Central Limit Theorems For Local Empirical Processes Near Boundaries of Sets
Einmahl, J.H.J.; Khmaladze, E.V.

Publication date:
2007

Link to publication

Citation for published version (APA):
No. 2007–66

CENTRAL LIMIT THEOREMS FOR LOCAL EMPIRICAL PROCESSES NEAR BOUNDARIES OF SETS

By John H.J. Einmahl, Estâte V. Khmaladze

September 2007

ISSN 0924-7815
Central limit theorems
for local empirical processes
near boundaries of sets

John H.J. Einmahl       Estáte V. Khmaladze
Tilburg University      Victoria University of Wellington

September 3, 2007

Abstract. 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. A continuous mapping theorem for our situation is also derived and some examples are presented.

Running title. Empiricals near set boundaries.

AMS 2000 subject classifications. 60F05, 60F17, 60G55, 62G30.

JEL code. C14.

Key words and phrases. Convex body, differentiation of sets, Gaussian behavior, local empirical process, set boundary, weak convergence.
1 Introduction

Let $X_1, \ldots, X_n$ 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 with $p$. For a Borel measurable subset $D$ of $\mathbb{R}^d$, write

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

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

Let $K \in \mathcal{D}$ be a convex body, i.e. a closed, bounded convex set, that has interior points; denote with $\partial K$ its boundary. The set $K$ will be fixed throughout. It is the aim of this paper to present appropriate central limit theorems for $\Psi_n$ in the neighborhood of $\partial K$. For this purpose set $V_\varepsilon(\partial K) = \{ z \in \mathbb{R}^d : \| z - \partial K \| \leq \varepsilon \}, \varepsilon > 0$, where $\| z - \partial K \| = \min_{x \in \partial K} \| z - x \|$. For a Borel set $A \subset V_\varepsilon(\partial K)$, define

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

with $a = P(V_\varepsilon(\partial K))$. Denoting the conditional probability distribution on $V_\varepsilon(\partial K)$ with $P_\varepsilon(A) = P(A)/a$, we can also write

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

This reflects that, on average, the effective sample size is equal to $na$. Let $\varepsilon = \varepsilon_n \to 0$, as $n \to \infty$, and, for $n \in \mathbb{N}$, let $\mathcal{A}_{\varepsilon_n}$ be a class of measurable subsets of $V_\varepsilon(\partial K)$. [The canonical example is constructed as follows. Let $\mathcal{K}$ be a fixed subset of $\mathcal{D}$ and define $\mathcal{A} = \{ K' \triangle K : K' \in \mathcal{K} \}$, where $\triangle$ denotes ‘symmetric difference’. Now take $\mathcal{A}_{\varepsilon_n} = \{ A \in \mathcal{A} : A \subset V_\varepsilon(\partial K) \}$.]

Our main results, the aforementioned central limit theorems, concern the local empirical process near $\partial K$ and indexed by $\mathcal{A}_{\varepsilon_n}$

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

(1)
where we also assume, in addition to $\varepsilon_n \to 0$, that

$$n\varepsilon_n \to \infty, \quad \text{as } n \to \infty.$$ 

The latter growth condition on $\varepsilon_n$ ensures that the sets in $A_{\varepsilon_n}$ contain enough observations to obtain Gaussian limit behavior; Poisson limit behavior of $\Psi_n$ has been studied in Khmaladze and Weil (2007).

Although here very natural, it is in general unusual that an empirical process is defined on a class of sets that depends on $n$. Moreover, all sets in our class shrink and have measure 0 in the limit. Therefore, it is not clear on which class of sets the limiting process ‘lives’. We will show that this process should be defined on a class of subsets 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 $A_{\varepsilon_n}$.

The local empirical process for one-dimensional $X_i$, i.e. the empirical process in the neighborhood of a point $c \in \mathbb{R} \cup \{-\infty, \infty\}$ is a classical object in probability theory, which has proven to be very valuable in statistics, see, e.g., Mason (1988), Dekkers, Einmahl and de Haan (1989), Deheuvels and Mason (1990, 1991), Einmahl (1992), the book Csörgő and Horváth (1993), and Khmaladze (1998). The one-dimensional local empirical process has been extended to the multivariate setup, but typically only the neighborhood of a point $c \in \mathbb{R}^d$ or the region outside a large sphere are considered, see, e.g., Deheuvels and Mason (1994), Rio (1994), Einmahl (1997), Einmahl and Mason (1997), Drees and Huang (1998), Mason (2004), and Davydov and Zitikis (2007). For a local empirical process for function-valued random elements, see Einmahl and Lin (2006).

In $\mathbb{R}^d$, the local empirical process in the neighborhood of the boundary of a set seems a very natural object, but it turns out to be a new concept in probability theory. Clearly, since a set is much ‘richer’ than a point, our local empirical processes must be more diverse and useful. Khmaladze and Weil (2007) and the present paper are the first steps in unraveling these processes.

Apart from the intrinsic motivation that the local empirical process near the boundary of a set is a very interesting probabilistic object, the study of this local empirical process is also very useful from a statistical point of view. Indeed, consider statistical problems
where the parameter is a set. If the parameter were a vector, as in parametric problems (see, e.g., Ibragimov and Has’minskii (1981)) or a function, as in nonparametric problems (see, e.g., Wasserman (2006)), or a combination of both, as in semiparametric problems (see, e.g., Bickel et al. (1993)), the local analysis of the likelihood ratio, or an other process which the inference is based upon, in the neighborhood of the true value of the parameter is a crucial step in asymptotic statistical theory. But it is equally important to be able to carry out such a local analysis when the parameter is a set.

