NONPARAMETRIC ESTIMATION OF BOUNDARY MEASURES AND RELATED FUNCTIONALS: ASYMPTOTIC RESULTS

INÉS ARMENDÁRIZ,* ** *Universidad de San Andrés and Universidade de São Paulo*

ANTONIO CUEVAS,*** *Universidad Autónoma de Madrid*

RICARDO FRAIMAN,* **** *Universidad de San Andrés and Universidad de la República*

Abstract

We study a nonparametric method for estimating the boundary measure of a compact body $G \subset \mathbb{R}^d$ (the boundary length when $d = 2$ and the surface area for $d = 3$) in the case when this measure agrees with the corresponding Minkowski content. The estimator we consider is closely related to the one proposed in Cuevas, Fraiman and Rodríguez-Casal (2007). Our method relies on two sets of random points, drawn inside and outside the set G, with different sampling intensities. Strong consistency and asymptotic normality are obtained under some shape hypotheses on the set G. Some applications and practical aspects are briefly discussed.

Keywords: Minkowski content; nonparametric set estimation; statistical image analysis

2000 Mathematics Subject Classification: Primary 60D05; 62G07

1. Introduction

The volume and the boundary measure of a compact body $G \subset \mathbb{R}^d$, $d > 1$, are the most basic quantitative measurements of its size and shape. We will denote by Γ and $\mu(G)$ the boundary and the Lebesgue measure of G, respectively.

The measure of Γ is usually defined as its $(d-1)$-dimensional Hausdorff measure. We are interested in estimating this measure using a random sample of points, which will require that the set G satisfy certain regularity properties. We will therefore restrict ourselves to those sets whose Hausdorff boundary measure coincides with the *Minkowski content* defined by

$$L_0(G) = \lim_{\varepsilon \to 0} \frac{\mu(B(\Gamma, \varepsilon))}{2\varepsilon}, \tag{1}$$

when this limit exists and is finite (see [11, p. 79] for definitions and properties). In this equation, given a set A, $B(A, \varepsilon)$ stands for the parallel set $B(A, \varepsilon) = \bigcup_{x \in A} B(x, \varepsilon)$, if $B(x, \varepsilon)$ is the closed ball with center $x \in \mathbb{R}^d$ and radius $\varepsilon > 0$.

Although the Minkowski content is more restrictive than the Hausdorff measure, it has the advantage that its definition (1) is much simpler to handle in statistical applications, and yet it is well defined for the sets encountered in many practical situations.

In what follows, $G \subset \mathbb{R}^d$ will denote a compact body whose (boundary) Minkowski content $L_0 = L_0(G)$ is well defined and finite. We assume throughout that G is the support of an absolutely continuous probability measure. Without loss of generality, we will further assume that G is a subset of the open unit square $(0, 1)^d$. We will denote by G^c the complementary set of G relative to $[0, 1]^d$, i.e. $G^c = [0, 1]^d \setminus G$.

In fact, in this paper we will work with a one-sided version of the Minkowski content given by

$$L_0^-(G) = \lim_{\varepsilon \to 0} \frac{\mu(B(G^c, \varepsilon) \cap G)}{\varepsilon}. \tag{2}$$

Our estimation procedures will rely on sample data taken inside and outside the set G, and the use of the one-sided Minkowski content (2) will make it possible to work with different sampling intensities for the inner and outer data, a crucial factor in our approach.

There has recently been some interest in the one-sided Minkowski content (2) as a convenient mathematical tool to apply definition (1). Ambrosio *et al.* [1] provided a thorough study of this concept. They showed that when G has a Lipschitz boundary, $L_0(G) = L_0^-(G)$, and these in turn coincide with the $(d-1)$-dimensional Hausdorff measure of Γ (see Corollary 1 of [1]). See also Remark 4.9 of [15].

Our aim is to devise statistical procedures to estimate $L_0(G)$, and study their consistency and asymptotic distribution. Such results could then be applied to testing statistical hypotheses and the search of confidence intervals. Problems of this type have already been considered in stereology; see, e.g. [2, pp. 17–21], [3], and [10]. In the stereological approach the estimation usually relies on using lower-dimensional sections of G. In the methods proposed in this paper, on the other hand, the information is obtained from samples of randomly selected observations in $\mathbb{R}^d$, under the assumption that we are able to determine whether each sample point belongs to G. One advantage of this approach lies in the generality of the sets G it can be applied to. In particular, these are not assumed to be members of any finite-dimensional family; hence, the method is nonparametric. As is quite standard in nonparametric estimation, our techniques require the use of a smoothing parameter.

This paper is a further development of the ideas proposed in [8]; see also [6] and [14] for related approaches. Cuevas and Rodríguez-Casal [7] (see also the references therein) studied the problem of boundary approximation from a nonparametric perspective as well, but they did not provide results on the estimation of boundary measures.

The main contribution of this paper is the derivation of the asymptotic nondegenerate distribution for an estimator of $L_0(G)$. These types of result are not very usual in the field of set estimation theory, where most articles are devoted to establishing consistency and convergence rates. For results on asymptotic distributions, see, e.g. the papers of Bräker *et al.* [5], who computed the Hausdorff distance between a convex set and the convex hull of an inner random sample, and Molchanov [12] on plug-in estimation of level sets.

