NONPARAMETRIC ESTIMATION OF SURFACE INTEGRALS

BY RAÚL JIMÉNEZ* AND J. E. YUKICH†

Universidad Carlos III de Madrid and Lehigh University

The estimation of surface integrals on the boundary of an unknown body is a challenge for nonparametric methods in statistics, with powerful applications to physics and image analysis among other fields. Provided one can determine whether random shots hit the body, Cuevas et al. [*Ann. Statist.* **35**:1031–1051] estimate the boundary measure (the boundary length for planar sets and the surface area for tridimensional objects) via consideration of shots at a box containing the body. The statistics considered by these authors, as well as those in subsequent papers, are based on the estimation of Minkowski content and depend on a smoothing parameter which must be carefully chosen. For the same sampling scheme, we introduce a new approach which bypasses this issue, providing strongly consistent estimators of both the boundary measure, together with the surface integral of scalar functions, provided one can collect the function values at the sample points. Examples arise in experiments in which the density of the body can be measured by physical properties of the impacts or in situations where such quantities as temperature and humidity are observed by randomly distributed sensors. Our method is based on random Delaunay triangulations and involves a simple procedure for surface reconstruction from a dense cloud of points inside and outside the body. We obtain basic asymptotics of the estimator, perform simulations, and discuss via Google Earth's data an application to the image analysis of the Aral Sea coast and its cliffs.

1. Introduction. The estimation of functionals defined on the boundary Γ of an unknown body $G \subset \mathbb{R}^d$ is a new branch of nonparametric statistics with powerful applications in several areas, including image analysis and stereology [9]. Cuevas et al. [8] address the estimation of the Minkowski content

$$\lim_{\varepsilon \to 0^+} \frac{\mu(\bigcup_{x \in \Gamma} B_\varepsilon(x))}{2\varepsilon}, \tag{1.1}$$

*Supported in part by MEC grant ECO-2008-05080 and CAM grant 2008-00059-002

†Supported in part by NSF grant DMS-0805570

AMS 2000 subject classifications: Primary 62G05; secondary 60D05

Keywords and phrases: Surface estimation, boundary measure, Delaunay triangulation, stabilization methods

$\mu := \mu_d$ being the Lebesgue measure on $\mathbb{R}^d$ and $B_\varepsilon(x)$ the closed d-dimensional ball with center at x and radius ε. When the limit (1.1) exists, it is the most basic measure of the content of Γ and it coincides with length and surface area in dimensions 2 and 3, respectively [16]. Minkowski content estimators are based on random point samples distributed on a d-dimensional rectangular solid containing G, for which one may determine whether a point is in G or not. Roughly speaking, they are empirical measures of the ε-approximation

$$\frac{\mu(\bigcup_{x\in\Gamma} B_\varepsilon(x))}{2\varepsilon} \tag{1.2}$$

of the Minkowski content of Γ. Both the statistic considered by Cuevas et al. [8], and other closely related ones [2, 10, 18], depend on the *smoothing parameter* ε, which must be chosen as a function of the size of the random point sample.

We propose a different nonparametric approach, free of smoothing parameters, to estimate not only boundary lengths of planar sets and surface areas of solids, but to also estimate surface integrals of those scalar functions whose values are knowable at the sample points. While this paper focusses on instances where the body G is unknown, the proposed method is also of interest for bodies having a known but complex boundary, one having a surface integral defying traditional estimators.

The nonparametric estimation of surface integrals has practical applications in image analysis, when the body G consists of some nonhomogeneous material and the density $h(x)$ of the material at x is collected at the sample points. Thus, one might for example be interested in the mass per unit thickness of Γ, which corresponds to the surface integral commonly represented by $\int_\Gamma h \, d\Gamma$. In some instances, quantities such as temperature and humidity can be measured by sensors randomly distributed on a set, in which an unknown body is embedded, and a fundamental problem is to estimate the surface integral of these quantities on the boundary of the body. In other situations, including those arising in medical imaging, oncology and cardiology, knowledge of boundary length is of importance in the prognosis of an infarction as well as in the assessment of the dissemination capacity of a tumor, as explained in Cuevas et al. [8].

Our statistics are based on the Delaunay triangulation of the sample points. Although the idea can be easily adapted to other graphs, in particular to the Voronoi diagram, we chose the Delaunay triangulation for two specific reasons:

1. It is a well-known tool in curve reconstruction methods. In particular, the β-skeleton [15], the Crust [1], and a wide range of related algorithms involve computing Delaunay triangulations.
2. It is computationally efficient. For example, Boissonnat and Cazals [5] report a 3-dimensional Delaunay triangulation code which can handle 500,000 randomly distributed points per minute.

The basic difference between previous methods and the one introduced here is that formerly used methods estimate a fattened boundary, namely $\bigcup_{x\in\Gamma} B_\varepsilon(x)$, whereas we directly estimate Γ by a surface (a curve if $d = 2$) properly selected among the polyhedra (polygons if $d = 2$) belonging to the Delaunay triangulation. Part of our methodology involves a new algorithm for surface reconstruction based on inner and outer sampling points, which differs substantially from the numerical methods for surface reconstruction in which the sample points are on (or close to) the surface to be reconstructed. Our method is described in Section 2, where we introduce the relevant statistics. Section 3 establishes basic asymptotics of these statistics and provides a strongly consistent estimator of the surface integral. In Section 4 we perform a simulation study. In particular, we estimate the Minkowski content of sets, comparing our method with existing ones. An application to image analysis of the Aral Sea coast and its cliffs from Google Earth's data is discussed in Section 5. Section 6 summarizes our conclusions. Our proofs, given in Section 7, rely on point process methods, including weak convergence of point processes and stabilization methods, a tool for establishing general limit theorems for sums of weakly dependent terms in geometric probability [3, 19–21].

2. The method: Sewing boundaries of unknown sets. Following the basic assumptions of [8, 10], we will assume that G is a compact subset of an open and bounded d-dimensional rectangular solid Q and that the closure of the interior of G has positive μ measure. The boundary of G will be denoted by Γ, that is,

$$\Gamma := \{x : \text{ for any } \varepsilon > 0,\ B_\varepsilon(x) \cap G \neq \emptyset \text{ and } B_\varepsilon(x) \cap G^c \neq \emptyset\},$$

with $G^c := Q \setminus G$. It is assumed that the *μ-boundary of G*, defined by

$$\{x : \text{ for any } \varepsilon > 0,\ \mu(B_\varepsilon(x) \cap G) > 0 \text{ and } \mu(B_\varepsilon(x) \cap G^c) > 0\}, \tag{2.1}$$

coincides with Γ. This rules out the existence of 'extremities' to G having null d-dimensional Lebesgue measure. Thus, if we randomly plot enough points in Q, there will be points inside and outside G close enough to any point on the boundary Γ. We assume that Γ is a $(d-1)$-rectifiable set. That is, there exists a countable collection of continuously differentiable maps $g_i : \mathbb{R}^{d-1} \to \mathbb{R}^d$ such that

$$\mathcal{H}^{d-1}(\Gamma \setminus \bigcup_{i=1}^{\infty} g_i(\mathbb{R}^{d-1})) = 0, \tag{2.2}$$

$\mathcal{H}^{d-1}$ being the $(d-1)$-dimensional Hausdorff measure [16]. We will assume that Γ has finite Hausdorff measure and thus Γ has tangent spaces which are defined almost everywhere.

As in [8, 10], the sampling model consists of n i.i.d. random variables $X_1, \ldots, X_n$ uniformly distributed on Q and n i.i.d. Bernoulli random variables $\delta_1, \ldots, \delta_n$ such that

$$\delta_k = \begin{cases} 1 & \text{if} \quad X_k \in G \\ 0 & \text{if} \quad X_k \notin G. \end{cases} \tag{2.3}$$

In other words, although G is unknown, we are able to know whether a sample point X_k is inside G or not. In addition, given measurable $h : Q \to \mathbb{R}$, we assume that we know the values $h(X_1), \ldots, h(X_n)$; in general h is unknown on all its domain but we assume that we are able to collect its value at each of the n sample points. Our goal is to estimate the surface integral of h, formally defined by

$$\int_{\Gamma} h \; d\Gamma = \int_{\Gamma} h(\gamma) \; \mathcal{H}^{d-1}(d\gamma), \tag{2.4}$$

$\mathcal{H}^{d-1}$ being the $(d-1)$-dimensional Hausdorff measure [16]. Since we are assuming G is compact and that Γ is $(d-1)$-rectifiable, its Minkowski content coincides up to a constant factor with its $(d-1)$-dimensional Hausdorff measure [14]. Recall that d-dimensional Hausdorff measure $\mathcal{H}^d$ satisfies $2^d \mathcal{H}^d / v_d = \mu$, where $v_d := \pi^{d/2}[\Gamma(1+d/2)]^{-1}$ is the volume of the unit radius ball in d dimensions. Thus the estimated target quantities in [2, 8, 10, 18] can be written as a surface integral (2.4) with $h(\gamma) \equiv 1$. This brings the issues and problems of [2, 8, 10, 18] within the compass of this paper.

Denote by $\mathcal{X}_n$ the set of sample points $\{X_1, \ldots, X_n\}$ and let $\mathcal{D}(\mathcal{X}_n)$ be the Delaunay triangulation for $\mathcal{X}_n$, namely the full collection of simplices (triangles when $d = 2$) with vertices in $\mathcal{X}_n$ satisfying the empty sphere criterion, namely no point in $\mathcal{X}_n$ is inside the circum-hypersphere (circumcircle if $d = 2$) of any simplex in $\mathcal{D}(\mathcal{X}_n)$. For any sample of absolutely continuous i.i.d. points on $\mathbb{R}^d$ there exists a unique Delaunay triangulation almost surely [17]. Each simplex $s \in \mathcal{D}(\mathcal{X}_n)$ is represented by a subset of $(d+1)$ vertices

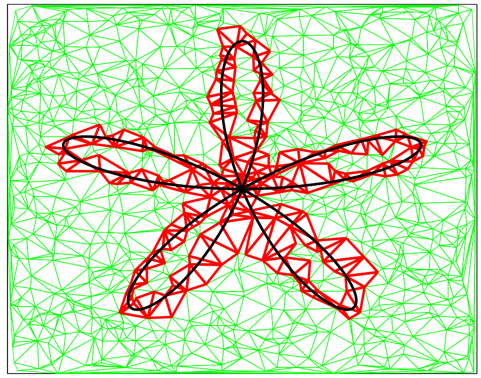

FIG 1. *Polar rose with polar coordinate equation* $\rho = \frac{4}{5}\sin(5\theta)$ *(black); Delaunay tessellation for* 10^3 *i.i.d. points, uniformly distributed on the square* $[-1,1]^2$ *(green and red). Red triangles are the sewing of the rose.*

belonging to $\mathcal{X}_n$, here denoted by $\{X_{s(1)}, \ldots, X_{s(d+1)}\}$. Recalling (2.3), we introduce the *sewing* of Γ, denoted by $S(\mathcal{X}_n, \Gamma)$, and defined by

$$\mathcal{S}(\mathcal{X}_n, \Gamma) := \left\{ s \in \mathcal{D}(\mathcal{X}_n) : 1 \leq \sum_{k=1}^{d+1} \delta_{s(k)} \leq d \right\}.$$

Thus $S(\mathcal{X}_n, \Gamma)$ is the collection of simplices in $\mathcal{D}(\mathcal{X}_n)$ with at least one vertex in G and at least one vertex in G^c. We may assume without loss of generality that there are no sample points on Γ. In this case, the sewing of Γ consists of the simplices in the triangulation which intersect the boundary of G. As an illustration, Figure 1 shows the sewing of a polar rose with five petals (with polar coordinate equation $\rho = \frac{4}{5}\sin(5\theta)$) based on a sample of 10^3 points.

