

TOPOLOGIES OF RANDOM GEOMETRIC COMPLEXES ON RIEMANNIAN MANIFOLDS IN THE THERMODYNAMIC LIMIT

ANTONIO AUFFINGER, ANTONIO LERARIO, ERIK LUNDBERG

ABSTRACT. We investigate the topologies of random geometric complexes built over random points sampled on Riemannian manifolds in the so-called “thermodynamic” regime. We prove the existence of universal limit laws for the topologies; namely, the random normalized counting measure of connected components (counted according to homotopy type) is shown to converge in probability to a deterministic probability measure. Moreover, we show that the support of the deterministic limiting measure equals the set of all homotopy types for Euclidean connected geometric complexes of the same dimension as the manifold.

1. INTRODUCTION

Sarnak and Wigman [19] recently established, utilizing methods developed by Nazarov and Sodin [15], the existence of universal limit laws for the topologies of nodal sets of random band-limited functions on Riemannian manifolds. In the current paper, we adapt these methods to the setting of random geometric complexes, that is, simplicial complexes with vertices arising from a random point process and faces determined by distances between vertices.

Kahle [13] made the first extensive investigation into the topology of random geometric complexes generated by a point process in Euclidean space (zero-dimensional homology of random geometric graphs were also investigated earlier in [17]). The expectation of each Betti number is studied within three main phases or regimes based on the relation between density of points and radius of the neighborhoods determining the complex: the subcritical regime (or “dust phase”) where there are many connected components with little topology, the critical regime (or “thermodynamic regime”) where topology is the richest (and where the percolation threshold appears), and the supercritical regime where the connectivity threshold appears. The thermodynamic regime is seen to have the most intricate topology. Many cycles of various dimensions begin to form as we enter this regime and many cycles become boundaries as we leave this regime.

Random geometric complexes on Riemannian manifolds were studied earlier in the influential work [16] of Niyogi, Smale, and Weinberger, where the manifold is embedded in Euclidean space and the distance between vertices is given by the ambient Euclidean distance.¹ The main question in [16] is motivated by applications in “manifold learning” and concerns the recovery of the topology of a manifold via a random sample of points on the manifold. Consequently, the authors only consider a certain window within the supercritical regime. The subsequent study [5] includes the thermodynamic regime where they provide upper and lower bounds of the same order of growth for each Betti number.

Yogeshwaran, Subag, and Adler [20] established limit laws (including a central limit theorem) in the thermodynamic regime for Betti numbers of random geometric complexes built over Poisson point processes in Euclidean space. More recently, Goel, Trinh, and Tsunoda [11] established a limit law in the thermodynamic regime for Betti numbers of random geometric complexes built over

¹In the current paper, where our manifold is not necessarily embedded, we use geodesic distance to build the complexes. If the manifold happens to be embedded, it can be seen, using Lemma 2.1, that the same limit law stated in Theorem 1.1 holds when using the ambient Euclidean distance.

(possibly inhomogeneous) Poisson point processes in Euclidean space, where they also addressed the case when the point process is supported on a submanifold.

Hiraoka, Shirai, and Trinh [12] proved a limit law for so-called “persistent” Betti numbers. Although this goes in a rather separate direction motivated by topological data analysis, the formalism they use for describing the convergence of the persistence diagram has a loose resemblance to the setup of the current paper in that they introduce a sequence of random measures and show that it converges in an appropriate sense to a deterministic measure.

A survey of other results on random geometric complexes is provided in [4]. Most progress in this area has been made only recently, but the problem of studying the topology of a random geometric complex (or equivalently the ε -neighborhood of a random point cloud) can be traced back to one of Arnold’s problems (see the historical note at the end of the introduction).

A novelty of the current paper is that, whereas previous studies of random geometric complexes have focused on Betti numbers, we consider enumeration of connected components according to homotopy type, a count that provides more refined topological information.

1.1. The Riemannian case. Let (M, g) be a compact Riemannian manifold of dimension $\dim(M) = d$, with normalized volume form $\text{Vol}(M) = 1$. Let $U_n = \{p_1, \dots, p_n\}$ be a random set of points independently sampled from the uniform distribution on M . We denote by $\hat{B}(x, r)$ the Riemannian ball² centered at $x \in M$ of radius $r > 0$. We fix a positive number $\alpha > 0$ and build the random set:

$$(1.1) \quad \mathcal{U}_n = \bigcup_{k=1}^n \hat{B}(p_k, \alpha n^{-1/d}).$$

We denote by $\check{C}(\mathcal{U}_n)$ the corresponding Čech complex (which for $n > 0$ large enough, is homotopy equivalent to \mathcal{U}_n itself, see Lemma 2.1 below).