The problem of estimating boundary measures arises quite naturally in the analysis of two-dimensional images. For example, the contour index of an injured area has sometimes been used in oncology and cardiology as an auxiliary diagnostic tool to assess the injury malignity. This index combines the volume and boundary measures of the set G in the ratio

$$C_0(G) = \frac{L_0(G)}{\mu(G)^{(d-1)/d}},$$

providing an adimensional measurement of how complicated the boundary of G is. The quantities $L_0 = L_0(G)$ and $C_0 = C_0(G)$ are clearly very closely related; we propose here

a consistent estimator for C_0 and derive its asymptotic normality from the analogous result for the estimator of L_0. We refer the reader to [8], where a case study in cardiology is described in detail using a variation of the approach considered in the present paper.

Another possible area of application has recently come to our attention in astrophysics, in connection with the Wilkinson microwave anisotropy probe. The goal of this project is to measure the temperature differences in the cosmic microwave background (CMB), an electromagnetic radiation whose discovery in 1965 is considered to provide strong support for the big bang cosmological model. Some maps have already been drawn identifying the cold and hot CMB regions (see [4]), and an important tool in the analysis of such maps is the use of Minkowski functionals, which include the area and boundary length. The statistical approach followed here (combined with a Monte Carlo methodology similar to that employed in the case study in [8]) could hopefully be applied to the estimation of these Minkowski functionals.

This paper is organized as follows. In Section 2 we present the sampling model, define the basic estimators, and introduce the required hypotheses. We state and discuss the main results in Section 3, and give the proofs in Section 4.

2. Model and hypotheses

2.1. Sampling model and estimators

Our goal is to estimate $L_0(G)$ and the contour index $C_0(G)$ from two independent samples of random variables $\{X_i\}$ and $\{Y_i\}$, drawn according to the following design, which we will refer to as model 2 (M2):

$\{X_i\}_{1 \leq i \leq n}$ are independent and identically uniformly distributed in $G^c = [0, 1]^d \setminus G$

and $\{Y_i\}_{1 \leq i \leq m}$ are independent and identically uniformly distributed in G.

The sample sizes m and n are chosen in advance, fulfilling some conditions to be specified below. This sampling design is closely related to that applied to estimate $L_0(G)$ in [8]: under that design, referred to here as model 1 (M1), the information is drawn from a sequence of independent and identically distributed observations $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$ of a random variable (Z, δ), where Z is uniformly distributed in $[0, 1]^d$. The parameter δ takes binary values: $\delta = 1$ if $Z \in G$ and $\delta = 0$ if $Z \notin G$; it is therefore assumed that it is possible to decide whether any given observation belongs to the set G.

In practice, we could use a sequential version, M1′, of M1 to obtain the data in M2: we just need to draw the data (Z_i, δ_i) sequentially, setting $Z_i = X_i$ if $\delta_i = 0$ and $Z_i = Y_i$ otherwise, until we have obtained m observations in G and n observations in G^c. It is then necessary to consider a random number of observations $N = N(n, m) \geq n + m$. Hence, the difference between M2 and M1′ is rather formal: under M2, we consider just the two independent samples, with sizes m and n, on G and G^c, whereas under M1′, we obtain $N - m - n$ extra observations in $[0, 1]^d$, as a by-product of the sequential process employed in drawing the samples. In most real cases we have a criterion, typically following from some sort of analysis, to decide whether a given observation $Z_i \in [0, 1]^d$ belongs to G, and then M1′ is the natural mechanism to obtain the data required by M2.

The sampling schemes M1 and M2 share an essential feature: they assume that positive fractions of the sample points fall both inside and outside G. This suffices to prove consistency and even to obtain convergence rates (see [8]) under some shape conditions on G. In the present paper we establish the asymptotic normality of the considered estimators of $L_0(G)$ and $C_0(G)$. The task becomes much easier by requiring that the inner and outer samples have different

intensities. This is the reason for the replacement of the original sampling scheme M1 with the more flexible version M2: the latter allows us to control the relative intensities (n and m) of the samples. For notational convenience, we will often omit the subindex m in the notation for the estimators, even though these still depend on the sizes of both samples, n and $m(n)$.

We now establish some notation necessary to define the estimators for $L_0(G)$ and $C_0(G)$. Given a Borel set $A \subseteq [0, 1]^d$, let $N_m(A; G)$ be the counting measure associated with the Y sample. That is, $N_m(A; G)$ is the number of Y points that belong to A:

$$N_m(A; G) = \#\{j \geq 1 \colon Y_j \in A\}.$$

We will denote by

$$\mu_{m,G}(A) = \mu(G)\frac{N_m(A; G)}{m}$$

the renormalized version of this empirical measure: $\mu_{m,G}(A)$ 'estimates' (if $\mu(G)$ were known) the Lebesgue measure of the intersection $A \cap G$.