We are particularly interested in the following two surfaces (curves if $d = 2$) contained in $\mathcal{S}(\mathcal{X}_n, \Gamma)$:

The inner sewing, here denoted by $\mathcal{S}^-(\mathcal{X}_n, \Gamma)$, and which consists of the union of all faces of the simplices of $\mathcal{S}(\mathcal{X}_n, \Gamma)$ having vertices in G.

The outer sewing, here denoted by $\mathcal{S}^+(\mathcal{X}_n)$, and which consists of the union of all faces of the simplices of $\mathcal{S}(\mathcal{X}_n)$ having vertices in G^c.

To be more precise: Let s be a simplex of $\mathcal{S}(\mathcal{X}_n, \Gamma)$ and f a face of s, *where here and henceforth by 'face' we mean a simplex of dimension* $(d-1)$. Every such face f may be represented by a vertex set of size d, denoted by

$$\mathcal{V}(f) := \{X_{f(1)}, \ldots, X_{f(d)}\}. \quad (2.5)$$

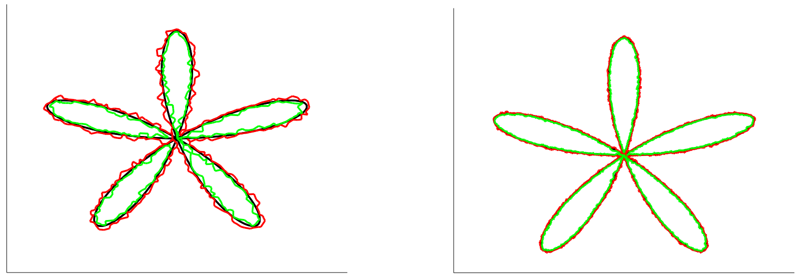

FIG 2. *Inner (green) and outer (red) sewing of the polar rose (black) with equation* $\rho = \frac{4}{5}\sin(5\theta)$ *based on* 10^4 *(left) and* 10^5 *(right) i.i.d. points uniformly distributed on the square* $[-1,1]^2$.

A face f of s is in the inner sewing $\mathcal{S}^-(\mathcal{X}_n, \Gamma)$ if and only if $s \in \mathcal{S}(\mathcal{X}_n, \Gamma)$ and $\mathcal{V}(f) \subset G$. The face itself need not lie wholly in G. On the other hand, a face f of s is in the outer sewing $\mathcal{S}^+(\mathcal{X}_n, \Gamma)$ if and only if $s \in \mathcal{S}(\mathcal{X}_n, \Gamma)$ and $\mathcal{V}(f) \subset G^c$; again the face need not lie wholly in G^c. Both the inner and outer sowing consist of polyhedral surfaces (polygons if $d = 2$) which can be used to estimate Γ. We show in Figure 2 the inner and outer sewing of the polar rose in Figure 1 based, this time, on samples with 10^4 and 10^5 points. As one expects, both sewings fit the rose when n is large.

Now that we have a way to estimate the surface Γ by sampling surfaces, we are ready to explore estimators for the surface integral at (2.4). Any numerical approximation of the integral involving either the inner or the outer sewing is a potential estimator of the integral on Γ. In this work, we approximate integrals by the trapezoidal rule. Thus, if h is continuous and bounded, the surface integral $\int_f h \; df$ of any face $f \in \mathcal{S}(\mathcal{X}_n, \Gamma)$, can be approximated by

$$\int_f h \; df \approx \mathcal{H}^{d-1}(f) \frac{1}{d} \left(\sum_{X_k \in \mathcal{V}(f)} h(X_k) \right),$$

$\mathcal{V}(f)$ being the vertex set of f defined at (2.5). Here $\mathcal{H}^{d-1}(f)$ is the length between two vertices if $d = 2$, the area of a triangle if $d = 3$, and the volume of a $(d-1)$-dimensional simplex otherwise. Therefore, the surface integrals of h with respect to the inner and outer sewings can be approximated by the sums

$$I_n^-(h,\Gamma) := \sum_{f\in\mathcal{S}^-(\mathcal{X}_n,\Gamma)} \mathcal{H}^{d-1}(f)\frac{1}{d}\left(\sum_{X_k\in\mathcal{V}(f)} h(X_k)\right) \tag{2.6}$$

and

$$I_n^+(h,\Gamma) := \sum_{f\in\mathcal{S}^+(\mathcal{X}_n,\Gamma)} \mathcal{H}^{d-1}(f)\frac{1}{d}\left(\sum_{X_k\in\mathcal{V}(f)} h(X_k)\right), \tag{2.7}$$

respectively. Next we study the basic properties of the statistics (2.6) and (2.7) and provide strongly consistent estimators of $\int_\Gamma h\, d\Gamma$.

3. Asymptotics. Let $\mathcal{X}\subset\mathbb{R}^d$ be a locally finite point set, i.e., for any compact set $K\subset\mathbb{R}^d$, $\mathcal{X}\cap K$ contains at most a finite number of points from $\mathcal{X}$. Let $B\subset\mathbb{R}^d$ be a body with boundary ∂B. For $x\in\mathcal{X}$, when $x\in\mathcal{S}^-(\mathcal{X},\partial B)$ we let $\xi^-(x,\mathcal{X},B)$ be the normalized sum of the Hausdorff measures of the faces belonging to $\mathcal{S}^-(\mathcal{X},\partial B)$ and containing x. If $x\notin\mathcal{S}^-(\mathcal{X},\partial B)$ then we put $\xi^-(x,\mathcal{X},B)=0$. Thus

$$\xi^-(x,\mathcal{X},B) := \begin{cases} \dfrac{1}{d}\displaystyle\sum_{\{f\in\mathcal{S}^-(\mathcal{X},\partial B):\ x\in\mathcal{V}(f)\}} \mathcal{H}^{d-1}(f) & \text{if } x\in \displaystyle\bigcup_{f\in\mathcal{S}^-(\mathcal{X},\partial B)} \mathcal{V}(f) \\ 0 & \text{otherwise} \end{cases}$$

Similarly, we define $\xi^+(x,\mathcal{X},B)$ to be the normalized sum of the Hausdorff measures of the faces belonging to $\mathcal{S}^+(\mathcal{X},\partial B)$ and containing x; if no such face exists then ξ^+ is defined to be zero.

We will use the functionals ξ^- and ξ^+ for several purposes; in particular, the statistics (2.6) and (2.7) can be expressed as the weighted sums

$$I_n^-(h,\Gamma) = \sum_{k=1}^n h(X_k)\xi^-(X_k,\mathcal{X}_n,G) \tag{3.1}$$

and

$$I_n^+(h,\Gamma) = \sum_{k=1}^n h(X_k)\xi^+(X_k,\mathcal{X}_n,G). \tag{3.2}$$

The functional ξ^- is translation invariant, that is for all $x\in\mathbb{R}^d$, all locally finite $\mathcal{X}$, and all bodies B, we have $\xi^-(x,\mathcal{X},B)=\xi^-(\mathbf{0},\mathcal{X}-x,B-x)$; here $\mathbf{0}$ is a point at the origin of $\mathbb{R}^d$ and for sets $F\subset\mathbb{R}^d$ and $x\in\mathbb{R}^d$

we put $\{F - x\} := \{y - x : \ y \in F\}$. Also, given $\alpha > 0$ and putting $\alpha F := \{\alpha y : \ y \in F\}$, we have $\mathcal{H}^{d-1}(\alpha f) = \alpha^{d-1}\mathcal{H}^{d-1}(f)$. Thus ξ^- satisfies the following *scaling property* for all $\eta > 0$

$$\eta^{d-1}\xi^-(\mathbf{0}, \mathcal{X}, B) = \xi^-(\mathbf{0}, \eta\mathcal{X}, \eta B), \tag{3.3}$$

which when combined with translation invariance, gives,

$$\eta^{d-1}\xi^-(x, \mathcal{X}, B) = \xi^-(\mathbf{0}, \eta(\mathcal{X} - x), \eta(B - x)). \tag{3.4}$$

Thus, by the definition of $I_n^-(h, \Gamma)$ we have

$$I_n^-(h, \Gamma) = n^{-(d-1)/d} \sum_{k=1}^{n} h(X_k)\xi^-(\mathbf{0}, n^{1/d}(\mathcal{X}_n - X_k), n^{1/d}(G - X_k)). \tag{3.5}$$

Similarly, ξ^+ is translation invariant and satisfies the scaling property (3.4) and therefore

$$I_n^+(h, \Gamma) = n^{-(d-1)/d} \sum_{k=1}^{n} h(X_k)\xi^+(\mathbf{0}, n^{1/d}(\mathcal{X}_n - X_k), n^{1/d}(G - X_k)). \tag{3.6}$$

Central to our results is the asymptotic behavior of the summands in (3.5) and (3.6); see Lemma 7.1. For this it is convenient to introduce the random variable $\xi(t)$ defined at (3.8) below.

Denote by $\mathcal{P}$ the homogeneous Poisson point process of intensity one on $\mathbb{R}^d$ and let $\mathcal{P}^{\mathbf{0}} := \mathcal{P} \cup \{\mathbf{0}\}$. In general, for any $\lambda > 0$, we denote by $\mathcal{P}_\lambda$ the homogeneous Poisson process of intensity λ on Q.

For all $t \in \mathbb{R}$, denote by $\mathbb{H}_t^d$ the *half-space*

$$\mathbb{H}_t^d := \mathbb{R}^{d-1} \times (-\infty, t]. \tag{3.7}$$

The homogeneity of the Poisson point process implies for all $t \in \mathbb{R}$ that $\xi^-(\mathbf{0}, \mathcal{P}^{\mathbf{0}}, \mathbb{H}_t^d)$ and $\xi^+(\mathbf{0}, \mathcal{P}^{\mathbf{0}}, \mathbb{H}_{-t}^d)$ are equally distributed.

Given $t \in \mathbb{R}$, denote by $\xi(t)$ the normalized sum of the Hausdorff measures of the faces incident to $\mathbf{0}$ belonging to the inner sewing $\mathcal{S}^-(\mathcal{P}^{\mathbf{0}}, \mathbb{H}_t^d)$; if there are no such faces, then ξ is set to be zero. That is

$$\xi(t) := \xi^-(\mathbf{0}, \mathcal{P}^{\mathbf{0}}, \mathbb{H}_t^d) \stackrel{\mathcal{D}}{=} \xi^+(\mathbf{0}, \mathcal{P}^{\mathbf{0}}, \mathbb{H}_{-t}^d). \tag{3.8}$$

Lemmas 7.2 and 7.3 imply that $\xi(\cdot)$ is dominated by an integrable function, therefore the constant

$$\alpha_d := \int_0^\infty \mathbb{E}[\xi(t)]\, dt, \tag{3.9}$$

which plays an important role in the asymptotics of our method, is well-defined.

It is also meaningful to consider the Poissonized versions of $I_n^-(h,\Gamma)$ and $I_n^+(h,\Gamma)$, namely

$$\mathcal{I}_\lambda^-(h,\Gamma) := \sum_{x\in\mathcal{P}_\lambda} h(x)\xi^-(x,\mathcal{P}_\lambda,G) \tag{3.10}$$

and

$$\mathcal{I}_\lambda^+(h,\Gamma) := \sum_{x\in\mathcal{P}_\lambda} h(x)\xi^+(x,\mathcal{P}_\lambda,G). \tag{3.11}$$

The following theorems are our main technical results. They take into account the possibility that h can be discontinuous on Γ from either outside or inside G, which is not infrequent in applications, including for example the case when h denotes density of the body G. Say that h is *inner continuous* if the restriction of h to G is continuous; likewise say that h is *outer continuous* if its restriction to G^c is continuous.