Important examples of such a model are provided by the class of spatial change-point problems or so-called change-set problems (see Khmaladze, Mnatsakanov and Toronjadze (2006)). 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, \( K \), on the boundary of which the distribution of the point process (e.g., the distribution of the marks) sharply changes. In most of the particular formulations of the change-set problem the likelihood ratio is some form 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.

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

## 2 Some Geometry and Differentiability of Sets

In this section we first briefly review some relevant notation and facts from geometry. Then we define the concept of ‘differentiation of sets in measure’. This differentiation has been developed in Khmaladze (2007). In that paper and the references therein (in particular Schneider (1993)) also more details about the required geometry can be found.

Let \( K \in \mathcal{D} \) be our convex body. Denote with \( \Pi(z) \) ‘the’ metric projection of \( z \in \mathbb{R}^d \) on \( \partial K \), i.e. \( \Pi(z) \) is a nearest point to \( z \) on \( \partial K \). The set of \( z \)-values for which such a nearest point is not unique has Lebesgue measure 0 and is called the skeleton \( S_K \) of \( K \) (a subset of \( K \)). 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 with $N(x)$ and the normal bundle of $K$ is defined as

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

The cylinder $\Sigma_1 = \text{Nor}(K) \times [-1, 1]$ will be important for describing our limiting processes. Also write $\Sigma_\infty = \text{Nor}(K) \times \mathbb{R}$.

We also need the so-called local magnification map $\tau_\varepsilon$, see Khmaladze (2007). Any point $z \in \mathbb{R}^d \setminus S_K$ can be written as $z = \Pi(z) + d_s(z)u$, with $d_s(z)$ the signed (‘+’ outside) distance between $z$ and $\Pi(z)$; $u$ 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 $\Sigma_1$.

We are now prepared to introduce the aforementioned differentiation of sets. Consider the first support measure on $\text{Nor}(K)$, see Schneider (1993). 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 on $\partial K$ in a one-to-one way. On $\partial K$ this map coincides with Hausdorff measure. For both measures we will use the same notation $\nu$. Define the measure $M = \nu \times \mu$ (\(\mu\) one-dimensional Lebesgue measure). Consider a (Borel) set-valued mapping $K(\varepsilon)$, $\varepsilon \in [0, 1]$, such that $K(0) = K$, with $K$ as before; write $A(\varepsilon) = K(\varepsilon) \Delta K$. The set-valued mapping $A(\varepsilon)$, $\varepsilon \in [0, 1]$, is called differentiable at $\partial K$ and $\varepsilon = 0$ if ($\mu_d$ denotes $d$-dimensional Lebesgue measure)

- there exists a finite $T > 0$ such that $\frac{1}{\varepsilon}\mu_d(A(\varepsilon) \cap [\mathcal{V}_{T\varepsilon}(\partial K)]^c) \to 0$ as $\varepsilon \to 0$, and
- there exists a Borel set $B \subset \Sigma_\infty$ such that $M(\tau_\varepsilon A(\varepsilon) \Delta B) \to 0$ as $\varepsilon \to 0$ (where $\tau_\varepsilon A = \{\tau_\varepsilon(z) : z \in A\}$).
The set $B$ is called the derivative of $A(\varepsilon)$ at $\partial K$. We also define the set-valued mapping $K(\varepsilon)$, $\varepsilon \in [0, 1]$, to be differentiable at $K$ and $\varepsilon = 0$ if $K(\varepsilon) K$ is differentiable at $\partial K$. In this case the derivative of $K(\varepsilon)$ at $\partial K$ is defined to be the same as that of $K(\varepsilon) K$ at $\partial K$. This is denoted as

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

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 $z \in K$ or not. More formally, we require the existence of two functions $p_+$ and $p_-$ on $\partial K$ such that, as $\varepsilon \to 0$,

$$\frac{1}{\varepsilon} \int_{V_\varepsilon(\partial K) \setminus K} |p(z) - p_+(\Pi(z))| d\mu_d(z) \to 0, \quad (2)$$

$$\frac{1}{\varepsilon} \int_{V_\varepsilon(\partial K) \cap K} |p(z) - p_-(\Pi(z))| d\mu_d(z) \to 0. \quad (3)$$

Now define a measure $M_p$ on $\Sigma_\infty$ as follows

$$dM_p(x, u, s) = p_+(x) d\nu(x, u) \times ds, \quad \text{for } s > 0,$$

$$dM_p(x, u, s) = p_-(x) d\nu(x, u) \times ds, \quad \text{for } s \leq 0.$$

For convenience assume $p_+$ and $p_-$ are bounded (although a weaker, integrability condition would suffice). An easy, but 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 Khmaladze (2007) shows the ‘differentiability of sets in measure’. If $\frac{1}{\varepsilon} P(A(\varepsilon) \cap [\mathcal{V}_\varepsilon(\partial K)]^c) \to 0$ and if $A(\varepsilon)$ is differentiable at $\partial K$, then

$$\frac{d}{d\varepsilon} P(A(\varepsilon)) |_{\varepsilon=0} = M_p \left( \frac{d}{d\varepsilon} A(\varepsilon) |_{\varepsilon=0} \right). \quad (4)$$