Our estimators will depend on a sequence $\{\varepsilon_n\}$ of positive smoothing parameters converging to 0 at a suitably slow rate, which we describe in the set of hypotheses below. They also rely on the nonparametric estimator H_n of G^c given by the parallel set of the outer sample:

$$H_n = B(\{X_1, \ldots, X_n\}, \varepsilon_n).$$

We are ready to introduce the estimators $\hat{L}_n = \hat{L}_{n,m}(G)$ of $L_0^-(G)$ and $\hat{C}_n$ of $C_0(G)$. First, we define an auxiliary statistic

$$A_n = \frac{\mu_{m,G}(H_n)}{\varepsilon_n}, \tag{3}$$

which will be a true estimator only in the unrealistic case where $\mu(G)$ is known. The numerator in (3) estimates the ε_n-'inwards dilation' of Γ by measuring H_n in terms of the inner sample points Y_j. Typically, the Lebesgue measure $\mu(G)$ is an unknown quantity, which must be estimated from the data as well. Under M1', a natural estimator of this value is given by

$$\hat{\mu}(G) = \frac{\#\{i \colon \delta_i = 1\}}{N},$$

which is just the relative frequency of observations falling inside G, when the total size of the sample is the random quantity $N = N(n, m)$. Since the Z sample is uniformly distributed in $[0, 1]^d$, $\hat{\mu}(G)$ is consistent by the law of large numbers. The estimator $\hat{L}_n$ is defined by replacing the value $\mu(G)$ by $\hat{\mu}(G)$ in (3). We thus obtain

$$\hat{L}_n = \hat{\mu}(G)\frac{N_m(H_n; G)}{m\varepsilon_n}.$$

In this setting, the natural choice for the estimator $\hat{C}_n$ is given by

$$\hat{C}_n = \frac{\hat{L}_n}{\hat{\mu}(G)^{(d-1)/d}}.$$

2.2. Hypotheses

We now state the main hypotheses to be used in our results. The first three are required for the consistency results (see Theorem 1, below). The remaining hypotheses are necessary for the derivation of the asymptotic normality (see Theorem 2, below).

(H0) The Minkowski content L_0 satisfies $L_0(G) = L_0^-(G) < \infty$.

As mentioned in the introduction, the relations between $L_0(G)$, $L_0^-(G)$, and the Hausdorff $(d-1)$-dimensional measure of G are discussed in detail in [1, Theorems 1, 2 and Corollary 1]; see also [15]. Note that all the discussion below could be alternatively written, with obvious changes, in terms of the lateral Minkowski content $L_0^+(G)$ defined in a dual way by just interchanging the roles of G and G^c.

(H1) The set G^c is standard in Γ. That is, there exists a positive constant K such that, for any small enough η, $P(X_1 \in B(x, \eta)) \geq K\mu(B(x, \eta))$ for all $x \in \Gamma$.

We recall that the variable X_1 above has uniform distribution in G^c. Condition (H1) is quite usual in set estimation theory; see, e.g. [7] and the references therein. It can be seen as a regularity assumption on the shape of the set G^c which rules out the existence of sharp peaks in the boundary.

(H2) The sequences $\{\varepsilon_n\}$ and $m = m(n)$ satisfy

$$\varepsilon_n \to 0, \qquad \frac{m\varepsilon_n}{\log n} \to \infty, \quad \text{and} \quad \frac{n\varepsilon_n^d}{\log n} \to \infty, \quad \text{as } n \to \infty.$$

This condition is analogous to that imposed on the sequence of smoothing parameters in the classical theory of nonparametric density estimation in order to obtain strong uniform consistency in kernel density estimators (see, e.g. [14]). It concerns the rate at which the smoothing parameters converge to 0.

(H3) $|L(\varepsilon) - L_0| = O(\varepsilon)$ as $\varepsilon \to 0$, where $L(\varepsilon) = \mu(B(G^c, \varepsilon) \cap G)/\varepsilon$.

This is a further regularity hypothesis on the set G. It clearly holds when $L(\varepsilon)$ is differentiable at $\varepsilon = 0$.

The reach of a closed set $S \subset \mathbb{R}^d$ is defined as the largest (possibly ∞) value r such that if $x \in \mathbb{R}^d$ is at a distance less than r from S then this distance is realized at a unique point in S. A generalization of Steiner's formula concerning the parallel sets of convex bodies (see [9, Theorem 5.6]) proves that when (H0) holds and G^c satisfies $\text{reach}(\overline{G^c}) > 0$, then the function $L(\varepsilon)$ is given, for $0 \leq \varepsilon < \text{reach}(\overline{G^c})$, by a polynomial whose independent term is L_0, and in particular (H3) holds. Moreover, the condition $\text{reach}(\overline{G^c}) > 0$ turns out to imply both the finiteness of the one-sided Minkowski content $L_0^-(G)$ (see the comments after Theorem 7 of [1]) and the equality between $L_0(G)$ and the corresponding $(d-1)$-dimensional Hausdorff measure (see Corollary 3 of [1]). A related discussion can be found in Subsection 3.2 of [8].

(H4) The sequences $m = m(n)$ and ε_n satisfy $m\varepsilon_n^3 \to 0$ as $n \to \infty$.

(H5) $\dfrac{1}{n} \left(\dfrac{m}{\varepsilon_n} \right)^{d/2} \log \left(\dfrac{m}{\varepsilon_n} \right) \to 0 \quad$ as $n \to \infty$.