THEOREM 3.1. *Let Γ be the boundary of a compact set $G\subset\mathbb{R}^d$. Assume Γ is a $(d-1)$-rectifiable set, that it has finite Hausdorff measure, and that it coincides with the μ-boundary of G, defined at (2.1). If h is inner continuous we have*

$$\lim_{\lambda\to\infty}\mathbb{E}[\mathcal{I}_\lambda^-(h,\Gamma)] = \lim_{n\to\infty}\mathbb{E}[I_n^-(h,\Gamma)] = \alpha_d\int_\Gamma h\ d\Gamma$$

and

$$\lim_{\lambda\to\infty}\mathcal{I}_\lambda^-(h,\Gamma) = \lim_{n\to\infty} I_n^-(h,\Gamma) = \alpha_d\int_\Gamma h\ d\Gamma\quad a.s.,$$

with α_d defined at (3.9). If h is outer continuous then we have the same asymptotics for $\mathcal{I}_\lambda^+$ and I_n^+.

THEOREM 3.2. *Let Γ be as in Theorem 3.1. If h is inner continuous we have*

$$\lim_{\lambda\to\infty}\lambda^{(d-1)/d}\mathrm{Var}[\mathcal{I}_\lambda^-(h,\Gamma)] = V_d\int_\Gamma h^2\ d\Gamma,$$

where

$$V_d := \int_0^\infty \mathbb{E}[\xi^2(t)]\ dt + \int_0^\infty\int_{\mathbb{R}^d} c_t(z)\ dz\ dt, \tag{3.12}$$

with

$$c_t(z) := \mathbb{E}[\xi^-(\mathbf{0},\mathcal{P}^{\mathbf{0}}\cup\{z\},\mathbb{H}_t^d)\xi^-(z,\mathcal{P}^{\mathbf{0}}\cup\{z\},\mathbb{H}_t^d)] - (\mathbb{E}[\xi(t)])^2. \tag{3.13}$$

If h is outer continuous we have the same asymptotics for $\mathrm{Var}[\mathcal{I}_\lambda^+(h,\Gamma)]$.

We expect that modifications of the stabilization arguments in [3, 19–21] will yield a de-Poissonized version of Theorem 3.2, that is variance asymptotics for $I_n^-(h,\Gamma)$. These involve non-trivial arguments and we postpone this to a later paper. Roughly speaking, since Poisson input introduces more variability than binomial input, we expect the variances of $I_n^-(h,\Gamma)$ and $I_n^+(h,\Gamma)$ are no larger than the variances of $\mathcal{I}_n^-(h,\Gamma)$ and $\mathcal{I}_n^+(h,\Gamma)$, respectively. Consequently, we have strong arguments to believe that, under the assumptions of Theorem 3.2, both $\mathrm{Var}[I_n^-(h,\Gamma)]$ and $\mathrm{Var}[I_n^+(h,\Gamma)]$ are $O\left(n^{-(d-1)/d}\right)$. We have not yet obtained analogous results for the rate of convergence for the bias $(\mathbb{E}[I_n^{\pm}(h,\Gamma)] - \alpha_d \int_\Gamma h \; d\Gamma)$. In this direction, we provide in Section 4 some experimental results for dimension $d = 2$.

In accordance with the assumptions of Theorems 3.1 and 3.2, we define

$$S_n(h,\Gamma) := \begin{cases} I_n^-(h,\Gamma), & \text{if } h \text{ is inner continuous on } G, \\ & \text{but not outer continuous on } G; \\ I_n^+(h,\Gamma), & \text{if } h \text{ is outer continuous on } G, \\ & \text{but not inner continuous on } G; \\ \dfrac{1}{2}\left(I_n^-(h,\Gamma) + I_n^+(h,\Gamma)\right), & \text{if } h \text{ is continuous everywhere.} \end{cases}$$

In light of our results and remarks, we propose to estimate the surface integral $\int_\Gamma h \; d\Gamma$ by the strongly consistent sewing based estimator

$$\alpha_d^{-1} S_n(h,\Gamma). \tag{3.14}$$

Since Theorems 3.1 and 3.2 only assume the rectifiability of Γ, the estimator (3.14) is applicable when the body G has sharp interior or and exterior cusps. In particular, it can be used for estimating the boundary measure of such sets, an estimation problem in which previous methods [2, 8, 18] have drawbacks. Under suitable conditions on the smoothing parameter, [10] discusses also the consistency of the estimators based on the Minkowski content for a general class of bodies.

4. Simulations. The estimator (3.14) depends on the constant α_d defined at (3.9). This constant can be estimated by Monte Carlo methods as follows. Consider large n and a surface Γ^0 such that $\int_{\Gamma^0} \; d\Gamma^0$ is known. Let $\mathbb{1}$ denote the function identically equal to 1 and recall the definition of our estimator S_n. Then simulate a random sample of $S_n(\mathbb{1},\Gamma^0)$ and estimate α_d by

$$\hat{\alpha}_d := \frac{\mathrm{mean}(S_n(\mathbb{1},\Gamma^0))}{\int_{\Gamma^0} \; d\Gamma^0},$$

with mean$(S_n(h, \Gamma^0))$ being the sample mean. Given $\hat{\alpha}_d$ and an arbitrary surface Γ, the natural estimator of $\int_\Gamma h \, d\Gamma$ is thus

$$\hat{\alpha}_d^{-1} S_n(h, \Gamma). \tag{4.1}$$

Taking into account this procedure, we estimated α_2 by performing a simulation with $1,000$ independent copies of $S_n(\mathbb{1}, \Gamma^0)$, with Γ^0 being the unit circle, and with sewings based on $n = 10^7$ points uniformly distributed on the square $[-2, 2]^2$. For this configuration, we obtained

$$\hat{\alpha}_2 = 1.1820. \tag{4.2}$$

To compare our estimator with those based on empirical approximation of the Minkowski content, we estimated the length $L = 12\sqrt{3} \approx 20.7846$ of $\mathcal{T}$, $\mathcal{T}$ being the Catalan's trisectrix (also called Tschirnhausen Cubic), with polar equations

$$r = \begin{cases} \sec^3(\theta/3), & \text{if } 0 < \theta \leq \pi, \\ \sec^3((2\pi - \theta)/3), & \text{if } \pi < \theta \leq 2\pi. \end{cases}$$

We used the same simulation framework used in [8], where n points ($n = 10,000$ and $n = 30,000$) were drawn from the square $[-9, 2] \times [-5.5, 5.5]$ 500 times. As we mentioned above, the estimator proposed in [8] depends on the smoothing parameter ε which must be carefully chosen. Of the fourteen possible values of the smoothing parameter considered in [8] and of the fourteen different corresponding estimators, we let $L_n^M(\mathcal{T})$ be the Minkowski content estimator having on average the smallest bias, i.e. the one minimizing the difference between the expectation of the estimator and its target value. In Table 1, we compare $L_n^M(\mathcal{T})$ with the *sewing-based estimator* $L_n^S(\mathcal{T})$, namely

$$L_n^S(\mathcal{T}) = \hat{\alpha}_d^{-1} S_n(\mathbb{1}, \mathcal{T}). \tag{4.3}$$

Roughly speaking, we improve the mean relative error (difference between

n	mean$(L_n^M(\mathcal{T}))$	std$(L_n^M(\mathcal{T})])$	mean$(L_n^S(\mathcal{T}))$	std$(L_n^S(\mathcal{T}))$
$10,000$	19.3679	1.0394	20.7030	0.3140
$30,000$	19.8237	1.0666	20.7328	0.2375

TABLE 1

Mean and standard deviation of estimators based on the Minkowski content, with optimal smoothing parameter, and the sewing-based estimators over 500 replications and simple sizes $n = 10,000$ and $n = 30,000$. The true value is $12\sqrt{3} \approx 20.7846$.

exact and approximate mean value/exact value) by a factor of 10^{-3} and we reduce significantly the standard deviation.

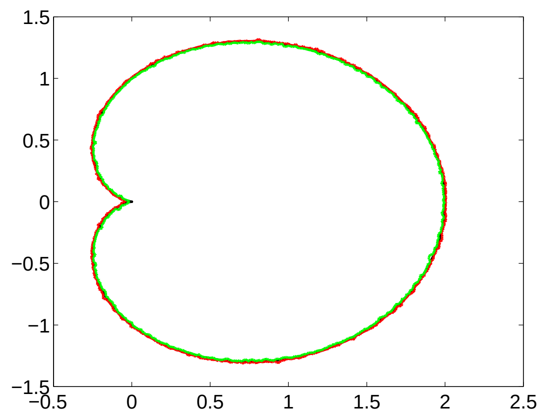

Fig 3. *Inner (green) and outer (outer) sewing based on* 10^5 *uniform points overlapping a cardioid.*

Consider now the planar cardioid $\mathcal{C}$ with polar equation $\rho = 1+\cos\theta,\ 0 \leq \theta \leq 2\pi$. We selected this curve since the sharp cusp at the origin precludes using methods based on empirical approximation of the Minkowski content. Figure 3 shows how the inner and outer sewing practically overlaps the curve with samples of size 10^5 of randomly distributed points on the square $[-0.5, 2.5] \times [-1.5, 1.5]$. The target now is to estimate the length of the cardioid, which equals 8, by the sewing-based estimator $L_n^S(\mathcal{C})$. We performed $1,000$ independent copies of $L_n^S(\mathcal{C})$ for $n = 10^3, 10^4, 10^5$ and 10^6. The results are summarized in Table 1 and Figure 4. Some important remarks can be

n	mean($L_n^S(\mathcal{C})$)	std($L_n^S(\mathcal{C})$)	$\sqrt{n}\times$MSE($L_n^S(\mathcal{C})$)
$1,000$	7.8862	0.1870	1.5157
$10,000$	7.9446	0.1009	1.3254
$100,000$	7.9772	0.0538	1.0801
$1,000,000$	7.9885	0.0323	1.1736

Table 2

Mean, standard deviation and $\sqrt{n}$ *times mean square error, of the sewing-based estimator* $L_n^S(\mathcal{C})$*, computed over* $1,000$ *independent replications.*

drawn from this case study:

1. As one might expect, the sharp inner cusp is slightly underestimated. The gain by increasing the sample size is for both bias and for standard deviation.
2. The simulations strongly suggest that $L_n^S(\mathcal{C})$ is very close to normal.

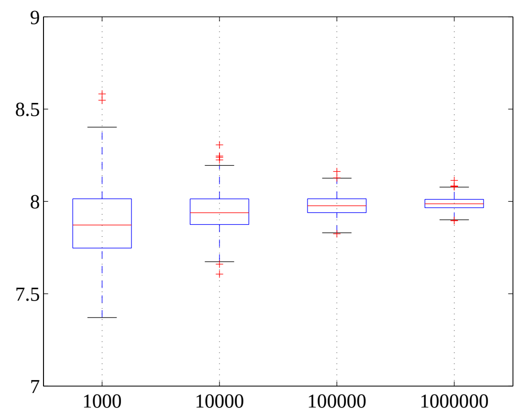

FIG 4. *Boxplots of* $1,000$ *replicas of the sewing-based estimator* $L_n^S(\mathcal{C})$*, for* $n = 10^3, 10^4, 10^5$ *and* 10^6.

This result can be used to provide confidence intervals for the integral that we are estimating. For this, in real applications where there are no replicas of the estimator, bootstrap procedures can be used to estimate $\mathrm{Var}[S_n(h, \Gamma)]$.