3 Main Results

Let $A_{\varepsilon_n}$ be as in Section 1 and assume $M_p(\Sigma_1) > 0$. Writing $a_n = P(\mathcal{V}_{\varepsilon_n}(\partial K))$, it easily follows, using (4), that $a_n/\varepsilon_n \to M_p(\Sigma_1)$. Hence we have, just as for $\varepsilon_n$,

$$a_n \to 0 \quad \text{and} \quad na_n \to \infty.$$
Denote with $\mathcal{B}$ 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 $M(\tau_{\varepsilon_n} A_n \triangle B) \rightarrow 0$. [Observe that for a thus converging sequence of Borel subsets of $\Sigma_1$ 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 $M(B_1 \triangle B_2) = 0$. Out of every such an equivalence class we choose one limit (Borel) set $B$, say. Whether 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 $z_n$, from Section 1. Write $\tau_{\varepsilon_n}^{-1} C = \{z \in \nu_{\varepsilon_n}(\partial K) : \tau_{\varepsilon_n}(z) \in C\}$, for a Borel set $C \subset \Sigma_1$, and define $\Phi_n(C) = \Psi_n(\tau_{\varepsilon_n}^{-1} C)$ and $Q_n(C) = P_{\varepsilon_n}(\tau_{\varepsilon_n}^{-1} C)$. Thus for any Borel set $C \subset \Sigma_1$, we can define

$$v_n(C) := \frac{1}{\sqrt{n a_n}}[\Phi_n(C) - n a_n Q_n(C)]$$

$$= \frac{1}{\sqrt{n a_n}}[\Psi_n(\tau_{\varepsilon_n}^{-1} C) - n a_n P_{\varepsilon_n}(\tau_{\varepsilon_n}^{-1} C)] = z_n(\tau_{\varepsilon_n}^{-1} C). \tag{5}$$

In particular, for each $n \in \mathbb{N}$, $v_n$ is defined on $\mathcal{B}_n := \{\tau_{\varepsilon_n} A : A \in \mathcal{A}_{\varepsilon_n}\}$, but also on $\mathcal{B}$. So, summarizing, we have three local empirical processes:

$$z_n, \mathcal{A}_{\varepsilon_n} := \{z_n(A) : A \in \mathcal{A}_{\varepsilon_n}\},$$

the transformed version of this process on $\Sigma_1$:

$$v_n, \mathcal{B}_n := \{v_n(B) : B \in \mathcal{B}_n\},$$

and the latter process on the limiting class of sets:

$$v_n, \mathcal{B} := \{v_n(B) : B \in \mathcal{B}\}.$$
Write $C_- = \{(x, u, s) \in C : (x, u) \in \text{Nor}(K), s \leq 0\}$ and $C_+ = C \setminus C_-$. In case $p_+(x) = c_+$ and $p_-(x) = c_-$ for all $x \in \partial K$, we have

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

When, e.g., $0 = c_- < c_+$ we obtain $Q(C) = Q(C_+) = M(C_+) / \nu(\partial K)$.

For Borel sets $C, C' \subset \Sigma_1$ define $d(C, C') = (Q(C \triangle C'))^{1/2}$. Throughout, we will assume that $(B, d)$ is totally bounded, and that

$$\sup_{A \in \mathcal{A}_n} \inf_{B \in \mathcal{B}} d(\tau_{\varepsilon_n} A, B) \to 0. \quad (7)$$

In particular, every sequence $(A_n)_{n=1}^\infty$, with $A_n \in \mathcal{A}_n$, has a subsequence $(A_{n_k})_{k=1}^\infty$ such that for some $B \in \mathcal{B}$, $d(\tau_{\varepsilon_{n_k}} A_{n_k}, B) \to 0$. Assumption (7) can be written as

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

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

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

Thus the Hausdorff distance between $B_n$ and $B$ tends to 0:

$$\gamma_n := \max \left( \sup_{B_n \in \mathcal{B}_n} \inf_{B \in \mathcal{B}} d(B_n, B), \sup_{B \in \mathcal{B}} \inf_{B_n \in \mathcal{B}_n} d(B_n, B) \right) \to 0. \quad (8)$$

It is the aim of this paper to present a central limit theorem for $z_n, A_{\varepsilon_n}$, or equivalently $v_{n, B_n}$, a local empirical process on a ‘moving’ class of sets. Motivated by the considerations in the previous paragraphs, we mean with ‘central limit theorem for $z_n, A_{\varepsilon_n}$’:

(a) $\sup_{B_n \in \mathcal{B}_n, B \in \mathcal{B}} |v_n(B_n) - v_n(B)|_\nu \to 0,$

and

(b) $v_{n, B} \stackrel{d}{\to} W_B := \{W(B), B \in B\}.$

Here $W_B$ 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 (1996). The following uniform version of (6) is very useful for proving this central limit theorem.
Lemma 1 From (2) and (3) it follows that $Q_n$ converges to $Q$ in total variation:

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

with the ‘sup’ taken over all Borel sets $C \subset \Sigma_1$.