Hypotheses (H4) and (H5) impose additional constraints on the rate of increase of $m(n)$ and the speed of convergence to 0 of the smoothing parameters ε_n. Note that (H5) implies that the

size n of the outer sample is of a larger order than that of the inner sample, m. This hypothesis is essential to our proof of the asymptotic normality (Theorem 2, below): the result follows from an application of the standard central limit theorem for independent and identically distributed random variables, once we condition on the set $H_n \cap G$. The fact that $n \gg m$ means that when we work at the scaling required to obtain the limiting distribution of the estimator, we are just perceiving the fluctuations in the spatial distribution of the inner sample in such a way that, for the purposes of determining the asymptotic distribution of the estimators, the set H_n may be considered as deterministic.

We can take $m = n^\alpha$ and $\varepsilon_n = n^{-\beta}$, and look for the values of $\alpha, \beta > 0$ that fulfill the conditions imposed on $m(n)$ and ε_n. A direct computation shows that, in this case, (H2), (H4), and (H5) are satisfied whenever

$$\alpha + \beta < \frac{2}{d} \quad \text{and} \quad \beta < \alpha < 3\beta. \tag{4}$$

3. Asymptotic results

We first establish the almost sure (a.s.) consistency of our estimators.

Theorem 1. *Under (H0), (H1), and (H2) and the sampling model M2 (i.e. independent samples of sizes m and n), we have $A_n \to L_0$ a.s. as $n \to \infty$.*
The same result holds for the estimator $\hat{L}_n$ under the sampling model M1$'$.

A similar consistency result is given in [8] under the sampling model M1.

As the estimator $\hat{C}_n$ for the contour index C_0 is obtained by just dividing $\hat{L}_n$ by $\hat{\mu}(G)^{(d-1)/d}$, the following result is immediate from Theorem 1.

Corollary 1. *Under (H0), (H1), and (H2) and the sampling model M1$'$, we obtain*

$$\hat{C}_n \to C_0 \quad \text{a.s. as } n \to \infty.$$

Our main result is the asymptotic normality established in the next theorem. We denote by $N(a, \sigma^2)$ the normal distribution with mean a and variance σ^2, and denote by '$\overset{w}{\to}$' the weak convergence.

Theorem 2. *Under the set of assumptions (H0)–(H5) and the sampling model M2, we have*

$$\sqrt{m\varepsilon_n}(A_n - L_0) \overset{w}{\to} N(0, \mu(G)L_0) \quad \text{as } n \to \infty, \tag{5}$$

and, under M1$'$, we have

$$\sqrt{m\varepsilon_n}(\hat{L}_n - L_0) \overset{w}{\to} N(0, \mu(G)L_0) \quad \text{as } n \to \infty. \tag{6}$$

As a direct consequence of Theorem 2 (see the comment preceding the statement of Corollary 1), we have the following result.

Corollary 2. *Under assumptions (H0)–(H5) and the sampling model M1$'$, we obtain*

$$\sqrt{m\varepsilon_n}(\hat{C}_n - C_0) \overset{w}{\to} N(0, \sigma^2) \quad \text{as } n \to \infty, \tag{7}$$

where $\sigma^2 = L_0\mu(G)^{(2-d)/d}$.

A few comments about the results are necessary.

The convergence rate. Set $m = n^\alpha$ and $\varepsilon_n = n^{-\beta}$ as discussed in the previous section. The conditions in (4) then imply that the convergence in (5), (6), and (7) can be realized at any rate slower than $n^{1/2d}$ by making a convenient choice of the parameters α and β. Indeed, the fastest rate for $m\varepsilon_n$ subject to $\alpha + \beta < 2/d$ and $\alpha < 3\beta$ corresponds to a choice of α very close to 3β, where $\beta = 1/2d$, which yields $\sqrt{m\varepsilon_n} \sim n^{1/2d}$. This coincides with the best rate found in [8] for the estimation of $L_0(G)$ under M1.

A faster convergence rate is obtained in [14], assuming additional shape conditions on G which are incorporated into the estimator. Our procedure yields rates that come arbitrarily close to $n^{1/2d} = m^{1/3}$; see the suggested values for the parameters m and n in (4). If no extra regularity assumptions on the set G are imposed, we do not expect that the rate $n^{1/2d}$ can be substantially improved. A simple informal argument illustrating this follows. Consider an ideal situation in which a set G fulfilling (H3) is completely known. In this case the parallel set $B(G^c, \varepsilon_n)$ is part of the data, and a natural estimator for L_0 is given by

$$\tilde{L}_n = \frac{1}{m\varepsilon_n} \sum_{j=1}^{m} \mathbf{1}_{\{Y_j \in B(G^c, \varepsilon_n) \cap G\}}.$$

A straightforward calculation shows that the mean-square error $\mathrm{E}[(\tilde{L}_n - L_0)^2]$ is of order $\varepsilon_n^2 + 1/m\varepsilon_n$, which leads to an optimal convergence rate of order $\sqrt{m\varepsilon_n} = m^{1/3}$ for $\tilde{L}_n$. Hence, this seems to be the best attainable performance.

The statistical use of asymptotic normality. We briefly illustrate the application of these results to problems arising in cardiology or oncology, where the contour index is used as a diagnostic tool. Similar applications to the computation of the boundary length, or other quantities that can be derived from this, follow in a similar way.