3. The mean square error scales with $\sqrt{n}$. Thus the sewing-based estimator seems to converge much faster than estimators based on the empirical approximation of the Minkowski content, which in general, we do not expect to converge faster than $n^{1/4}$ [2].

We finalize our simulation study by estimating line integrals of a parametric family $\{h_\zeta, \zeta \in \mathbb{R}\}$ of scalar functions on the cardioid $\mathcal{C}$ given by $\rho = 1 + \cos\theta, 0 \le \theta \le 2\pi$. We chose $h_\zeta(x, y) = \rho^{\zeta+1/2}$, for two reasons:

(i) We have simple expressions for the integrals

$$I(\zeta) = \int_{\mathcal{C}} h_\zeta \, d\mathcal{C} = \sqrt{2}\int_0^{2\pi} (1 + \cos\theta)^{1+\zeta} \, d\theta$$

where we use $d\mathcal{C} = \sqrt{(\rho(\theta))^2 + (\rho'(\theta))^2} d\theta = \sqrt{2\rho(\theta)} d\theta$.

(ii) The contribution to the total integral of the integral near the sharp inner cusp increases with ζ. The variability of h_ζ also increases with ζ.

As before, we considered $1,000$ independent replications of the sewing-based estimator of $I(\zeta)$ based on $n = 10^3, 10^4, 10^5, 10^6$ points uniformly distributed on the square $[-0.5, 2.5] \times [-1.5, 1.5]$. Similar to the previous case study, increased sample size resulted in smaller bias and smaller standard deviation. Also the simulations strongly suggested that the estimator is very close to normal and that the mean square error scales with $\sqrt{n}$. Our goal now is to highlight how the estimation depends on the parameter zeta. For this, we summarize the results for $n = 10^4$ in Table 3. This behavior on ζ is common

ζ	$I(\zeta)$	$\mathrm{mean}(\hat{I}(\zeta))$	$\mathrm{std}(\hat{I}(\zeta))$
$-1/2$	8.0000	7.9446	0.1009
0	8.8858	8.8786	0.1138
1	13.3286	13.3349	0.1827
2	22.2144	22.2191	0.3470

TABLE 3
Mean and standard deviation, computed over 1,000 *independent replications, of the sewing-based estimator of* $I(\zeta)$ *based on* $n = 10^4$ *points uniformly distributed on the square* $[-0.5, 2.5] \times [-1.5, 1.5]$.

for all the n values that we considered. We remark that, on the one hand, the bias is slightly smaller to the extent that the sharp inner cusp less affects the value of the integral. On the other hand, the more the function h differs from being a constant (i.e., the case $\zeta = -1/2$), the more the variance of $\hat{I}(\zeta)$ grows.

5. Application to image analysis from Google Earth's data. Our simulations show that we require about 10^5 points or more to get attractive estimators of the full image of a planar set. But they also show that 1,000 points are enough to obtain good estimates of integrals along curves. This sample size is quite manageable for online applications for any new-generation personal computer, in which a compiled version of our algorithm would run in less than a second. Even $n = 10^4$ could be implemented for snapshot queries. Next we discuss an application to image analysis from Google Earth's data. Quite possibly Google provides a sharp image of the area around your residence or workplace, but what can you infer from the image of the most inaccessible regions of the planet, including for example the coast of the Aral Sea?

According to the United Nations, the disappearance of the Aral Sea, once the fourth largest inland body of water in our planet, is the worst man-made environmental disaster of the 20th century. It is estimated that the size of the Aral Sea has fallen by more than 60% since the 1960's. Water withdrawals for irrigation have caused a dramatic fall in the water level, revealing a fascinating geology of cliffs overlooking the Aral Sea. What is the mean height of the cliffs along a long waterfront? How irregular are they? Google Earth provides elevation information for any pixel and from

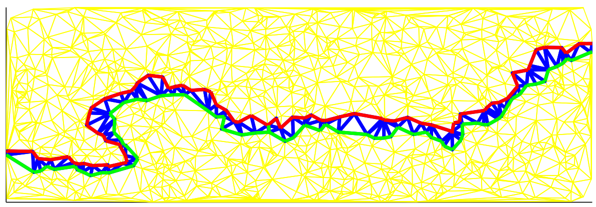

FIG 5. *Google Earth image of the land area bounded between latitudes 45.9 and 46.04 and longitudes 58.9 and 59.27 (top) and Delaunay triangulation for 1,000 points uniformly distributed on the rectangle* $[58.9, 59.27] \times [45.9, 46.04]$ *and sewings of the Aral Sea (bottom).*

its color we can determine whether it is in the Aral Sea or not. This scenario fits our sampling model and thus we may use sewings to estimate the mean and variance of the elevation of cliffs along waterfronts. Both parameters are related with common topographic measures. According to [13], the standard deviation of elevation provides one of the most stable measures of the vertical variability of a topographic surface. On the other hand, the elevation mean is related with the elevation-relief ratio, namely

$$\frac{\text{elevation mean - elevation min}}{\text{elevation max - elevation min}},$$

which has already been computed in the past using a point-sampling technique rather than planimetry [22].

We focus our analysis on the land area bounded between latitudes 45.9 and 46.04 and longitudes 58.9 and 59.27. We chose this surface, approximately 445 km^2, in part because of its complex shape, with capes and bays. Figure 5 (top) shows the Google image of the surface under study. We emphasize three important points:

1. This surface is not a planar rectangle but we proceed as though it were. The land area that concerns us is relatively small and therefore our statistics do not depend on whether we use the Euclidean or geodesic distance. The error made by considering land areas as planar sets is briefly discussed later.

2. One of our basic assumptions is that the set to be estimated is compact and contained in the interior of the rectangle to be scanned. This is not the case now because the boundary between sea and land is not a loop contained by the scanned rectangle. In these cases boundary effects can induce spurious long faces in the extremities of the inner and outer sewings. It is appropriate to exclude these two faces from each sewing. We remark that this procedure does not affect the result for large n.
3. The fractal-like structure of coast lines is not detectable with the limited resolution of Google Earth and thus we may safely assume that the Aral Sea coast, even if naturally fractal-like, is a rectifiable curve.

To summarize and clarify, we follow this procedure: Drop $1,000$ points uniformly distributed on the longitude/latitude rectangle $Q = [58.9, 59.27] \times [45.9, 46.04]$, obtain their corresponding elevations from Google Earth, determine whether they are in the sea or not, and, finally, compute the sewings of the sample. The results are shown in Figure 5 (bottom). In this context, distances between points are found as follows. Putting the earth's radius to be 6371 kilometers, we approximate the distance in kilometers between two latitude/longitude points $(\mathrm{long}_1, \mathrm{lat}_1)$ and $(\mathrm{long}_2, \mathrm{lat}_2)$ of Q by the spherical law of cosines

$$6371 \left[\cos^{-1}\left(\sin(\mathrm{lat}_1)\sin(\mathrm{lat}_2) + \cos(\mathrm{lat}_1)\cos(\mathrm{lat}_2)\cos(\mathrm{long}_2 - \mathrm{long}_1)\right)\right].$$

Alternatively, we obtain the same numerical results using the appropriately scaled Euclidean distance or the Haversine formula.

Call Γ the waterfront, that is to say the curve contained in Q separating land from water in the image, and $L(\Gamma)$ the length of Γ. Let G be the land and $h(x, y)$ the height in meters from the water of the latitude/longitude point $(x, y) \in G$. We are here assuming that the discrete data given by Google can be extended to yield a continuous approximation to height at those points where there is not enough available information. It is reasonable to assume that h is inner continuous on G but not outer continuous on G.

We estimate the mean of the height of the cliffs along Γ, namely

$$\bar{h} := \int_\Gamma h \frac{d\Gamma}{L(\Gamma)},$$

and the standard deviation

$$s_h := \sqrt{\int_\Gamma \left(h - \bar{h}\right)^2 \frac{d\Gamma}{L(\Gamma)}},$$

by

$$\frac{I^-(h,\Gamma)}{I^-(\mathbb{1},\Gamma)} = 10.2787 \text{ m} \quad \text{and} \quad \sqrt{\frac{I^-\left((h-\bar{h})^2,\Gamma\right)}{I^-(\mathbb{1},\Gamma)}} = 9.8102 \text{ m}$$

respectively. Note that $\bar{h} \approx s_h$, which certainly means a significant variance of the height of cliffs along the waterfront.

We remark that one can compute the above estimators without knowing the value of α_d. Using our estimation of this constant, the length of the coastline $L(\Gamma)$ is estimated using (4.2) and (4.3), yielding

$$L^{\mathcal{S}}(\Gamma) = \frac{I^-(\mathbb{1},\Gamma) + I^+(\mathbb{1},\Gamma)}{2\hat{\alpha}_2} = 40.977 \text{ km}.$$

This length may be contrasted with the distance between the lower vertices of Q, which is approximately 28.6313 km.

All the above estimates could be easily implemented in Google Earth. This requires that the user enter the southwest and northeast coordinates and that there is a rule (the pixel color in our case) characterizing the set to estimate. Of course, the larger the land area the greater the differences between spherical and Euclidean measures. Thus the larger the area the worse the estimation. We have to mention that more sophisticated application with which we could provide approximations of the estimation errors, using re-sampling for example, could be still costly for online consultation with current processors.

6. Conclusions.

Random scanning of unknown bodies is a good alternative to regular scanning when the underlying morphology is complex. The paper of Cuevas et al. [8] addressed the problem of estimating the boundary measure of a set for the former type of these scanning setups. For this sampling scheme, we introduce an efficient computational method for estimating not only the boundary measure but also surface integrals of scalar functions, provided one can collect the function values at the sample points.

We discuss conditions for getting strong consistency as well as some issues related with the rate of convergence of our estimators. Our proofs rely on point process methods, including weak convergence of point processes and stabilization methods.

We perform a simulation study to compare our estimators with previous estimators of boundary length of sets, concluding that the sewing-based estimators (4.1) significantly reduce the errors and computation times while increasing precision. We complete our simulation study by estimating boundary length of sets with sharp cusps as well as the integral of scalar functions on the boundary these sets, always obtaining good results.

An online application to image analysis from Google Earth's data is discussed. In particular a complex waterfront of the Aral Sea, approximately 41 km according to the our estimators, is analyzed. Specifically, we estimate surface integrals related with mean and standard deviation of the height of the irregular cliffs facing the sea line.

7. Proofs and technical results. As indicated at the outset, our approach makes use of stabilizing functionals ξ, where here $\xi(x, \mathcal{X})$ is a translation invariant functional defined on pairs $(x, \mathcal{X})$, where $x \in \mathbb{R}^d$ and $\mathcal{X} \subset \mathbb{R}^d$ is locally finite. Translation invariance means that $\xi(x, \mathcal{X}) = \xi(x+y, x+\mathcal{X})$ for all $y \in \mathbb{R}^d$. We recall a few facts about such functionals [3, 19–21]. As in [19] we consider the following metric on the space $\mathcal{L}$ of locally finite subsets of $\mathbb{R}^d$:

$$D(\mathcal{A}, \mathcal{A}') := (\max\{K \in \mathbb{N} : \mathcal{A} \cap B_K(\mathbf{0}) = \mathcal{A}' \cap B_K(\mathbf{0})\})^{-1}. \tag{7.1}$$

We say that $\xi(\cdot, \cdot)$ stabilizes on a homogeneous Poisson point process $\mathcal{P}$ if for all $x \in \mathbb{R}^d$ there is an a.s. finite $R := R(x, \mathcal{P})$ such that

$$\xi(x, \mathcal{P} \cap B_R(x)) = \xi(x, [\mathcal{P} \cap B_R(x)] \cup \mathcal{A})$$

for all $\mathcal{A} \subset (B_R(x))^c$. Now whenever $\xi(\cdot, \cdot)$ satisfies stabilization, then $\mathcal{P}^{\mathbf{0}}$ is a continuity point for the function $g(\mathcal{A}) := \xi(\mathbf{0}, \mathcal{A})$ with respect to the topology on $\mathcal{L}$ induced by D. Thus by the continuous mapping theorem (Prop. A2.3V on page 394 of [11]), if $\mathcal{Y}_n, n \geq 1$, is a sequence of random point measures satisfying $\mathcal{Y}_n \xrightarrow{\mathcal{D}} P^{\mathbf{0}}$ as $n \to \infty$, then $\xi(\mathbf{0}, \mathcal{Y}_n) \xrightarrow{\mathcal{D}} \xi(\mathbf{0}, \mathcal{P}^{\mathbf{0}})$. See [19] for details.