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 $\delta > 0$ there exists a finite collection of pairs (brackets) $[A(\delta), A(\delta)]$ of Borel sets in $\mathcal{V}_{\varepsilon_n}(\partial K)$, with $d_n(A(\delta), A(\delta)) \leq \delta$, such that any set $A \in \mathcal{A}_{\varepsilon_n}$ can be placed in a bracket from this collection: $A(\delta) \subset A \subset A(\delta)$. Consider such a class of brackets with minimal cardinality; denote this cardinality (the bracketing number) with $N[\cdot,n](\delta)$ and let $N[\cdot,n](\delta)$ be the set of $A(\delta)$’s of this class. We assume the same for $\tau_{\varepsilon_n}B := \{\tau_{\varepsilon_n}^{-1}B : B \in \mathcal{B}\}$ and use the notations $\tilde{N}[\cdot,n](\delta)$ and $\tilde{N}[\cdot,n](\delta)$. We will require

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \int_{0}^{\delta} \sqrt{\log N[\cdot,n](x)} \, dx = 0, \quad (9)$$
$$\lim_{\delta \to 0} \limsup_{n \to \infty} \int_{0}^{\delta} \sqrt{\log \tilde{N}[\cdot,n](x)} \, dx = 0. \quad (10)$$

Theorem 1 Under the aforementioned assumptions, in particular the growth conditions on $\varepsilon_n$, the approximation of $p$ by $p_+$ or $p_-$ in (2) and (3), the relation between $\mathcal{A}_{\varepsilon_n}$ and $\mathcal{B}$ specified in (7), and the entropy conditions (9) and (10), the central limit theorem for $z_{n,\mathcal{A}_{\varepsilon_n}}$ holds, i.e. 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-Červonenkis (VC) classes (see, e.g., van der Vaart and Wellner (1996), 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 $\mathcal{B}$ is a VC class. Assume we have $\varepsilon_n \to 0$ and $n\varepsilon_n \to \infty$ and (2), (3), and (7), then the central limit theorem for $z_{n,\mathcal{A}_{\varepsilon_n}}$ holds, i.e. statements (a) and (b) hold true.

9
Remark 1 The limiting process \( W_B \) is defined on subsets of the cylinder \( \Sigma_1 = \text{Nor}(K) \times [-1,1] \). This cylinder is not easy to visualize. For a Borel set \( C \subset \Sigma_1 \), write

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

Since \( Q(C) \) only depends on \( C \) through \( C_0 \) we can define \( Q(C_0) = Q(C) \). Therefore, if convenient, we will replace \( \Sigma_1 \) by \( \Gamma_1 = \partial K \times [-1,1] \) and view \( W_B \) as a process defined on \( B_0 = \{B_0 : B \in \mathcal{B}\} \), a class of subsets of \( \Gamma_1 \). However, we cannot do this with \( v_{n,B_n} \).

Remark 2 Consider the canonical example of Section 1 and let \( \mathcal{K} \) be a VC class. Then \( \mathcal{A} \) is a VC class too, with index \( t \), say. Since \( \mathcal{A}_{\varepsilon_n} \subset \mathcal{A} \), the index \( t_n \) of \( \mathcal{A}_{\varepsilon_n} \) indeed satisfies \( t_n \leq t \).

Weak convergence in function spaces is important because of the continuous mapping theorem, which states that continuous functionals of the random elements involved inherit the weak convergence. Now we 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 \( x_n \in \ell^\infty(\mathcal{B}_n) \) and \( x \in \ell^\infty(\mathcal{B}) \) and assume \( x \) is \( d \)-continuous. Also assume the functionals \( \varphi_n : \ell^\infty(\mathcal{B}_n) \to \mathbb{R} \) and \( \varphi : \ell^\infty(\mathcal{B}) \to \mathbb{R} \) are such that (with \( \gamma_n \) as in (8))

\[
\sup_{B_n \in \mathcal{B}_n, B \in \mathcal{B}, d(B_n, B) \leq \gamma_n} |x_n(B_n) - x(B)| \to 0,
\]

implies

\[
\varphi_n(x_n) \to \varphi(x).
\]

Then we have

\[
\varphi_n(v_{n,B_n}) \xrightarrow{d} \varphi(W_B).
\]

For the proof we only mention that a Skorohod almost sure representation theorem yields the existence of \( \tilde{v}_{n,B} \xrightarrow{d} v_{n,B} \) and \( \tilde{W}_B \xrightarrow{d} W_B \) such that

\[
\sup_{B \in \mathcal{B}} |\tilde{v}_n(B) - \tilde{W}(B)| \to 0 \quad \text{a.s.}
\]
If we extend $\tilde{v}_{n,B}$ to $B_n$, we obtain from (a)

$$\sup_{B_n \in B_n, B \in B; d(B_n, B) \leq \gamma_n} |\tilde{v}_{n,B_n} - \tilde{W}(B)| \leq \sup_{B_n \in B_n, B \in B; d(B_n, B) \leq \gamma_n} \left| \tilde{v}_{n,B_n} - \tilde{v}_{n,B} \right| + \sup_{B \in B} \left| \tilde{v}_{n,B_n} - \tilde{W}(B) \right| \overset{P}{\to} 0.$$  

Now compare this with (11). The rest of the proof is elementary. (Here and in the sequel we assume for convenience that our classes of sets are such that the various ‘suprema’ are measurable, i.e. that they are random variables.)

Using this continuous mapping theorem, we obtain the following corollary to Theorems 1 and 2.

**Corollary 1** Let $q_n : \ell^\infty(B_n) \to \mathbb{R}^+$ and $q : \ell^\infty(B) \to \mathbb{R}^+$ be functionals such that

$$\sup_{B_n \in B_n, B \in B; d(B_n, B) \leq \gamma_n} |q_n(B_n) - q(B)| \to 0 \quad \text{and} \quad \inf_{B \in B} q(B) > 0,$$

then

$$\sup_{B_n \in B_n} \left| v_{n,B_n} \right| \to \sup_{B \in B} \left| W(B) \right|.$$  

We have in particular by taking $q_n \equiv 1$ and $q \equiv 1$

$$\sup_{B_n \in B_n} \left| v_{n,B_n} \right| \to \sup_{B \in B} \left| W(B) \right|.$$  

**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 $V_\varepsilon(\partial K) = \{(x, y) \in \mathbb{R}^2 : (1 - \varepsilon)^2 \leq x^2 + y^2 \leq (1 + \varepsilon)^2\}$.