Theorem 2 and Corollary 2 can be applied to determine from an estimate C_n of C_0 whether there is enough statistical evidence to decide that the true C_0 is larger than a prescribed value C_0^*. According to (7), we may accept the alternative hypothesis $C_0 > C_0^*$, at an approximate significance level α, whenever

$$\sqrt{m\varepsilon_n}\,\frac{(\hat{C}_n - C_0^*)}{\hat{\sigma}_n} > z_\alpha,$$

where $\hat{\sigma}_n^2 = \hat{L}_n \hat{\mu}(G)^{(2-d)/d}$ and z_α is the $100(1-\alpha)\%$ percentile in the $N(0, 1)$ distribution.

Suppose, on the other hand, that it is necessary to compare two images and decide whether there is sufficient statistical evidence in favor of a relationship $C_0^{(1)} > C_0^{(2)}$ between the two contour indices. Using the estimators $\hat{C}_n^{(1)}$, $\hat{C}_n^{(2)}$, $\hat{\sigma}_n^{(1)}$, and $\hat{\sigma}_n^{(2)}$ of the contour indices and asymptotic standard deviations, we could then consider a classical two-sample test with approximate α-critical region given by

$$\frac{\sqrt{m\varepsilon_n}(\hat{C}_n^{(1)} - \hat{C}_n^{(2)})}{\sqrt{[\hat{\sigma}_n^{(1)}]^2 + [\hat{\sigma}_n^{(2)}]^2}} > z_\alpha.$$

It is clear that this analysis deals only with the statistical variability due to the sampling variation in the estimators. When applying these techniques to estimate the contour index of a set which has in turn been randomly chosen among some population, an additional stochastic model should be considered in order to account for the inter-individual variability of C_0.

4. Proofs

In the sequel, P will denote the measure on the space $([0, 1]^d)^n \times ([0, 1]^d)^m$ determined by the sequence $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$, and E will stand for the expectation taken with respect to P. Hence, P is the product of $n + m$ probability measures: the first n are uniformly distributed in G^c and the remaining m are uniformly distributed in G.

We will need to use several auxiliary constants in the following proofs, most often as part of chains of inequalities. These constants will be independent of the parameters n and m, which will eventually tend to ∞. They will be all denoted by the letter C, provided that this does not lead to confusion.

4.1. Proof of Theorem 1

Recall that the statistic A_n is given by $A_n = \mu_{m,G}(H_n)/\varepsilon_n$. We will henceforth use the acronym i.o. to mean infinitely often.

The theorem will follow if we can show that, for any $\delta > 0$,

$$P\left(\left| \frac{\mu_{m,G}(H_n)}{\varepsilon_n} - L_0 \right| > \delta \text{ i.o.} \right) = 0.$$

We have

$$\left| \frac{\mu_{m,G}(H_n)}{\varepsilon_n} - L_0 \right| \leq \left| \frac{\mu_{m,G}(H_n)}{\varepsilon_n} - \frac{\mu(H_n \cap G)}{\varepsilon_n} \right| + \left| \frac{\mu(H_n \cap G)}{\varepsilon_n} - \frac{\mu(B(G^c, \varepsilon_n) \cap G)}{\varepsilon_n} \right|$$
$$+ \left| \frac{\mu(B(G^c, \varepsilon_n) \cap G)}{\varepsilon_n} - L_0 \right|$$
$$=: T_1(n, m) + T_2(n) + T_3(n). \tag{8}$$

We study these terms separately in the following lemmas.

Lemma 1. *Under (H0)–(H2),*

$$P\left(T_1(n, m) > \frac{\delta}{3} \text{ i.o.} \right) = P\left(\left| \frac{\mu_{m,G}(H_n)}{\varepsilon_n} - \frac{\mu(H_n \cap G)}{\varepsilon_n} \right| > \frac{\delta}{3} \text{ i.o.} \right) = 0.$$

Proof. We first condition on the σ-algebra generated by the sample $\{X_i\}$, so that in particular the set H_n is fixed.

The inequality

$$P\left(T_1(n, m) \geq \frac{\delta}{3} \text{ i.o.} \right) \leq \lim_{n_0 \to \infty} \sum_{n=n_0}^{\infty} P\left(T_1(n, m) \geq \frac{\delta}{3} \right) \tag{9}$$

implies that the left-hand side vanishes: $P(T_1(n, m) \geq \delta/3 \text{ i.o.}) = 0$ whenever the series on the right-hand side of (9) is convergent.