Thus if $U_k, k \geq 1$, are i.i.d. uniform on the unit cube with $\mathcal{U}_n := \{U_k\}_{k=1}^n$, and if ξ is translation invariant, so that $\xi(n^{1/d}U_1, n^{1/d}\mathcal{U}_n) = \xi(\mathbf{0}, n^{1/d}(\mathcal{U}_n - U_1))$, then since the shifted and dilated random point measures $n^{1/d}(\mathcal{U}_n - U_1)$ satisfy the convergence $n^{1/d}(\mathcal{U}_n - U_1) \xrightarrow{\mathcal{D}} P^{\mathbf{0}}$, it follows from stabilization that as $n \to \infty$

$$\xi(n^{1/d}U_k, n^{1/d}\mathcal{U}_n) \xrightarrow{\mathcal{D}} \xi(\mathbf{0}, \mathcal{P}^{\mathbf{0}}). \tag{7.2}$$

This result is central to proving the weak law of large numbers [21]

$$n^{-1} \sum_{k=1}^{n} \xi(n^{1/d}U_k, n^{1/d}\mathcal{U}_n) \xrightarrow{P} \mathbb{E}[\xi(\mathbf{0}, \mathcal{P}^{\mathbf{0}})]. \tag{7.3}$$

The analogous convergence result for the two-dimensional vector

$$[\xi(n^{1/d}U_k, n^{1/d}\mathcal{U}_n), \xi(n^{1/d}U_j, n^{1/d}\mathcal{U}_n)],\ k \neq j$$

is likewise central to showing variance asymptotics and central limit theorems [3, 19] for the sums in (7.3).

One of the main features of our proofs is that if ξ is stabilizing and if $\mathcal{Y}_n, n \geq 1$, is a sequence of point processes on sets increasing up to the *half-space* H which is a translation/rotation of $\mathbb{H}_0^d := \mathbb{R}^{d-1} \times (-\infty, 0]$, and if $\mathcal{Y}_n \xrightarrow{\mathcal{D}} \mathcal{P} \cap H$ as $n \to \infty$, then subject to the proper scaling, there exist limit results for ξ analogous to those at (7.3). Our goal is to make these ideas precise.

We prove our main results for the functional ξ^-. Identical results hold for the functional ξ^+ and the proofs for it are analogous. We henceforth shorthand ξ^- by ξ and I^- by I.

To show the required asymptotics for (3.1) and (3.2) and to take advantage of the ideas just discussed, it is natural to parametrize points in Q as follows. First of all, assume without loss of generality that Q has Lebesgue measure equal to 1. Let $M(G)$ be the medial axis of G, that is to say the closure of the set of all points in G with more than one closest point on Γ. In general, $M(G)$ is a non-regular $(d-1)$-dimensional surface with null d-dimensional Lebesgue measure. We refer the readers to [7] for all matters concerning the theory of medial axis.

Let $\Gamma_0 \subset \Gamma$ be the subset of Γ where there is no uniquely defined tangent plane. Then $\mathcal{H}^{d-1}(\Gamma_0) = 0$ by assumption. Consider the subset G_1 of $G \setminus M(G)$ consisting of points x such that there is a $\gamma \in \Gamma \setminus \Gamma_0$ with $x - \gamma$ orthogonal to the tangent plane at γ and $\gamma := \gamma(x)$ is the boundary point closest to x. Let G_0 be the complement of G_1 with respect to $G \setminus M(G)$.

For all $t > 0$, consider the level sets $\Gamma(t) := \{x \in G \setminus M(G) : \ d(x, \Gamma) = t\}$, where $d(x, \Gamma)$ denotes the distance between x and Γ. Except possibly on a Lebesgue null subset of G_1, we may uniquely parameterize points $x \in G_1$ as $x(\gamma, t)$, where γ is the boundary point closest to x and $t := ||x - \gamma||$. Put $\Gamma(0) = \Gamma$. Let $\Gamma_1(t) := \Gamma(t) \cap G_1$ and $\Gamma_0(t) := \Gamma(t) \cap G_0$. For each $\gamma \in \Gamma \setminus \Gamma_0$, let R_γ be the distance between γ and $M(G)$, measured along the orthogonal to the tangent plane to Γ at γ. Put $D := \sup_{\gamma \in \Gamma} R_\gamma$.

For any integrable function $g : Q \to \mathbb{R}$, an application of the co-area formula (see Theorem 3.2.12 and Lemma 3.2.34 in [14]), for the distance function $f(x) := d(x, \Gamma)$ gives

$$\int_G g(x) dx = \int_0^D \int_{\Gamma(t)} g(x) \mathcal{H}^{d-1}(dx) dt,$$

since $|\nabla(f(x)| \equiv 1$ a.e. This yields the *scaled volume identity*

$$n^{1/d}\int_G g(x)\ dx = \int_0^{Dn^{1/d}}\left[\int_{\Gamma_1(tn^{-1/d})} g(x(\gamma, tn^{-1/d}))\mathcal{H}^{d-1}(dx)\right] dt + \int_0^{Dn^{1/d}}\left[\int_{\Gamma_0(tn^{-1/d})} g(x)\mathcal{H}^{d-1}(dx)\right] dt. \tag{7.4}$$

We may similarly write

$$n^{1/d}\int_{G^c} g(x)\ dx = \int_0^{D'n^{1/d}}\left[\int_{\Gamma_1'(tn^{-1/d})} g(x(\gamma, tn^{-1/d}))\mathcal{H}^{d-1}(dx)\right] dt + \int_0^{D'n^{1/d}}\left[\int_{\Gamma_0'(tn^{-1/d})} g(x)\mathcal{H}^{d-1}(dx)\right] dt, \tag{7.5}$$

where now $D', \Gamma_1'(\cdot)$, and $\Gamma_0'(\cdot)$ are the analogs of $D', \Gamma_1(\cdot)$, and $\Gamma_0(\cdot)$.

To simplify the notation and to set the stage for developing the analog of (7.2), we define for all $x \in G_1$ the shifted and $n^{1/d}$-dilated binomial point measures

$$\mathcal{P}_n(\gamma, t) := n^{1/d}(\mathcal{X}_{n-1} - x(\gamma, tn^{-1/d})) \cup \{\mathbf{0}\}, \tag{7.6}$$

together with the shifted and dilated bodies

$$G_n(\gamma, t) := n^{1/d}(G - x(\gamma, tn^{-1/d})). \tag{7.7}$$

We similarly define the shifted and $\lambda^{1/d}$-dilated Poisson point measures

$$\mathcal{P}_\lambda(\gamma, t) := \lambda^{1/d}(\mathcal{P}_\lambda - x(\gamma, t\lambda^{-1/d})) \cup \{\mathbf{0}\} \tag{7.8}$$

and the shifted and dilated bodies

$$G_\lambda(\gamma, t) := \lambda^{1/d}(G - x(\gamma, t\lambda^{-1/d})). \tag{7.9}$$

More generally, for all $x \in G$ we write

$$\mathcal{P}_n(x) := n^{1/d}(\mathcal{X}_{n-1} - x) \cup \{\mathbf{0}\}$$

and similarly for $G_n(x), \mathcal{P}_\lambda(x)$, and $G_\lambda(x)$.

Roughly speaking, the dilated bodies $G_n(\gamma, t)$ are converging locally around $x(\gamma, tn^{-1/d})$ to a half-space, whereas the restriction of the point processes $\mathcal{P}_n(\gamma, t)$ to $G_n(\gamma, t)$ are converging to a homogeneous point process on this half-space (see (7.12) below).

The next result, a consequence of this observation, is similar to (7.2) and is the key to all that follows. It shows that for all $\gamma \in \Gamma \setminus \Gamma_0$ and $t \in (0, \infty)$, the Hausdorff measure of faces incident to $\mathbf{0}$ and belonging to the inner sewing $S^-(\mathcal{P}_n(\gamma, t), G_n(\gamma, t))$ converges in distribution to the Hausdorff measure of faces incident to $\mathbf{0}$ and belonging to the inner sewing $S^-(\mathcal{P}^{\mathbf{0}}, \mathbb{H}^d_{-t})$.

LEMMA 7.1. *For all $\gamma \in \Gamma \setminus \Gamma_0$ and $t \in (0, \infty)$ we have*

$$\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t)) \xrightarrow{\mathcal{D}} \xi(t);\ \ \xi(\mathbf{0}, \mathcal{P}_\lambda(\gamma, t), G_\lambda(\gamma, t)) \xrightarrow{\mathcal{D}} \xi(t). \tag{7.10}$$

PROOF. We will only prove the first result as the second follows by identical methods. By making a rotation if necessary, we will without loss of generality assume that the vector $\mathbf{0} - (\gamma, 0)$ is orthogonal to the boundary of the half space $\mathbb{H}^d_{-t}$.

For an arbitrary point set $\mathcal{Y} \subset \mathbb{R}^d$ and a body D, let $\xi_B(y, \mathcal{Y} \cap D)$ be the sum of the Hausdorff measure of the faces of the Delaunay triangulation of $\mathcal{Y} \cap D$ lying wholly inside D and incident to $y \in \mathcal{Y}$. Recall that by construction, the faces giving non-zero contribution to $\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t))$ have vertices belonging to $G_n(\gamma, t)$. Since the boundary of $G_n(\gamma, t)$ is differentiable, it follows for large enough n that these faces will eventually belong entirely to $G_n(\gamma, t)$. In other words

$$\lim_{n\to\infty} |\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t)) - \xi_B(\mathbf{0}, \mathcal{P}_n(\gamma, t) \cap G_n(\gamma, t))| = 0 \ \ a.s. \tag{7.11}$$

Since $\mathbf{0}$ is distant t from the boundary of $G_n(\gamma, t)$, it follows that the value of ξ_B at $\mathbf{0}$ with respect to $\mathcal{H} \cap (\mathbb{R}^{d-1} \cap (-\infty, t])$ is determined by the points of $\mathcal{H} \cap (\mathbb{R}^{d-1} \cap (-\infty, t])$ in a ball centered at $\mathbf{0}$ of radius $R := \max(t, R_0)$ where R_0 is the stabilization radius at $\mathbf{0}$ for the graph of the standard Delaunay triangulation of $\mathcal{H} \cup \mathbf{0}$. Thus, reasoning exactly as in (7.1)-(7.2), we have that $\mathcal{H} \cap (\mathbb{R}^{d-1} \cap (-\infty, t])$ is a continuity point for the function $g(\mathcal{A}) := \xi_B(\mathbf{0}, \mathcal{A})$ with respect to the topology on $\mathcal{L}$ induced by the metric D at (7.1). Since

$$\mathcal{P}_n(\gamma, t) \cap G_n(\gamma, t) \xrightarrow{\mathcal{D}} \mathcal{H} \cap (\mathbb{R}^{d-1} \cap (-\infty, t]), \tag{7.12}$$

it therefore follows by the continuous mapping theorem that

$$\xi_B(\mathbf{0}, \mathcal{P}_n(\gamma, t) \cap G_n(\gamma, t)) \xrightarrow{\mathcal{D}} \xi_B(\mathbf{0}, \mathcal{H} \cap (\mathbb{R}^{d-1} \cap (-\infty, t])) = \xi(t).$$

Combining this with (7.11), we have $\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t)) \xrightarrow{\mathcal{D}} \xi(t)$, which concludes the proof.