(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. So $A = \{E \Delta K : E \in \mathcal{E}\}$ and $A_{\varepsilon_n} = \{A \in A : A \subset V_{\varepsilon_n}(\partial K)\}$. By Remark 2, $A_{\varepsilon_n}$ is a VC class with uniformly bounded index.

We parametrize $\partial K$ with the angle $\theta \in [0, 2\pi)$ and rewrite the cylinder $\Gamma_1 = \partial K \times [-1, 1]$ as $[0, 2\pi) \times [-1, 1]$. Consider the functions $f_{a,b,c,d} : [0, 2\pi) \to [-1, 1]$, defined by

$$f_{a,b,c,d}(\theta) = f(\theta) = a + b \sin^2(\theta - \alpha) + c \sin(\theta - \alpha) + d \cos(\theta - \alpha),$$
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 with \( \mathcal{F}_\xi \). A tedious calculation shows:

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

Since \( \mathcal{B}_0 \) is a limit class, it can be shown, using the definition of VC class directly, that \( \mathcal{B}_0 \) is a VC class too. 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, e.g., \( \{(x, 0) : 0 \leq x \leq 1\} \subset \partial K \) 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 \( \mathcal{A} = \{Q \triangle K : Q \in \mathcal{Q}\} \) 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 rewrite the cylinder \( \Gamma_1 \) as \([0, 4) \times [-1, 1]\). Consider the functions \( f_{a,b} : [0, 4) \to [-1, 1] \), with \( a = (a_0, a_1, a_2, a_3) \) and \( b = (b_0, b_1, b_2, b_3) \), defined by

\[
f_{a,b}(\theta) = f(\theta) = a_m(\theta - m) + b_m, \quad \text{for } m \leq \theta < m + 1; \quad m = 0, 1, 2, 3,
\]
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 with $\mathcal{F}_Q$. 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 $Q$, namely $C$, the class of all convex bodies in $\mathbb{R}^2$. For convenience let $P$ be the uniform distribution on $[-1, 2]^2$. The class $C$ is again an example of the general $K$ in the canonical example in Section 1, but it is not a VC class. We have $\mathcal{A} = \{C \triangle K : C \in \mathcal{C}\}$ and $\mathcal{A}_{e_n} = \{A \in \mathcal{A} : A \subset V_{e_n}(\partial K)\}$.

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

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

with $f_m : [0, 1) \to [-1, 1]$ a concave function. Denote the class of all such functions with $\mathcal{F}_C$. 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 Corollary 2.7.9 in van der Vaart and Wellner (1996), it can be derived that (9) and (10) hold true.

**Remark 3** Similar to the discussions in Khmaladze (2007) and Khmaladze and Weil (2007), 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 when $K$ is closed and bounded and has a boundary of positive reach (intuitively: $K$ has a ‘smooth’ boundary). Indeed, the key objects like 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 Hug, Last and Weil (2004). This offers perspectives for considering such a general $K$ in the statements of our results.
4 Proofs

Proof of Lemma 1 Based on the local Steiner formula, in the proof of Theorem 2 of Khmaladze and Weil (2007) it is shown that the measure \( P(\tau_{\varepsilon_n}^{-1})/\varepsilon_n \) converges in total variation to the measure \( M_p \). This implies that \( P(V_{\varepsilon_n}(\partial K))/\varepsilon_n \to M_p(\Sigma_1) \) and hence that \( Q_n = P(\tau_{\varepsilon_n}^{-1})/P(V_{\varepsilon_n}(\partial K)) \) converges in total variation to \( Q = M_p/M_p(\Sigma_1) \). \( \square \)

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

\[
\sup_{B_n \in B_n, B \in B_n: d(B_n, B) \leq \gamma_n} [v_n(B_n) - v_n(B)] \to 0. 
\]

From relation (5), Lemma 1, and the Markov inequality, it follows that it is sufficient to show that

\[
\lim_{\delta \downarrow 0} \limsup_{n \to \infty} E \sup_{A \in A_n, \tilde{A} \in \tau_{\varepsilon_n}^{-1}B} |z_n(A) - z_n(\tilde{A})| = 0. \tag{12}
\]

The proof of (12) goes through several steps.

Step 1. Let \( \eta > 0 \) and take \( 0 < \delta_n \leq \eta^{1/2}/(na_n)^{1/4} \). Then

\[
\sup_{A \in A_n} |z_n(A) - z_n(\tilde{A}(\delta_n))| \leq \max_{A \in A_n} |z_n(\tilde{A}(\delta_n)) - z_n(A(\delta_n))| + \eta.
\]

The same holds true for \( \tau_{\varepsilon_n}^{-1}B \).

The proof follows easily from \( A(\delta) \subset A \subset \tilde{A}(\delta) \) and the fact that \( \Psi_n \) and \( P_{\varepsilon_n} \) are measures and therefore respect this monotonicity.