Consider the sequence of random variables

$$\alpha_j = \mathbf{1}_{\{Y_j \in H_n\}} - \frac{\mu(H_n \cap G)}{\mu(G)}, \qquad 1 \leq j \leq m,$$

and note that

$$E[\alpha_j \mid H_n] = 0, \qquad E[\alpha_j^2 \mid H_n] = \frac{\mu(H_n \cap G)}{\mu(G)} - \left(\frac{\mu(H_n \cap G)}{\mu(G)} \right)^2.$$

We have

$$P\left(T_1(n,m) \geq \frac{\delta}{3}\right) = P\left(\left|\sum_{j=1}^{m} \alpha_j\right| \geq \frac{m\varepsilon_n}{3\mu(G)}\delta\right) = E\left[P\left(\left|\sum_{j=1}^{m} \alpha_j\right| \geq \frac{m\varepsilon_n}{3\mu(G)}\delta \,\middle|\, H_n\right)\right]$$

in terms of the sequence $\{\alpha_j\}$. Using Bernstein's inequality, the last expression above can be bounded by

$$P\left(\left|\sum_{j=1}^{m} \alpha_j\right| \geq \frac{m\varepsilon_n}{3\mu(G)}\delta \,\middle|\, H_n\right) \leq 2\exp\{-C(G,\delta)m\varepsilon_n\}.$$

In order to derive the last inequality, we used the fact that $\mu(H_n \cap G) \leq \mu(B(G^c, \varepsilon_n) \cap G) = O(\varepsilon_n)$, which holds by hypothesis (H0). The bound does not depend on H_n; hence, it also applies to the unconditional probability. The lemma now follows from assumption (H2) on the relation between ε_n and m.

Lemma 2. *Under (H0)–(H2),*

$$P\left(T_2(n) > \frac{\delta}{3} \ i.o.\right) = P\left(\left|\frac{\mu(H_n \cap G)}{\varepsilon_n} - \frac{\mu(B(G^c, \varepsilon_n) \cap G)}{\varepsilon_n}\right| > \frac{\delta}{3} \ i.o.\right) = 0.$$

Proof. As in the proof of Lemma 1, it will be enough to show that the series with the nth term given by $P(T_2(n) \geq \delta/3)$ is summable. Then let

$$R_n = (B(G^c, \varepsilon_n) \cap G) \setminus (H_n \cap G).$$

We first remark that in the event $T_2(n) > \delta/3$, or, equivalently,

$$\mu(R_n) \geq \frac{\delta\varepsilon_n}{3}, \tag{10}$$

there exists $z \in \Gamma$ (the boundary of G) such that

$$B(z, r\varepsilon_n) \cap \{X_1, \ldots, X_n\} = \varnothing \tag{11}$$

for large enough n and $r = \delta/6L_0$. We assume at this point that δ is small enough so that $r < 1$. Indeed, from assumption (H0) we have

$$\mu(B(G^c, \varepsilon) \cap G) = \mu(B(\Gamma, \varepsilon) \cap G) = \varepsilon L_0 + \rho(\varepsilon), \qquad 0 \leq \varepsilon \leq \varepsilon_n, \tag{12}$$

where the residue term $\rho(\cdot)$ satisfies $\rho(\varepsilon) = o(\varepsilon)$ as $\varepsilon \to 0$. If (11) does not hold, we then have $\Gamma \subseteq B(\{X_1, \ldots, X_n\}, r\varepsilon_n)$, from where it follows that the parallel set of Γ of radius $(1-r)\varepsilon_n$ must satisfy the relation $B(\Gamma, (1-r)\varepsilon_n) \cap G \subseteq H_n \cap G$. In particular,

$$\mu(H_n \cap G) \geq (1-r)\varepsilon_n L_0 + \rho((1-r)\varepsilon_n), \tag{13}$$

where the function $\rho(\cdot)$ is as in (12).

However, (10) and (13) are incompatible. In order to see this, note that the sets $H_n \cap G$ and R_n are disjoint by construction, and that $(H_n \cap G) \cup R_n = B(G^c, \varepsilon_n) \cap G$. We take measures in this set equality and apply the bounds in (10) and (13) to conclude that, with our choice of r,

$$\mu(B(G^c, \varepsilon_n) \cap G) = \mu(H_n \cap G) + \mu(R_n)$$

$$\geq \frac{\delta\varepsilon_n}{3} + (1-r)\varepsilon_n L_0 + \rho((1-r)\varepsilon_n)$$

$$\geq \varepsilon_n L_0 + \frac{\delta\varepsilon_n}{6} + \rho((1-r)\varepsilon_n),$$

which violates (12). Hence, (11) follows.

Let M be the minimal cardinality of a covering of Γ of type $\{B(z_i, r\varepsilon_n/2)\colon z_i \in \Gamma\}$. Clearly, $M = O(1/\varepsilon_n^d)$ since the minimal number of balls of radius $r\varepsilon_n/4$ needed to cover $[0, 1]^d$ is an upper bound for M.

Choose $z \in \Gamma$ as in (11), and let i_0, $1 \le i_0 \le M$, be such that $z \in B(z_{i_0}, r\varepsilon_n/2)$. Then $B(z_{i_0}, r\varepsilon_n/2) \subseteq B(z, r\varepsilon_n)$, and, therefore, $B(z_{i_0}, r\varepsilon_n/2) \cap \{X_1, \ldots, X_n\} = \varnothing$. We conclude that $\mathrm{P}(T_2(n) \ge \delta/3)$ can be bounded by the probability of the event that none of the n independent, uniformly distributed points in G^c fall into one of the balls $B(z_j, r\varepsilon_n/2)$. By (H1), $\mathrm{P}(X_1 \in B(z_j, r\varepsilon_n/2)) \ge K\omega_d(r\varepsilon_n/2)^d$, where ω_d is the volume of the d-dimensional unitary ball. We obtain

$$
\begin{aligned}
\mathrm{P}\left(T_2(n) \ge \frac{\delta}{3}\right) &\le M \max_j \mathrm{P}\left(B\left(z_j, \frac{r\varepsilon_n}{2}\right) \cap \{X_1, \ldots, X_n\} = \varnothing\right) \\
&= M \max_j \mathrm{P}\left(X_1 \notin B\left(z_j, \frac{r\varepsilon_n}{2}\right)\right)^n \\
&\le M(1 - C(d, G)(r\varepsilon_n)^d)^n \\
&\le \exp\{-C(d, G, \delta)n\varepsilon_n^d\}.
\end{aligned}
$$