□

Given (7.10), one has convergence of the corresponding expectations in (7.10) provided the random variables satisfy the customary uniform integrability condition. This is the content of the next lemma.

LEMMA 7.2. *For all $\gamma \in \Gamma \setminus \Gamma_0$ and all $t > 0$ we have*

$$\lim_{n\to\infty} \mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t))] = \lim_{\lambda\to\infty} \mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_\lambda(\gamma, t), G_\lambda(\gamma, t))] = \mathbb{E}[\xi(t)].$$

PROOF. It is a straightforward that the empty sphere criterion characterizing Delaunay triangulations (see e.g. Chapter 4 of [24]) implies that, uniformly in t, the volume of a Delaunay simplex in $\mathcal{P}_n(\gamma, t)$ has exponentially decaying tails, this being equivalent to the tail probability that a circumcircle (circumsphere in dimension greater than 2) contains no points from the binomial point process $\mathcal{P}_n(\gamma, t)$. It follows that for $p > 0$ there is a constant $C(p)$ such that

$$\sup_n \sup_{(\gamma,t)} \mathbb{E}[|\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t))|^p] \le C(p), \tag{7.13}$$

showing that

$$\{\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t)),\ n \ge 1\}$$

are uniformly integrable. By Lemma 7.1, we obtain the desired convergence of $\mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t))]$. The convergence of $\mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_\lambda(\gamma, t), G_\lambda(\gamma, t))]$ follows from identical methods. □

The next two lemmas are also a consequence of exponential decay of the volume of the circumspheres not containing points from $n^{1/d}\mathcal{X}_n$. The next result shows that the expectations appearing in Lemma 7.2 are uniformly bounded, a technical result foreshadowing the upcoming use of the dominated convergence theorem.

LEMMA 7.3. *There is an integrable function* $F : [0, \infty) \to [0, \infty)$*, with exponentially decaying tails, such that*

$$\sup_n \sup_{x \in \Gamma(tn^{-1/d})} \mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_n(x), G_n(x))] \le F(t)$$

and

$$\sup_\lambda \sup_{x \in \Gamma(t\lambda^{-1/d})} \mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_\lambda(x), G_\lambda(x))] \le F(t).$$

PROOF. For all $x \in \Gamma(tn^{-1/d})$ let

$$E(x) := \{\mathbf{0} \text{ belongs to a face } f \text{ of a simplex in } \mathcal{D}(\mathcal{P}_n(x)),\ f \cap \partial\mathbb{H}_t^d \neq \emptyset\}.$$

By definition ξ vanishes on $E^c(x)$. It is easy to see that $P[E(x)]$ has exponentially decaying tails in t uniformly in n and γ. The first result follows by considering $\xi(\mathbf{0}, \mathcal{P}_n(x), G_n(x))\mathbf{1}(E(x))$ and applying the Cauchy-Schwarz inequality and the bound (7.13) with $p = 2$. The second result follows similarly. □

Our last lemma provides a high probability bound on the diameter of Delaunay simplices, a result which follows immediately from the exponential decay of the volume of spheres not containing points from $\mathcal{X}_n$ or $\mathcal{P}_\lambda$. In other words, making use of the bounds $P[\mathcal{P}_\lambda \cap B_r(x) = \emptyset] = \exp(-\lambda r^d v_d)$, where v_d is the volume of the unit radius d-dimensional ball, as well as the bounds $P[\mathcal{X}_n \cap B_r(x) = \emptyset] = (1 - r^d v_d)^n$, we obtain:

LEMMA 7.4. *For any constant $A > 0$, there is a constant $C > 0$ such that with probability exceeding $1-n^{-A}$ the diameter of all Delaunay simplices in $\mathcal{D}(\mathcal{X}_n)$ is bounded by $C(\log n/n)^{1/d}$ and thus with probability exceeding $1 - n^{-A}$ we have*

$$\sup_{i\leq n} \xi(X_i, \mathcal{X}_n, G) \leq C(\log n/n)^{(d-1)/d}$$

and

$$\sup_{x\in\mathcal{P}_\lambda} \xi(x, \mathcal{P}_\lambda, G) \leq C(\log \lambda/\lambda)^{(d-1)/d}.$$

We now have all the ingredients needed to prove Theorem 3.1.

PROOF. (**Theorem 3.1**) The proof has two parts: the first shows expectation convergence and the second uses concentration inequalities to deduce the a.s. convergence.

Part I. We use the above lemmas (binomial input) to first show that $\lim_{n\to\infty} \mathbb{E}[I_n(h, \Gamma)] = \alpha_d \int_\Gamma h \, d\Gamma$ for h inner continuous, recalling that we shorthand ξ^- by ξ and I_n^- by I_n. We omit the proof of $\lim_{\lambda\to\infty} \mathbb{E}[\mathcal{I}_\lambda^-(h, \Gamma)] = \alpha_d \int_\Gamma h \, d\Gamma$, as it follows verbatim, using instead the Poisson input versions of the above lemmas.

Note that $\xi(X_1, \mathcal{X}_n, G) = 0$ if $X_1 \notin G$. Using (3.4) and conditioning on X_1, we have

$$\begin{aligned}
&\mathbb{E}[I_n(h, \Gamma)] = n \; \mathbb{E}[h(X_1) \; \xi(X_1, \mathcal{X}_n, G)] \\
&= n^{1/d} \; \mathbb{E}[h(X_1) \; \xi(\mathbf{0}, n^{1/d}(\mathcal{X}_n - X_1), n^{1/d}(G - X_1))] \\
&= n^{1/d} \int_G h(x) \; \mathbb{E}[\xi(\mathbf{0}, n^{1/d}(\mathcal{X}_{n-1} - x) \cup \{\mathbf{0}\}, n^{1/d}(G - x))] \, dx.
\end{aligned}$$

Using the scaled volume identity (7.4) and putting for all $x \in G$,

$$G_n(x) := h(x)\mathbb{E}[\xi(\mathbf{0}, n^{1/d}(\mathcal{X}_{n-1} - x) \cup \{\mathbf{0}\}, n^{1/d}(G - x))]$$

we get that

$$\mathbb{E}[I_n(h,\Gamma)] = \int_0^{Dn^{1/d}} \int_{\Gamma_1(tn^{-1/d})} G_n(x)\,\mathcal{H}^{d-1}(dx)\,dt$$

$$+ \int_0^{Dn^{1/d}} \int_{\Gamma_0(tn^{-1/d})} G_n(x)\,\mathcal{H}^{d-1}(dx)\,dt. \tag{7.14}$$

The proof of expectation convergence will be complete once we show that the two integrals in (7.14) converge to $\alpha_d \int_\Gamma h(\gamma)\mathcal{H}^{d-1}(d\gamma)$ and zero, respectively.

We first consider the first integral in (7.14). For fixed $t > 0$ and for all n there is an a.e. C^1 mappping $f_n : \Gamma \to \Gamma_1(tn^{-1/d})$ of Γ onto the level set $\Gamma_1(tn^{-1/d})$. To prepare for the dominated convergence theorem, we next show for each $t > 0$ that as $n \to \infty$

$$\int_{\Gamma_1(tn^{-1/d})} G_n(x)\mathcal{H}^{d-1}(dx) \to \int_\Gamma h(\gamma)\mathbb{E}[(\xi(t)]\mathcal{H}^{d-1}(d\gamma). \tag{7.15}$$

To see this, write the difference of the integrals in (7.15) as the sum of

$$\int_{\Gamma_1(tn^{-1/d})} G_n(x)\mathcal{H}^{d-1}(dx) - \int_\Gamma G_n(f_n(\gamma))\mathcal{H}^{d-1}(d\gamma) \tag{7.16}$$

and

$$\int_\Gamma G_n(f_n(\gamma))\mathcal{H}^{d-1}(d\gamma) - \int_\Gamma h(\gamma)\mathbb{E}[(\xi(t)]\mathcal{H}^{d-1}(d\gamma). \tag{7.17}$$

By change of variables the difference (7.16) equals

$$\int_\Gamma G_n(f_n(\gamma))[\mathcal{H}^{d-1}(df_n(\gamma)) - \mathcal{H}^{d-1}(d\gamma)].$$

By the a.e. smoothness of Γ, we have $\mathcal{H}^{d-1}(df_n(\gamma)) = (1+\epsilon_n(\gamma))\mathcal{H}^{d-1}(d\gamma)$ a.e., where $\epsilon_n(\gamma)$ goes to zero as $n \to \infty$. By (7.13) we have

$$\sup_n \sup_{\gamma\in\Gamma} |G_n(f_n(\gamma))| \le C(1)||h||_\infty$$

and so the difference (7.16) goes to zero. Next consider the difference (7.17). For all $\gamma \in \Gamma$, let $f_n(\gamma) = x(\gamma, tn^{-1/d})$. As $n \to \infty$ we have $h(x(\gamma, tn^{-1/d})) \to h(x(\gamma, 0)) = h(\gamma)$ by inner continuity of h. Combining this with Lemma 7.2, we get for all $x \in \Gamma_1(tn^{-1/d})$ with $x = x(\gamma, tn^{-1/d})$, that $G_n(x)$ converges to $h(\gamma)\mathbb{E}[\xi(t)]$, and so by the bounded convergence theorem, the difference (7.17) goes to zero as $n \to \infty$. Thus (7.15) holds.

By Lemma 7.3 we have for fixed $t > 0$

$$\int_{\Gamma_1(tn^{-1/d})} G_n(x)\mathcal{H}^{d-1}(dx) \le ||h||_\infty \times \sup_n(\mathcal{H}^{d-1}(\Gamma(tn^{-1/d})) \times F(t). \tag{7.18}$$

By the boundedness of h and $\mathcal{H}^{d-1}(\Gamma)$, the right hand side of (7.18) is integrable in t and thus the dominated convergence theorem implies

$$\int_0^{Dn^{1/d}} \int_{\Gamma(tn^{-1/d})} \Big[h(x(\gamma, tn^{-1/d}))\ \mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_n(\gamma, t), G_n(\gamma, t))]\Big]\ \mathcal{H}^{d-1}(dx)\, dt$$

converges to

$$\int_0^\infty \int_\Gamma h(\gamma)\mathbb{E}[\xi(t)]\, \mathcal{H}^{d-1}(dx)dt = \int_0^\infty \mathbb{E}[\xi(t)]dt \int_\Gamma h(\gamma)\mathcal{H}^{d-1}(d\gamma),$$

as desired.

Finally we show that the second integral in (7.14) goes to zero as $n \to \infty$. For fixed $t > 0$ the inside integrals in (7.14) are also bounded by the right hand side of (7.18). Since $\mathcal{H}^{d-1}(\Gamma_0(tn^{-1/d})) \to 0$ as $n \to \infty$, the dominated convergence theorem gives that the second integral in (7.14) converges to zero. This completes the proof of expectation convergence.