Step 2. Let \( \delta > 0 \). If \( \delta_n \leq \frac{1}{2}\delta \) is as in Step 1, then

\[
\sup_{A \in A_n, \tilde{A} \in \tau_{\varepsilon_n}^{-1}B} |z_n(A) - z_n(\tilde{A})| 
\leq \max_{A \in A_n} |z_n(\tilde{A}(\delta_n)) - z_n(A(\delta_n))| + \max_{A \in \tau_{\varepsilon_n}^{-1}B} |z_n(\tilde{A}(\delta_n)) - z_n(A(\delta_n))| + 2\eta
\]

\[
+ \max_{A \in \mathcal{N}_{\varepsilon_n}(\delta_n), \tilde{A} \in \mathcal{N}_{\varepsilon_n}(\delta_n)} |z_n(A) - z_n(\tilde{A})|.
\]

This follows from adding and subtracting \( A(\delta_n) \) and \( \tilde{A}(\delta_n) \) within the absolute value on the left, applying the triangle inequality, and then using Step 1.
Step 3. If Borel sets \( A, A' \subset V_{\varepsilon_n}(\partial K) \) satisfy \( d_n(A, A') \leq \delta' \), then
\[
\mathbb{E} \max_{A, A'} |z_n(A) - z_n(A')| \leq 2K \left[ \frac{1}{3^{2/3}} \log(1 + m) + \delta' \sqrt{\log(1 + m)} \right],
\]
where \( K \) is a universal constant and \( m \) is the number of pairs \((A, A')\) considered.

First use \( |z_n(A) - z_n(A')| \leq |z_n(A \setminus A')| + |z_n(A' \setminus A)| \). The rest follows from Lemma 2.2.10 in van der Vaart and Wellner (1996), in conjunction with Bernstein’s inequality and the property that the \( L_1 \)-norm is bounded by the \( \psi_1 \) Orlicz norm (with \( \psi_1(x) = e^x - 1 \)).

Step 4. Let \( \eta^{1/2}/(2(na_n)^{1/4}) \leq \delta_n \leq \eta^{1/2}/(na_n)^{1/4} \), with \( a_n \to 0 \) and \( na_n \to \infty \), and let \([\underline{A}(\delta_n), \overline{A}(\delta_n)]\) be a \( \delta_n \)-bracket of \( A \), then for large \( n \),
\[
\mathbb{E} \max_{\underline{A} \in \underline{A}_n} |z_n(\overline{A}(\delta_n)) - z_n(\underline{A}(\delta_n))| \leq 3K \delta_n \sqrt{\log(1 + N_{\underline{1},n}(\delta_n))}.
\]

Also
\[
\mathbb{E} \max_{\underline{A} \in \tau_{\varepsilon_n}^{-1}B} |z_n(\overline{A}(\delta_n)) - z_n(\underline{A}(\delta_n))| \leq 3K \delta_n \sqrt{\log(1 + N_{\underline{1},n}(\delta_n))}.
\]

The first inequality follows direct from Step 3. For the second inequality we first note that \( 1/\sqrt{na_n} \leq 4\delta_n^2/\eta \) and therefore it is sufficient to show that \( 3\delta_n \sqrt{\log(1 + N_{\underline{1},n}(\delta_n))} \leq \eta \), or \( \lim_{n \to \infty} \delta_n \sqrt{\log(1 + N_{\underline{1},n}(\delta_n))} = 0 \). Observe that (9) and the fact that \( \sqrt{\log N_{\underline{1},n}(x)} \) is non-increasing in \( x \), imply that
\[
\lim_{\delta \to 0} \limsup_{n \to \infty} \delta \sqrt{1 + \log N_{\underline{1},n}(\delta)} = 0,
\]
which yields \( \lim_{n \to \infty} \delta_n \sqrt{\log(1 + N_{\underline{1},n}(\delta_n))} = 0 \). The result for \( \tau_{\varepsilon_n}^{-1}B \) follows in the same way.

Step 5 (Chaining). Let \( \delta > 0 \) and define \( \delta_j = 2^{-j}, j = 0, 1, \ldots, k_n \), with \( k_n = \min\{j : \delta_j \leq \eta^{1/2}/(na_n)^{1/4}\} \). For \( A \in N_{\underline{1},n}(\delta_{k_n}) \), there exists an \( A' \in A_{\varepsilon_n} \) with \( d_n(A, A') \leq \delta_{k_n} \). Let \( A_0, A_1, \ldots, A_{k_n} = A \) be a ‘chain’ for \( A' \), meaning that \( A_j \) is an \( A'(\delta_j) \). Construct a similar
chain for $\tilde{A}'$ corresponding to $\tilde{A} \in \tilde{N}_{[|]}(\delta_{k_n})$. Then we have

$$\max_{A \in N_{[|]}(\delta_{k_n}), \tilde{A} \in \tilde{N}_{[|]}(\delta_{k_n}), d_n(A, \tilde{A}) \leq 2\delta} |z_n(A) - z_n(\tilde{A})|$$

$$\leq \sum_{j=1}^{k_n} \max_{A_j \in N_{[|]}(\delta_j)} |z_n(A_j) - z_n(A_{j-1})| + \sum_{j=1}^{k_n} \max_{\tilde{A}_j \in \tilde{N}_{[|]}(\delta_j)} |z_n(\tilde{A}_j) - z_n(\tilde{A}_{j-1})|$$

$$+ \max_{A_0 \in N_{[|]}(\delta_0), \tilde{A}_0 \in \tilde{N}_{[|]}(\delta_0), d_n(A_0, \tilde{A}_0) \leq 6\delta} |z_n(A_0) - z_n(\tilde{A}_0)|.$$ 

The proof follows from repeated use of the triangle inequality.