The series with the nth term given by the expression on the right-hand side above is summable due to (H2). This completes the proof of Lemma 2.

It remains to control the third term in expression (8). Clearly, from (H0),

$$
T_3(n) = \left|\frac{\mu(B(G^c, \varepsilon_n) \cap G)}{\varepsilon_n} - L_0\right| \le \frac{\delta}{3} \quad \text{for sufficiently large } n. \tag{14}
$$

The proof of the first statement in Theorem 1 follows from (8) together with Lemmas 1 and 2 and (14).

In order to see that $\hat{L}_n \to L_0$ a.s. as $n \to \infty$ as well, it suffices to observe that in $\hat{L}_n$ the true value $\mu(G)$ has been replaced by the consistent estimator $\hat{\mu}(G)$. This completes the proof of Theorem 1.

4.2. Proof of Theorem 2

The proof of (5) follows the steps of the proof of Theorem 1. Briefly, we will decompose the difference $|\mu_{m,G}(H_n) - L_0|$ as in (8), and study each term separately. The first term will yield the central limit theorem we are looking for, which will follow in the standard way by conditioning on the set H_n. The second and third terms in (8) are negligible. In order to prove this assertion for the second term, we will show that when it does not hold, we can find a (random) set in G^c with the property that it does not intersect the sample $\{Y_i\}$, and that the probability of this event can be easily controlled. The third term is deterministic, the result is straightforward from the assumptions on G.

As some of the arguments to be presented now are similar to those employed in the proofs of Lemmas 1 and 2, we give an outline here and refer the reader to those lemmas for details.

Lemma 3. *Under (H0)–(H2),*

$$
\sqrt{m\varepsilon_n}\left(\frac{\mu_{m,G}(H_n)}{\varepsilon_n} - \frac{\mu(H_n \cap G)}{\varepsilon_n}\right) \xrightarrow{\text{w}} N(0, \mu(G)L_0) \quad \text{as } n \to \infty.
$$

Proof. We compute the Laplace transform of the expression on the left-hand side. Let $\gamma > 0$. We have

$$\mathrm{E}\left[\exp\left\{-\gamma\sqrt{m\varepsilon_n}\left(\frac{\mu_{m,G}(H_n)}{\varepsilon_n} - \frac{\mu(H_n \cap G)}{\varepsilon_n}\right)\right\}\right]$$

$$= \mathrm{E}\left[\mathrm{E}\left[\exp\left\{-\frac{\tilde{\gamma}}{\sqrt{m\varepsilon_n}}(\mathbf{1}_{\{Y_1 \in H_n\}} - \beta_n)\right\}\ \bigg|\ H_n\right]^m\right]$$

$$= \mathrm{E}\left[\left(1 + \frac{1}{2}\frac{\tilde{\gamma}^2}{m\varepsilon_n}(\beta_n - \beta_n^2) + O\left[\frac{1}{(m\varepsilon_n)^{3/2}}\right]\right)^m\right], \tag{15}$$

where $\tilde{\gamma} = \mu(G)\gamma$ and $\beta_n = \mu(H_n \cap G)/\mu(G) = \mathrm{E}[\mathbf{1}_{\{Y_i \in H_n\}}\ |\ H_n]$. The last line in (15) follows by expanding the exponential term in its Taylor series up to the second order, and taking conditional expectations. Note that since $|\mathbf{1}_{\{Y_i \in H_n\}} - \beta_n| \leq 1$, the remainder term in the expansion is uniformly bounded. On the other hand, $\beta_n/\varepsilon_n \to L_0/\mu(G)$ a.s. by Lemma 2 and (14), and the difference can be bounded uniformly as well. We apply the bounded convergence theorem to conclude the result.

Lemma 4. *Under (H0)–(H5),*

$$\sqrt{m\varepsilon_n}\left(\frac{\mu(H_n \cap G)}{\varepsilon_n} - \frac{\mu(B(G^{\mathrm{c}}, \varepsilon_n) \cap G)}{\varepsilon_n}\right) \xrightarrow{\mathrm{w}} 0 \quad as\ n \to \infty.$$

Proof. We just need to slightly adapt the proof of Lemma 2 and apply the extra assumptions. We maintain the same notation from this lemma.