Let now $\hat{\mathcal{G}}$ be the set of equivalence classes of M -geometric, connected simplicial complexes, up to homotopy equivalence (observe that this is a countable set). In other words, $\hat{\mathcal{G}}$ consists of all the connected simplicial complexes that arise as Čech complexes of some finite family of balls in M . Note that different manifolds give rise to different sets $\hat{\mathcal{G}}$. For example, among all \mathbb{R}^d -geometric complexes we cannot find complexes with nonzero d -th Betti number; but if $M = S^d$, such complexes belong to $\hat{\mathcal{G}}$. When $M = \mathbb{R}^d$ we simply denote this set by \mathcal{G} .

Given \mathcal{U}_n as above, we define the random probability measure $\hat{\mu}_n$ on $\hat{\mathcal{G}}$:

$$\hat{\mu}_n = \frac{1}{b_0(\check{C}(\mathcal{U}_n))} \sum \delta_{[s]},$$

where the sum is over all connected components s of \mathcal{U}_n , $[s]$ denotes the type of s (i.e., the equivalence class of all connected complexes homotopy equivalent to s), and b_0 denotes the number of connected components.

Remark 1. The next theorem deals with the convergence of the random measure $\hat{\mu}_n$ in the limit $n \rightarrow \infty$. We endow the set \mathcal{P} of probability measures on the countable set $\hat{\mathcal{G}}$ with the *total variation distance*:

$$d(\mu_1, \mu_2) = \sup_{A \subset \hat{\mathcal{G}}} |\mu_1(A) - \mu_2(A)|.$$

In this way $\hat{\mu}_n$ is a random variable with values in the metric space (\mathcal{P}, d) . Convergence *in probability* (which is used in Theorem 1.1 and Theorem 1.3) of a sequence of random variables $\{\mu_n\}_{n \in \mathbb{N}}$ to a limit μ means that for every $\varepsilon > 0$ we have $\lim_{n \rightarrow \infty} \mathbb{P}\{d(\mu_n, \mu) > \varepsilon\} = 0$.

²In this paper we adopt the convention that when an object is denoted with a “hat” sign, then it is related to M . Analogous objects related to Euclidean space will have no “hat”. For example a ball in M is denoted by $\hat{B}(x, r)$ and a ball in \mathbb{R}^d by $B(x, r)$.

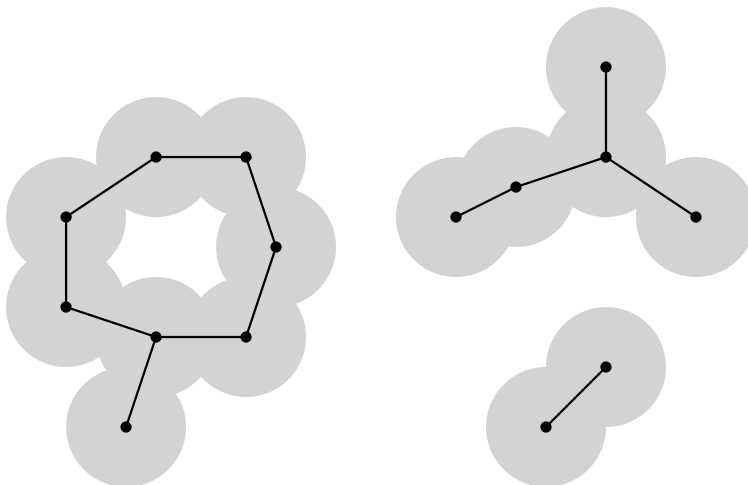


FIGURE 1. A geometric complex in the plane which is homotopy equivalent to $S^1 \cup \{p_1\} \cup \{p_2\}$. The corresponding measure on the set of homotopy classes of connected, simplicial complexes is: $\hat{\mu} = \frac{1}{3} (\delta_{[S^1]} + 2\delta_{[\text{pt}]})$. Theorem 1.1 says that as $n \rightarrow \infty$ the random measure $\hat{\mu}_n$ converges to a deterministic measure.

Theorem 1.1. *The random measure $\hat{\mu}_n$ converges in probability to a universal deterministic probability measure $\mu \in \mathcal{P}$ supported on the set \mathcal{G} of connected \mathbb{R}^d -geometric complexes.*

The “universal” in the previous statement means that μ does not depend on the manifold M (but it depends on its dimension d and on the parameter α).

Remark 2. Since \mathcal{G} is a proper subset of $\hat{\mathcal{G}}$, the measure μ does not charge some points in $\hat{\mathcal{G}}$. This is consistent with the findings of [6] where it was shown that an additional factor of $\log n$ is needed in the radii of the balls defining \mathcal{U}_n in order to see the so-called “connectivity threshold” where nontrivial d -dimensional homology appears.