**Step 6.** We have

$$\mathbb{E} \max_{A_j \in N_{[|]}(\delta_j)} |z_n(A_j) - z_n(A_{j-1})| \leq 20K \int_{0}^{\delta_j-1} \sqrt{\log N_{[|]}(x)} \ dx,$$

and hence

$$\mathbb{E} \sum_{j=1}^{k_n} \max_{A_j \in N_{[|]}(\delta_j)} |z_n(A_j) - z_n(A_{j-1})| \leq 20K \int_{0}^{\delta} \sqrt{\log N_{[|]}(x)} \ dx.$$ 

Also

$$\mathbb{E} \sum_{j=1}^{k_n} \max_{\tilde{A}_j \in \tilde{N}_{[|]}(\delta_j)} |z_n(\tilde{A}_j) - z_n(\tilde{A}_{j-1})| \leq 20K \int_{0}^{\delta} \sqrt{\log \tilde{N}_{[|]}(x)} \ dx.$$ 

This follows readily from combination of Step 3 and similar arguments as in the proof of Step 4.

Now, using (9) and (10), by Steps 2 and 4 and then Step 5 (where $\delta_{k_n}$ is the $\delta_n$ of Step 2) and Step 6, it follows that (12) holds if we can show that

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathbb{E} \max_{A_0 \in N_{[|]}(\delta_0), \tilde{A}_0 \in \tilde{N}_{[|]}(\delta_0), d_n(A_0, \tilde{A}_0) \leq 6\delta} |z_n(A_0) - z_n(\tilde{A}_0)| = 0.$$ 

This, however, readily follows from Step 3 and a similar reasoning as in the proof of Step 4, which renders the proof of (a) complete.

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

For proving tightness we use Theorem 2.11.9 in van der Vaart and Wellner (1996), a general bracketing central limit theorem. We will choose \( d \) for the semimetric \( \rho \) on \( B \), required in that theorem. For the 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 (10). It remains to show the second condition:

\[
s_n := \sup_{B, B' \in B, d(B, B') \leq \delta_n} \frac{1}{n a_n} \sum_{i=1}^{n} \text{E} \left( \frac{1}{\sqrt{n a_n}} 1_{\tau_n^{-1} B}(X_i) - \frac{1}{\sqrt{n a_n}} 1_{\tau_n^{-1} B'}(X_i) \right)^2 \to 0, \text{ for every } \delta_n \downarrow 0.
\]

But

\[
s_n = \frac{1}{n a_n} \sup_{d(B, B') \leq \delta_n} \sum_{i=1}^{n} \text{E} 1_{\tau_n^{-1} B \Delta \tau_n^{-1} B'}(X_i)
= \frac{1}{a_n} \sup_{d(B, B') \leq \delta_n} P\left(\tau_n^{-1}(B \Delta B') \leq \delta a_n \right) = \sup_{Q(B \Delta B') \leq \delta_n^2} Q_n(B \Delta B').
\]

Now Lemma 1 immediately yields \( s_n \to 0 \). \( \square \)

**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 \( n \)

\[
P\left( \sup_{\underset{d_n(A, \tilde{A}) \leq \sqrt{\delta}}{A \in A_n, \tilde{A} \in \tau_n^{-1} B}} |z_n(A) - z_n(\tilde{A})| > 2\eta \right) \leq 2\eta. \tag{13}
\]

We have for \( n \) large enough

\[
P\left( \sup_{\underset{d_n(A, \tilde{A}) \leq \sqrt{\delta}}{A \in A_n, \tilde{A} \in \tau_n^{-1} B}} |z_n(A) - z_n(\tilde{A})| > 2\eta \right) = P\left( \sup_{\underset{P(\Delta \tilde{A}) \leq \delta a_n}{A \in A_n, \tilde{A} \in \tau_n^{-1} B}} |z_n(A) - z_n(\tilde{A})| > 2\eta \right)
= P\left( \sup_{\underset{P(\Delta \tilde{A}) \leq \delta a_n}{A \in A_n, \tilde{A} \in \tau_n^{-1} B}} |z_n(A \setminus \tilde{A}) - z_n(\tilde{A} \setminus A)| > 2\eta \right)
\leq P\left( \sup_{\underset{P(\Delta \tilde{A}) \leq \delta a_n}{A \in A_n, \tilde{A} \in \tau_n^{-1} B}} |z_n(A \setminus \tilde{A})| > \eta \right) + P\left( \sup_{\underset{P(\Delta \tilde{A}) \leq \delta a_n}{A \in A_n, \tilde{A} \in \tau_n^{-1} B}} |z_n(\tilde{A} \setminus A)| > \eta \right)
\leq 2P\left( \sup_{\underset{P(C) \leq \delta a_n}{C \in C_n}} |z_n(C)| > \eta \right), \tag{14}
\]
where $C_n = \{ A \setminus \tilde{A} : A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_n^{-1}B \} \cup \{ \tilde{A} \setminus A : A \in \mathcal{A}_{\varepsilon_n}, \tilde{A} \in \tau_n^{-1}B \}$. It can be shown (see, e.g., van der Vaart and Wellner (1996), p. 147), using $A_1 \setminus A_2 = A_1 \cap A_2^c$, that $C_n$ 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 $k = na_n$,