Part II. Next we show a.s. convergence of $I_n(h, \Gamma)$ and $I_\lambda(h, \Gamma)$. We will do this by appealing to a variant of the Azuma-Hoeffding concentration inequality. We only prove the convergence of $I_n(h, \Gamma)$ as the proof of the convergence of $I_\lambda(h, \Gamma)$ is identical. Let C_1 be a positive constant. For all n define the 'thickened boundary'

$$G(n) := \{x \in G :\ d(x, \Gamma) \le C_1(\log n/n)^{1/d}\}.$$

By smoothness of Γ it follows that $v(n) := \text{volume}(G(n)) = O((\log n/n)^{1/d})$. Put

$$\hat{I'}_n(h, \Gamma) := \sum_{k=1}^{nv(n)} h(X_k)\xi(X_k, \mathcal{X}_{nv(n)}, G), \tag{7.19}$$

where $X_k, k \ge 1$, belong to $G(n)$, $\mathcal{X}_{nv(n)} := \{X_1, ..., X_{nv(n)}\}$. In contrast to $I_n(h, \Gamma)$, notice that $\hat{I'}_n(h, \Gamma)$ contains a deterministic number of non-zero terms and is therefore more amenable to analysis. Our goal is to show that $\hat{I'}_n(h, \Gamma)$ well approximates $I_n(h, \Gamma)$ both a.s. and in L^1, and then obtain concentration results for $\hat{I'}_n(h,\Gamma)$. The proof of a.s. convergence proceeds in the following four steps.

Step (a). With high probability, the summands $h(X_k)\xi(X_k, \mathcal{X}_n, G)$ contributing a non-zero contribution to $I_n(h,\Gamma)$ arise when X_k belongs to the thickened boundary $G(n)$. There are roughly $nv(n)$ such summands and thus the convergence of $I_n(h,\Gamma)$ may be obtained by restricting attention to the statistic $\hat{I'}_n(h,\Gamma)$ defined at (7.19). Our first goal is to make this precise.

The number of sample points in $\mathcal{X}_n$ belonging to $G(n)$ is a binomial random variable $B(n, v(n))$. Re-labeling we may without loss of generality assume that $X_1, ..., X_{B(n,v(n))}$ belong to $G(n)$, where we suppress the dependency of X_k on n.

Put

$$\hat{I}_n(h,\Gamma) := \sum_{k=1}^{B(n,v(n))} h(X_k)\xi(X_k, \mathcal{X}_{B(n,v(n))}, G),$$

where $\mathcal{X}_{B(n,v(n))} := \{X_1, ..., X_{B(n,v(n))}\}$.

By Lemma 7.4 and recalling the definition of $G(n)$, if C_1 is large enough then with high probability the simplices defining the inner sewing of G belong to $G(n)$ and thus it follows that for any constant A there is a C_1 large enough so that

$$P[\hat{I}_n(h,\Gamma) \neq I_n(h,\Gamma)] \leq n^{-A}. \tag{7.20}$$

We will return to this bound in the sequel.

Step (b). We next approximate $\hat{I}_n(h,\Gamma)$ by $\hat{I'}_n(h,\Gamma)$. Given $\hat{I}_n(h,\Gamma)$, we replace $B(n, v(n))$ with its mean, which we assume without loss of generality is integral (otherwise use integer part thereof). Observe that

$$|\hat{I'}_n(h,\Gamma) - \hat{I}_n(h,\Gamma)|$$

is bounded by the product of four factors:

- (i) the difference between the cardinalities of the defining index sets, namely $|B(n, v(n)) - nv(n)|$,
- (ii) the maximal number N of summands affected by either deleting or inserting a single point into either $\mathcal{X}_{B(n,v(n))}$ or $\mathcal{X}_{nv(n)}$,
- (iii) $\sup_{k\leq nv(n)} \xi(X_k, \mathcal{X}_{nv(n)}, G) + \sup_{k\leq B(n,v(n))} \xi(X_k, \mathcal{X}_{B(n,v(n))}, G)$, and
- (iv) the sup norm of h on G, namely $||h||_\infty$.

However, with high probability we have these bounds:

$$|B(n, v(n)) - nv(n)| \leq C_2 \log(nv(n))(nv(n))^{1/2},$$

$N \le C_3 \log n$ (by Lemma 7.4), and, by the analog of Lemma 7.4,

$$\sup_{k \le nv(n)} \xi(X_k, \mathcal{X}_{nv(n)}, G) + \sup_{k \le B(n,v(n))} \xi(X_k, \mathcal{X}_{B(n,v(n))}, G) \le C_4 (\log n/n)^{(d-1)/d}.$$

Here and elsewhere $C_1, C_2, ...$ denote generic constants. We consequently obtain the high probability bound

$$|\hat{I'}_n(h,\Gamma) - \hat{I}_n(h,\Gamma)|$$

$$\le C_2[\log(nv(n))(nv(n))^{1/2}][C_3 \log n][C_4(\log n/n)^{(d-1)/d}]||h||_\infty$$

$$\le C_5(\log n)^3 n^{-(d-1)/2d}. \tag{7.21}$$

Step (c). We combine (7.20) and (7.21) and take A large enough in (7.20) to get the high probability bound

$$|\hat{I'}_n(h,\Gamma) - I_n(h,\Gamma)| \le C_6(\log n)^3 n^{-(d-1)/2d}.$$

Since $|\hat{I'}_n(h,\Gamma) - I_n(h,\Gamma)|$ is deterministically bounded by a multiple of n, it follows that

$$\mathbb{E}[|\hat{I'}_n(h,\Gamma) - I_n(h,\Gamma)|] \to 0$$

whence

$$\mathbb{E}[\hat{I'}_n(h,\Gamma)] \to \alpha_d \int_\Gamma h(x)\ d\Gamma.$$

It thus suffices to show

$$|\hat{I'}_n(h,\Gamma) - \mathbb{E}[\hat{I'}_n(h,\Gamma)]| \to 0 \ \ a.s. \tag{7.22}$$

Step (d). We conclude the proof by showing (7.22). We do this by using a variant of the Azuma-Hoeffding inequality, due to Chalker et al. [6]. Write $I(X_1, ..., X_{nv(n)})$ instead of $\hat{I'}_n(h,\Gamma)$. Consider the martingale difference representation

$$I(X_1, ..., X_{nv(n)}) - \mathbb{E}[I(X_1, ..., X_{nv(n)})] = \sum_{i=1}^{nv(n)} d_i,$$

where $d_i := \mathbb{E}[I(X_1, ..., X_{nv(n)})|\mathcal{F}_i] - \mathbb{E}[I(X_1, ..., X_{nv(n)})|\mathcal{F}_{i-1}]$, where $\mathcal{F}_i$ is the sigma-field generated by $X_1, ..., X_i$. Observe

$$d_i := \mathbb{E}[I(X_1, ..., X_{nv(n)})|\mathcal{F}_i] - \mathbb{E}[I(X_1, ..., X'_i, ..., X_{nv(n)})|\mathcal{F}_i],$$

where X_i' signals an independent copy of X_i. By the conditional Jensen inequality and Lemma 7.4 it follows that

$$|d_i| \le \mathbb{E}[|I(X_1,...,X_{nv(n)}) - I(X_1,...,X_i',...,X_{nv(n)})| \mathcal{F}_i] \le C_7(\log n/n)^{(d-1)/d}$$

holds on a high probability set, i.e., for all $A > 0$ there is a C_7 such that $P[|d_i| \ge C_7(\log n/n)^{(d-1)/d}] \le n^{-A}$. If the $(d_i)_i$ were uniformly bounded in sup norm by $o(1)$, then we could use the Azuma-Hoeffding inequality. Since this is not the case, we use the following variant (see Lemma 1 of [6]), valid for all positive scalars $w_i,\ i \ge 1$:

$$\begin{aligned} P\left[\left|\sum_{i=1}^{nv(n)} d_i\right| > t\right] \quad &\le \quad 2\exp\left(\frac{-t^2}{32\sum_{i=1}^{nv(n)} w_i^2}\right) \\ &+ \left(1 + 2t^{-1}\sup_{i\le nv(n)} ||d_i||_\infty\right)\sum_{i=1}^{nv(n)} P[|d_i| > w_i]. \end{aligned} \tag{7.23}$$

Let $w_i = C_7(\log n/n)^{(d-1)/d}$. We have $\sum_{i=1}^{nv(n)} w_i^2 = C_8(\log n)^{2-(1/d)}n^{-(d-1)/d}$, showing that the first term in (7.23) is summable in n. Since $||d_i||_\infty \le C_9$ and $P[|d_i| \ge w_i] \le n^{-A}$, the second term in (7.23) is summable in n. Since t is arbitrary, this gives (7.22) as desired.

□

Proof. (**Theorem 3.2**) We will follow ideas given in [23], which also involves functionals ξ whose expectations decay exponentially fast with the distance to the boundary. For completeness we provide the details, following in part [19] and [4].

Our goal is to show

$$\lim_{\lambda\to\infty} \lambda^{(d-1)/d}\mathrm{Var}[I_\lambda(h,\Gamma)] = V_d \int_\Gamma h^2(\gamma)\mathcal{H}^{d-1}(d\gamma),$$

where $I_\lambda(h,\Gamma)$ and V_d are defined at (3.10) and (3.12), respectively.

Let $\xi_\lambda(x,\mathcal{P}_\lambda,G) := \xi(\lambda^{1/d}x, \lambda^{1/d}\mathcal{P}_\lambda, \lambda^{1/d}G)$. By scaling (3.4)we have

$$\lambda^{(d-1)/d}\mathrm{Var}[I_\lambda(h,\Gamma)] = \lambda^{-(d-1)/d}\mathrm{Var}\left[\sum_{x\in\mathcal{P}_\lambda} h(x)\xi_\lambda(x,\mathcal{P}_\lambda,G)\right].$$

On the other hand, Campbell's theorem (see Chapter 13 of [12]) gives

$$\begin{aligned} \lambda^{-(d-1)/d}\mathrm{Var}\left[\sum_{x\in\mathcal{P}_\lambda} h(x)\xi_\lambda(x,\mathcal{P}_\lambda,G)\right] \quad &= \quad \lambda^{(d+1)/d}\int_G\int_{\mathbb{R}^d}[...]h(x)h(y)dy\ dx \\ &+ \lambda^{1/d}\int_G \mathbb{E}[\xi_\lambda^2(x,\mathcal{P}_\lambda,G)]h^2(x)dx, \end{aligned} \tag{7.24}$$

where

$$[...] := \mathbb{E}[\xi_\lambda(x, \mathcal{P}_\lambda \cup y, G)\xi_\lambda(y, \mathcal{P}_\lambda \cup x, G)] - \mathbb{E}[\xi_\lambda(x, \mathcal{P}_\lambda, G)]\mathbb{E}[\xi_\lambda(y, \mathcal{P}_\lambda, G)].$$