Remark 3. In the one-dimensional case $d = 1$, the set \mathcal{G} contains only one element: the class of the point (since any connected geometric complex in \mathbb{R} is contractible). The case $d = 2$ is already more interesting, since in this case $\mathcal{G} = \{[w_k]\}_{k \in \mathbb{N}}$ where w_k is the wedge of k -circles ($k = 0$ is the point). In general the support of μ is more difficult to describe.

Remark 4. We can write the limiting measure μ as:

$$\mu = \sum_{\gamma \in \mathcal{G}} a_\gamma \delta_\gamma$$

for some non-negative constants a_γ , $\gamma \in \mathcal{G}$, which depend on the $\alpha > 0$ appearing in (1.1), and satisfy $a_\gamma = c_\gamma/c$ with c_γ, c defined in Proposition 3.1. All of the coefficients a_γ are strictly positive by Proposition 3.2.

The following result is related to the positivity of all coefficients a_γ . While it is not needed for showing such positivity (which follows from Proposition 3.2), it provides additional information on the prevalence of localized components with prescribed homotopy type throughout the manifold, see Section 6.

Proposition 1.2 (Existence of all topologies). *Let $\mathcal{P}_0 \subset \mathbb{R}^d$ be a finite geometric complex and $\alpha > 0$. There exist $R, a > 0$ (depending on \mathcal{P}_0 and α but independent of M and n) such that for every $p \in M$ and for n large enough:*

$$\mathbb{P} \left\{ \mathcal{U}_n \cap \hat{B}(p, Rn^{-1/d}) \simeq \mathcal{P}_0 \right\} > a.$$

Remark 5. Let us point out an interesting consequence of the previous Proposition 1.2: given a compact, embedded manifold $M_0 \hookrightarrow \mathbb{R}^d$, then for $R > 0$ large enough with positive probability the pair (\mathbb{R}^d, M_0) is homotopy equivalent to the pair $(\hat{B}(p, Rn^{-1/d}), \mathcal{U}_n \cap \hat{B}(p, Rn^{-1/d}))$. This follows from the fact that, by [16, Proposition 3.1], one can cover M_0 with (possibly many) small Euclidean balls $M_0 \subset \bigcup_{k=1}^{\ell} B(p_k, \varepsilon) = \mathcal{U}$ with the inclusion $M_0 \hookrightarrow \mathcal{U}$ a homotopy equivalence – hence the pair (\mathbb{R}^d, M_0) is homotopy equivalent to a pair $(\mathbb{R}^d, \mathcal{P}_0)$ with \mathcal{P}_0 a \mathbb{R}^d -geometric complex.

1.2. The local model (the Euclidean case). The proof of Theorem 1.1 for the Riemannian case involves a study of a rescaled version of the problem in a small neighborhood of a given point. Specifically, one can fix $R > 0$ and a point $p \in M$ and study the asymptotic structure of our random complex in the ball $\hat{B}(p, Rn^{-1/d})$. The random geometric complex that we obtain in the $n \rightarrow \infty$ limit can be described as follows.

Let $P = \{p_1, p_2, \dots\}$ be a set of points sampled from the standard spatial Poisson distribution on \mathbb{R}^d and for $\alpha > 0$ consider the random set:

$$\mathcal{P} = \bigcup_{p \in P} B(p, \alpha).$$

We also define \mathcal{P}_R to be the subset of \mathcal{P} consisting of all the connected components of \mathcal{P} that are completely contained in the interior of $B(0, R)$. Note that each $B(p, \alpha)$ is now convex, and, by the Nerve Lemma, \mathcal{P}_R is homotopy equivalent to the simplicial complex $\check{C}(\mathcal{P}_R)$. The relation between $\mathcal{U}_n \cap \hat{B}(p, Rn^{-1/d})$ and \mathcal{P}_R is described in Theorem 4.1.

Similarly to what we have done above, we define the random probability measure μ_R on the set \mathcal{G} of homotopy types of finite and connected \mathbb{R}^d -geometric complexes:

$$\mu_R = \frac{1}{b_0(\check{C}(\mathcal{P}_R))} \sum \delta_{[s]},$$

where the sum is over all connected components s of \mathcal{P}_R . The following result provides a limit law for μ_R .

Theorem 1.3. *The family of random measures μ_R converges in probability to a deterministic probability measure $\mu \in \mathcal{P}$ whose support is all of \mathcal{G} .*

It is important to note that the limiting measure μ appearing in Theorem 1.3 is the same one appearing in Theorem 1.1 (this explains the statement on the support of the limiting measure in Theorem 1.1).