$$
\mathbb{P}(\sup_{C \in C_n, P(C) \leq \delta a_n}|z_n(C)| > \eta) = \sum_{m=0}^{n} \mathbb{P}(\sup_{C \in C_n, P(C) \leq \delta a_n}|z_n(C)| > \eta | N = m) \mathbb{P}(N = m) = \sum_{m=0}^{n} \mathbb{P}\left(\sup_{C \in C_n, P(C) \leq \delta a_n} \left| \frac{1}{\sqrt{k}} |\Psi_n(C) - nP(C)| - \eta \right| N = m \right) \mathbb{P}(N = m) \leq \sum_{m=[k+C_n\sqrt{k}]}^{n} \mathbb{P}\left(\sup_{C \in C_n, P(C) \leq \delta a_n} \left| \frac{1}{\sqrt{k}} |\Psi_n(C) - nP(C)| - \eta \right| N = m \right) \mathbb{P}(N = m) + \mathbb{P}(\{ |N - k| \geq C_n\sqrt{k} \}),
$$

where $C_n$ 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 $n$

$$
\mathbb{P}(\sup_{C \in C_n, P(C) \leq \delta a_n}|z_n(C)| > \eta) \leq \sum_{m=[k-C_n\sqrt{k}]}^{n} \mathbb{P}\left(\sup_{C \in C_n, P_n(C) \leq \delta} \left| \frac{1}{\sqrt{m}} \sum_{j=1}^{m} 1_C(Y_j) - mP_n(C) \right| > \eta \right) \mathbb{P}(N = m) + \mathbb{P}(\{ |N - k| \geq C_n\sqrt{k} \}) \mathbb{P}(N = m) + \frac{\eta}{2}, \quad (15)
$$

where the $Y_i$ are i.i.d. random vectors on $\mathcal{V}_{\varepsilon_n}(\partial K)$ distributed according to $P_n$. 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 Corollary 2.9 in Alexander (1984). Using $\max_{n \in \mathbb{N}} w_n < \infty$, this leads to the following upper bound for the left hand side of (15):

$$
\sum_{m=[k+C_n\sqrt{k}]}^{n} 16 \exp(-\eta^2/(36\delta)) \mathbb{P}(N = m) + \frac{\eta}{2} \leq 16 \exp(-\eta^2/(36\delta)) + \frac{\eta}{2} \leq \eta,
$$
for small enough $\delta$. So because of (14) we have proved (13) and hence (a).

For a proof of (b), we only need to show tightness of $v_n, B$, 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 $\eta > 0$

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

(see, e.g., Theorem 1.5.7 in van der Vaart and Wellner (1996)). Again, from (5) and Lemma 1, it suffices to show that

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathbb{P}\left( \sup_{A, A' \in \tau - \frac{1}{\varepsilon n} B} d_n(A, A') \leq 2\delta \middle| z_n(B) - z_n(B') > \eta \right) = 0. \quad (16)$$

The proof of (16) can be given along the lines of the proof of (13).

\begin{proof}[Proof of Corollary 1]
Define

$$\varphi_n(x_n) = \sup_{B_n \in B_n} \frac{|x_n(B_n)|}{q_n(B_n)} \quad \text{and} \quad \varphi(x) = \sup_{B \in B} \frac{|x(B)|}{q(B)}.$$ 

We have to show that (11) implies $\varphi_n(x_n) \to \varphi(x)$.

Note that

$$\left| \sup_{B_n \in B_n} \frac{|x_n(B_n)|}{q_n(B_n)} - \sup_{B \in B} \frac{|x(B)|}{q(B)} \right| \leq \sup_{B_n \in B_n, B \in B} \left| \frac{x_n(B_n)}{q_n(B_n)} - \frac{x(B)}{q(B)} \right|,$$

and let us prove that the latter quantity tends to 0. Suppose it does not. Then we can find a subsequence $n_k$ such that for some $\varepsilon > 0$

$$\left| \frac{x_n(B_{n_k})}{q_n(B_{n_k})} - \frac{x(B'_{n_k})}{q(B'_{n_k})} \right| > \varepsilon,$$

with $B_{n_k} \in B_{n_k}, B'_{n_k} \in B$ and $d(B_{n_k}, B'_{n_k}) \leq \gamma_{n_k}$. Write $M = \sup_{B \in B} |x(B)|$ and $\delta = \inf_{B \in B} q(B)$. For large enough $k$ we have

$$|x_n(B_{n_k}) - x(B'_{n_k})| < \frac{\varepsilon \delta}{4} \quad \text{and} \quad |q_n(B_{n_k}) - q(B'_{n_k})| < \frac{\varepsilon^2 \delta}{4M}.$$
So
\[
\frac{|x_{n_k}(B_{n_k}) - x(B'_{n_k})|}{q_{n_k}(B_{n_k})} - \frac{x(B'_{n_k})}{q(B'_{n_k})} \leq \frac{|x_{n_k}(B_{n_k}) - x(B'_{n_k})|}{q_{n_k}(B_{n_k})} + \frac{|q_{n_k}(B_{n_k}) - q(B'_{n_k})|}{q_{n_k}(B_{n_k})q(B'_{n_k})} < \frac{1}{q_{n_k}(B_{n_k})} \left[ \frac{\varepsilon \delta}{4} + \frac{M \varepsilon \delta^2}{\delta} \frac{4 \delta^2}{4 M} \right] = \frac{1}{q_{n_k}(B_{n_k})} \frac{\varepsilon \delta}{2}.
\]

But for \( \varepsilon \) small enough, \( q_{n_k}(B_{n_k}) > \delta - \varepsilon \delta^2/(4M) > \delta/2 \), which yields that \( \varepsilon \delta/(2q_{n_k}(B_{n_k})) < \varepsilon \). Contradiction. \( \square \)

**References**