As in [19], in the double integral in (7.24) we put $y = x + \lambda^{-1/d}z$, thus giving

$$\lambda^{(d+1)/d}\int_G\int_{\mathbb{R}^d}[...]h(x)h(y)dy\,dx = \lambda^{1/d}\int_G\int_{\mathbb{R}^d}F_\lambda(z,x)h(x)h(x+\lambda^{-1/d}z)dz\,dx, \tag{7.25}$$

where

$$F_\lambda(z, x) := \mathbb{E}[\xi_\lambda(x, \mathcal{P}_\lambda \cup \{x + \lambda^{-1/d}z\}, G)\xi_\lambda(x + \lambda^{-1/d}z, \mathcal{P}_\lambda \cup x, G)]$$
$$- \mathbb{E}[\xi_\lambda(x, \mathcal{P}_\lambda, G)]\mathbb{E}[\xi_\lambda(x + \lambda^{-1/d}z, \mathcal{P}_\lambda, G)]$$

and where we adopt the convention that $\xi(x, \mathcal{Y}, G)$ is short for $\xi(x, \mathcal{Y} \cup x, G)$ when x is not in $\mathcal{Y}$. By definition of ξ_λ and by translation invariance, we obtain $F_\lambda(z, x)$ is equal to

$$\mathbb{E}[\xi(\mathbf{0}, \lambda^{1/d}(\mathcal{P}_\lambda - x) \cup \{z\}, \lambda^{1/d}(G - x))\xi(z, \lambda^{1/d}(\mathcal{P}_\lambda - x) \cup \{\mathbf{0}\}, \lambda^{1/d}(G - x))]$$
$$- \mathbb{E}[\xi(\mathbf{0}, \lambda^{1/d}(\mathcal{P}_\lambda - x), \lambda^{1/d}(G - x))]\mathbb{E}[\xi(z, \lambda^{1/d}(\mathcal{P}_\lambda - x), \lambda^{1/d}(G - x))].$$

Write $x := x(\gamma, t\lambda^{-1/d})$ and recall the definitions of $\mathcal{P}_\lambda(\gamma, t)$ and $G_\lambda(\gamma, t)$ from (7.8) and (7.9), respectively, so that the above becomes

$$F_\lambda(z, x) = \mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_\lambda(\gamma, t) \cup \{z\}, \lambda^{1/d}(G - x))\xi(z, \mathcal{P}_\lambda(\gamma, t), \lambda^{1/d}(G - x))]$$
$$- \mathbb{E}[\xi(\mathbf{0}, \mathcal{P}_\lambda(\gamma, t), \lambda^{1/d}(G - x))]\mathbb{E}[\xi(z, \mathcal{P}_\lambda(\gamma, t), \lambda^{1/d}(G - x))].$$

Recalling that $\xi_B(y, \mathcal{Y} \cap D)$ is the normalized Hausdorff measure of the faces of the Delaunay triangulation of $\mathcal{Y} \cap D$ lying inside D and incident to y, the above becomes

$$F_\lambda(z, x) = \mathbb{E}[\xi_B(\mathbf{0}, [\mathcal{P}_\lambda(\gamma, t) \cup \{z\}] \cap G_\lambda(\gamma, t))\xi_B(z, [\mathcal{P}_\lambda(\gamma, t) \cup \{z\}] \cap G_\lambda(\gamma, t))]$$
$$- \mathbb{E}[\xi_B(\mathbf{0}, \mathcal{P}_\lambda(\gamma, t) \cap G_\lambda(\gamma, t))]\mathbb{E}[\xi_B(z, [\mathcal{P}_\lambda(\gamma, t) \cup \{z\}] \cap G_\lambda(\gamma, t))].$$

Next, we have a two dimensional version of (7.10), namely

$$[\xi_B(\mathbf{0}, [\mathcal{P}_\lambda(\gamma, t) \cup \{z\}] \cap G_\lambda(\gamma, t)), \xi_B(z, [\mathcal{P}_\lambda(\gamma, t) \cup \{z\}] \cap G_\lambda(\gamma, t))]$$
$$\xrightarrow{\mathcal{D}} [\xi_B(\mathbf{0}, [\mathcal{P}^{\mathbf{0}} \cup \{z\}] \cap \mathbb{H}_t^d), \xi_B(z, [\mathcal{P}^{\mathbf{0}} \cup \{z\}] \cap \mathbb{H}_t^d)],$$

from which it follows from uniform integrability that as $\lambda \to \infty$

$$F_\lambda(z, x) \to c_t(z), \tag{7.26}$$

where $c_t(z)$ is as in (3.13).

We now find the large λ behavior of the integrals at (7.25). Recalling the scaled volume identity (7.4) and recalling that $x = x(\gamma, t\lambda^{-1/d})$, we get after substitution that

$$\lambda^{(d+1)/d} \int_G \int_{\mathbb{R}^d} [...] h(x)h(y) dy\ dx =$$

$$= \int_0^{D\lambda^{1/d}} \int_{\Gamma_1(t\lambda^{-1/d})} \int_{\mathbb{R}^d} J_\lambda(z,t,\gamma)\ dz\ \mathcal{H}^{d-1}(dx)\, dt + \tag{7.27}$$

$$+ \int_0^{D\lambda^{1/d}} \int_{\Gamma_0(t\lambda^{-1/d})} \int_{\mathbb{R}^d} J_\lambda(z,t,\gamma)\ dz\, \mathcal{H}^{d-1}(dx)\, dt,$$

where

$$J_\lambda(z,t,\gamma) := F_\lambda(z, x(\gamma, t\lambda^{-1/d}))h(x(\gamma, t\lambda^{-1/d}))h(x(\gamma, t\lambda^{-1/d}) + \lambda^{-1/d} z).$$

Notice that $J_\lambda(z,t,\gamma)$ is dominated uniformly in λ by a function $F(z,t,\gamma)$ decaying exponentially fast in $|z|$ and t. By the a.e. continuity of h and the convergence (7.26), the integrand in the first integral tends to $c_t(z)h^2(\gamma)$ as $\lambda \to \infty$. Bounding the integrand by $||h||_\infty^2 F(z,t,\gamma)$ and applying dominated convergence we obtain as $\lambda \to \infty$ that the first integral in (7.27) converges to

$$\int_0^\infty \int_\Gamma \int_{\mathbb{R}^d} h^2(\gamma) c_t(z) dz\ \mathcal{H}^{d-1}(d\gamma) dt. \tag{7.28}$$

The second integral in (7.27) converges to zero, by same methods used to show that the second integral in (7.14) goes to zero. Indeed,

$$\int_{\Gamma_0(t\lambda^{-1/d})} \int_{\mathbb{R}^d} J_\lambda(z,t,\gamma)\ dz\, \mathcal{H}^{d-1}(dx)$$

is bounded by an integrable function of t which is going to zero as $\lambda \to \infty$, since $\int_{\mathbb{R}^d} J_\lambda(z,t,\gamma)\, dz$ are bounded uniformly in γ and t and $\mathcal{H}^{d-1}(\Gamma_0(t\lambda^{-1/d})) \to 0$.

On the other hand, the single integral at (7.24) satisfies the identity

$$\lambda^{1/d} \int_G \mathbb{E}[\xi_\lambda^2(x, \mathcal{P}_\lambda, G)] h^2(x) dx = \lambda^{1/d} \int_G \mathbb{E}[\xi_\lambda^2(x, \mathcal{P}_\lambda, G)] h^2(x)\, dx,$$

which, as $\lambda \to \infty$, tends to

$$\int_0^\infty \int_\Gamma \mathbb{E}[\xi^2(t)] dt\ h^2(\gamma) \mathcal{H}^{d-1}(d\gamma)\, dt. \tag{7.29}$$

Combining (7.28) and (7.29) and recalling the definition of V_d at (3.12), we obtain

$$\lim_{\lambda\to\infty} \lambda^{(d-1)/d}\mathrm{Var}[I_\lambda(h,\Gamma)] = V_d \int_\Gamma h^2(\gamma)\mathcal{H}^{d-1}(d\gamma).$$

This completes the proof of Theorem 3.2. □

Acknowledgements. Part of this work was done while R. Jiménez was visiting the Department of Mathematics of Lehigh University. He thanks the great hospitality of all faculty and staff of this department. The authors gratefully thank Jan Rataj, Joe Fu, as well as anonymous referees for comments leading to an improved exposition.

References.

[1] Amenta, N., Bern, J. and Eppstein, D. (1998). The Crust and the β-skeleton: Combinatorial curve reconstruction. *Graphical Models and Image Processing*, **60**,125-135.

[2] Armendáriz, I., Cuevas, A. and Fraiman, R. (2009). Nonparametric estimation of baundary measures and related functional: asymptotics results. *Adv. Appl. Prob.*, **41**, 311–322.

[3] Baryshnikov, Y. and Yukich, J. E. (2005). Gaussian limits for random measures in geometric probability. *Ann. Appl. Probab.*, **15**, 213–253.

[4] Baryshnikov, Y., Penrose, M. and Yukich, J. E. (2009). Gaussian limits for generalized spacings, *Ann. Appl. Probab.*, 2009, 19, No. 1, 158 - 185.

[5] Boissonnat, J-D. and Cazals, F. (2000). Natural coordinates of points on a surface. *Proceedings of the 16th Annual ACMSymposium on Computational Geometry*, 223–232.

[6] Chalker, T. K., Godbole, A. P., Hitczenko, P., Radcliff, J. and Ruehr O. G. (1999). On the size of a random sphere of influence graph. *Adv. Appl. Prob.*, **31**, 596–609.

[7] Choi, H.I., Choi, S.W. and Moon, H.P. (1997). Mathematical theory of medial axis transform. *Pacific Journal of Mathematics*, **181**, 57-88.

[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] Cuevas, A. and Fraiman, R. (2010). Set estimation, In *New Perspectives in Stochastic Geometry*, 374-397, Eds. Molchanov, I. and Kendall, W. Oxford University Press.

[10] Cuevas, A., Fraiman, R. and L. Györfi (2010). Towards a universally consistent estimator of the Minkowski content, preprint.

[11] Daley, D. J. and D. Vere-Jones (2003), An Introduction to the Theory of Point Processes, vol. I, second ed., Springer-Verlag.

[12] Daley, D. J. and D. Vere-Jones (2008), An Introduction to the Theory of Point Processes, vol. II, second ed., Springer-Verlag.

[13] Evans, I. S. (1972). General geomorphometry, derivatives of altitude, and descriptive statistics. In *Spatial Analysis in Geomorphology*, Eds. Chorley R. J.. Mathuen & Co. Ltd., London.

[14] Federer, H. (1969) *Geometric measure theory.* Springer-Verlag, New York.

[15] Kirkpatrick, D. G. and Radke, J. D. (1998). A framework for computational morphology. In *Computational Geometry*, Ed. Toussaint, G., North-Holland, 217–248.

[16] Mattila, P.(1999). *Geometry of sets and measures in Euclidean spaces: Fractals and rectifiability.* Cambridge University Press.

[17] Möller, J. (1994). *Lectures on Random Voronoi Tessellations. Lectures Notes in Statistics*, **87**. Springer-Verlag.
[18] Pateiro-López, B. and Rodríguez-Casal, A. (2008). Length and surface area estimation under convexity type restrictions. *Adv. Appl. Prob.*, **40**, 348–358.
[19] Penrose, M. D. (2007). Gaussian limits for random geometric measures, *Electron. J. Probab.* **12**, 989-1035.
[20] Penrose, M. D. (2007). Laws of large numbers in stochastic geometry with statistical applications, *Bernoulli*, **13**, 4, 1124-1150.
[21] Penrose, M. D. and Yukich, J. E. (2003). Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.*, **13**, 277–303.
[22] Pike, R. J. and Wilson S. E. (1971). Elevation-Relief Ratio, Hypsometric Integral, and Geomorphic Area-Altitude Analysis. *Geological Society of America Bulletin*, **82**,1079–1084.
[23] Schreiber, T. and Yukich, J. E. (2008). Variance asymptotics and central limit theorems for generalized growth processes with applications to convex hulls and maximal points, *Ann. Probab.*, **36**, 363-396.
[24] Small, C. (1996). *The statistical theory of shape*, Springer Series in Statistics, Springer-Verlag, New York.

Department of Statistics
Universidad Carlos III de Madrid
C./ Madrid, 126 – 28903 Getafe (Madrid), Spain
E-mail: raul.jimenez@uc3m.es

Department of Mathematics
Lehigh University
Bethlehem PA 18015, USA
E-mail: joseph.yukich@lehigh.edu