantile processes. *Ann. Statist.* **20** 1062–1078. [MR1165606](#)
- [18] Einmahl, U. and Mason, D.M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Relat. Fields* **107** 283–311. [MR1440134](#)
- [19] Ferger, D. (2004). Boundary estimation based on set-indexed empirical processes. *J. Nonparametric Stat.* **16** 245–260. [MR2053073](#)
- [20] Grübel, R. (1988). The length of the shorth. *Ann. Statist.* **16** 619–628. [MR0947565](#)
- [21] Hug, D., Last, G. and Weil, W. (2004). A local Steiner-type formula for general closed sets and applications. *Math. Z.* **246** 237–272. [MR2031455](#)

- [22] Ibragimov, I.A. and Has'minskii, R.Z. (1981). *Statistical Estimation. Asymptotic Theory*. New York: Springer. [MR0620321](#)
- [23] Khmaladze, E.V. (1998). Goodness of fit tests for “Chimeric” alternatives. *Statist. Neerlandica* **52** 90–111. [MR1615550](#)
- [24] Khmaladze, E.V. (2007). Differentiation of sets in measure. *J. Math. Anal. Appl.* **334** 1055–1072. [MR2338647](#)
- [25] Khmaladze, E.V., Mnatsakanov, R.M. and Toronjadze, N. (2006). Local covering numbers and change-set problem. *Math. Methods Statist.* **15** 289–308. [MR2278291](#)
- [26] Khmaladze, E.V. and Weil, W. (2008). Local empirical processes near boundaries of convex bodies. *Ann. Inst. Statist. Math.* **60** 813–842. [MR2453573](#)
- [27] Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. [MR1041391](#)
- [28] Korostelev, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction*. New York: Springer. [MR1226450](#)
- [29] Mason, D.M. (1988). A strong invariance theorem for the tail empirical process. *Ann. Inst. H. Poincaré Probab. Statist.* **24** 491–506. [MR0978022](#)
- [30] Mason, D.M. (2004). A uniform functional law of the logarithm for the local empirical process. *Ann. Probab.* **32** 1391–1418. [MR2060302](#)
- [31] Müller, D.W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746. [MR1147099](#)
- [32] Müller, H.-G. and Stadtmüller, U. (1999). Discontinuous versus smooth regression. *Ann. Statist.* **27** 299–337. [MR1701113](#)
- [33] Nolan, D. (1991). The excess-mass ellipsoid. *J. Multivariate Anal.* **39** 348–371. [MR1147127](#)
- [34] Rio, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21–45. [MR1254823](#)
- [35] Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics with Applications* (W. Grossman, G. Pflug, I. Vincze and W. Wertz, eds.) 283–297. Dordrecht: Reidel. [MR0851060](#)
- [36] Schneider, R. (1993). *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge: Cambridge Univ. Press. [MR1216521](#)
- [37] Schneider, R. and Weil, W. (2008). *Stochastic and Integral Geometry*. New York: Springer. [MR2455326](#)
- [38] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge Univ. Press. [MR1652247](#)
- [39] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes, With Applications to Statistics*. New York: Springer. [MR1385671](#)
- [40] Wasserman, L. (2006). *All of Nonparametric Statistics*. New York: Springer. [MR2172729](#)

Received February 2009 and revised July 2009