The result will follow if we show that $\sqrt{m/\varepsilon_n}\mu(R_n) \to 0$ in probability as $n \to \infty$. Then let $\delta > 0$, and suppose that

$$\frac{\sqrt{m}}{\sqrt{\varepsilon_n}}\mu(R_n) \geq \delta. \tag{16}$$

Define $\delta_n = \delta/2\sqrt{m\varepsilon_n}$. Assumptions (H2) and (H3) now imply that

$$\mu(B(G^{\mathrm{c}}, \varepsilon_n(1 - \delta_n)) \cap G) = L_0\varepsilon_n(1 - \delta_n) + O(\varepsilon_n^2),$$

and, hence, in the event that (16) holds, there exists $z \in \Gamma$ such that

$$B(z, \varepsilon_n\delta_n) \cap \{X_1, \ldots, X_n\} = \varnothing.$$

We now consider a covering of Γ consisting of M balls $\{B_j\}_{1 \leq j \leq M}$ with radii $\varepsilon_n\delta_n/2$, centered at points $\{z_j\}_{1 \leq j \leq M}$ in Γ. It is easy to see that the centers can be chosen so that $M = O(1/(\varepsilon_n\delta_n)^d)$. Finally, if j_0 is such that $\|z - z_{j_0}\| \leq \varepsilon_n\delta_n/2$ then $B_{j_0} \cap \{X_1, \ldots, X_n\} = \varnothing$. The problem is thus reduced to estimating the probability of this last event,

$$\mathrm{P}\left(\frac{\sqrt{m}}{\sqrt{\varepsilon_n}}\mu(R_n) \geq \delta\right) \leq \mathrm{P}\left(\bigcup_{j=1}^{M}\{\{X_1, \ldots, X_n\} \cap B_j = \varnothing\}\right)$$

$$\leq M \max_j \mathrm{P}(\{X_1, \ldots, X_n\} \cap B_j = \varnothing)$$

$$\leq C(G)\left(\frac{m}{\varepsilon_n}\right)^{d/2}\left(1 - C(d, G)\left(\frac{\varepsilon_n}{m}\right)^{d/2}\right)^n$$

$$\to 0 \quad \text{as } n \to \infty,$$

after replacing M and δ_n by their values and applying conditions (H1) and (H5). This completes the proof of Lemma 4.

Let us finally note that (H4) implies that

$$\sqrt{m\varepsilon_n}\, T_3(n) = \sqrt{m\varepsilon_n}\left|\frac{\mu(B(G^{\mathrm{c}}, \varepsilon_n) \cap G)}{\varepsilon_n} - L_0\right| \le \delta$$

if n is large enough. Together with Lemmas 3 and 4, this concludes the proof of (5).

In order to prove (6), it suffices to note that the difference $\hat{\mu}(G) - \mu(G)$ is of order (in probability)

$$\frac{1}{\sqrt{N}} \le \frac{1}{\sqrt{n+m}} \ll \frac{1}{\sqrt{m\varepsilon_n}};$$

hence, we can replace A_n by $\hat{L}_n$ in (5) while keeping the convergence rate and the asymptotic distribution.

Acknowledgements

References

[1] AMBROSIO, L., COLESANTI, A. AND VILLA, E. (2008). Outer Minkowski content for some classes of closed sets. *Math. Ann.* **342,** 727–748.

[2] BADDELEY, A. J. AND VEDEL-JENSEN, E. B. V. (2005). *Stereology for Statisticians* (Monogr. Statist. Appl. Prob. **103**). Chapman and Hall, Boca Raton, FL.

[3] BADDELEY, A. J., GUNDERSEN, H. J. G. AND CRUZ-ORIVE, L. M. (1986). Estimation of surface area from vertical sections. *J. Microscopy* **142,** 259–276.

[4] BENNETT, C. L. *et al.* (2003). First-year Wilkinson microwave anisotropy probe (WMAP) observations: preliminary maps and basic results. *Astro. J. Sup. Ser.* **148,** 1–27.

[5] BRÄKER, H., HSINGH, T. AND BINGHAM, N. H. (1998). On the Hausdorff distance between a convex set and an interior random convex hull. *Adv. Appl. Prob.* **30,** 295–316.

[6] CARLSTEIN, E. AND KRISHNAMOORTHY, C. (1992). Boundary estimation. *J. Amer. Statist. Assoc.* **87,** 430–438.

[7] CUEVAS, A. AND RODRÍGUEZ-CASAL, A. (2004). On boundary estimation. *Adv. Appl. Prob.* **36,** 340–354.

[8] 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.

[9] FEDERER, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93,** 418–491.

[10] GOKHALE, A. M. (1990). Unbiased estimation of curve length in 3D using vertical slices. *J. Microscopy* **159,** 133–141.

[11] MATTILA, P. (1995). *Geometry of Sets and Measures in Euclidean Spaces* (Camb. Stud. Adv. Math. **44**). Cambridge University Press.

[12] MOLCHANOV, I. S. (1998). A limit theorem for solutions of inequalities. *Scand. J. Statist.* **25,** 235–242.

[13] PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic Press, New York.

[14] PATEIRO-LÓPEZ, B. AND RODRÍGUEZ-CASAL, A. (2008). Length and surface area estimation under smoothness restrictions. *Adv. Appl. Prob.* **40,** 348–358.

[15] VILLA, E. (2009). On the outer Minkowski content of sets. To appear in *Ann. Mat. Pura Appl.*