Besides their positivity, little is known about the coefficients a_γ in μ , and a worthwhile computational problem would be to perform Monte Carlo simulations in order to estimate their numerical values and how they depend on α . Concerning dependence on α , a direction that has been suggested to us by Matthew Kahle is to study whether the dependence of μ on α exhibits any interesting behavior related to the “percolation threshold” (recalling that our random geometric complex is associated to continuum percolation with disks for which existence of a percolation threshold $\alpha = \alpha_c$ is known [14]).

While this paper was under review, K. A. Dowling and the third author posted a preprint [10] further adapting these methods to study the limiting homotopy distribution for random cubical complexes associated to Bernoulli site percolation on a cubical grid, where it was shown that the limiting homotopy measure has an exponentially decaying tail for subcritical percolation and a subexponential tail (slower than exponential decay) for supercritical percolation. It is then natural to pose a specific version of the above problem suggested by M. Kahle, namely, to investigate the tail decay of μ in the current setting of random geometric graphs and to determine whether it exhibits a phase transition at the percolation threshold $\alpha = \alpha_c$ (see [10, Concluding Remarks]).

Outline of the paper. We prove Theorem 1.3 addressing the Euclidean setting in Section 3. In Section 4, we establish the “semi-local” result involving a double-scaling limit within a neighborhood on the manifold, and in Section 5 we collect the semi-local information throughout the manifold in order to prove the global result Theorem 1.1 for the manifold setting. We prove Proposition 1.2 in Section 6. Section 2 contains some basic tools used throughout the paper, including the integral geometry sandwiches that play an essential role.

Historical Note. The study of the topology of random simplicial complexes has taken shape only recently with intense activity in the past few years, but it is worth mentioning (as it seems to have been forgotten) that this theme was proposed by V.I. Arnold in the early 1970s, with specific attention given to random geometric complexes in the thermodynamic regime. In the collection [1] of Arnold’s problems, the 28th problem from 1973 states (notice that the set considered is homotopy equivalent to a geometric complex by the nerve lemma):

Consider a random set of points in \mathbb{R}^d with density λ . Let $V(\alpha)$ be the α -neighborhood of this set. Consider the averaged Betti numbers

$$\beta_i(\alpha, \lambda) := \lim_{R \rightarrow \infty} \frac{b_i(V(\alpha) \cap B(0, R))}{R^d}.$$

Investigate these numbers.

Acknowledgements. This work was initiated during the conference “Stochastic Topology and Thermodynamic Limits” that was hosted at ICERM, and part of the work was completed during a second week-long visit to ICERM through the collaborate@ICERM program. The authors wish to thank the institute for their support and for a pleasant and hospitable work environment. The authors would also like to thank the anonymous referee for a careful reading of the paper and many helpful comments regarding the exposition. This research was conducted while A.A. was supported by NSF Grant CAREER DMS-1653552.

2. PRELIMINARY MATERIAL

In this section we collect some basic tools used throughout the paper.

2.1. Geometry. A subset A of a Riemannian manifold (M, g) is called *strongly convex* if for any pair of points $y_1, y_2 \in \text{clos}(A)$ there exists a unique minimizing geodesic joining these two points such that its interior is entirely contained in A (see [7, 9]).

Lemma 2.1. *Let (M, g) be a compact Riemannian manifold. There exists $r_0 > 0$ such that for every point $x \in M$ and every $r < r_0$ the ball $\hat{B}(x, r)$ is strongly convex and contractible. Moreover for every $x_1, \dots, x_k \in M$ and $0 < r_1, \dots, r_k < r_0$ the set $\bigcap_{j=1}^k \hat{B}(x_j, r_j)$ is also strongly convex and contractible. In particular, by the Nerve Lemma, the set $\bigcup_{j=1}^k \hat{B}(x_j, r_j)$ is homotopy equivalent to its associated Čech complex.*

Proof. By [7, Theorem 5.14] there exists a positive and continuous function $r : M \rightarrow (0, \infty)$ such that if $r < r(x)$, then $\hat{B}(x, r)$ is strictly convex (this is in fact due to Whitehead). Since M is compact, then $r_0 = \min r > 0$. Any strongly convex set in a Riemannian manifold is contractible with respect to any of its point (star-shaped in exponential coordinates), hence it follows that for $r < r_0$ the ball $\hat{B}(x, r)$ is also contractible. To finish the proof, we simply observe that the intersection of strongly convex sets A_1, A_2 is still strongly convex: in fact given two points $y_1, y_2 \in A_1 \cap A_2$, by strong convexity of the sets, the unique minimizing geodesic joining the two points is contained in both sets. \square

From now on, the notation $\{\ell\}_k$ denotes the collection of all k -element subsets of $\{1, 2, \dots, \ell\}$.

We will say that a \mathbb{R}^d -geometric complex $\bigcup_{k=1}^{\ell} B(y_k, r)$ is *nondegenerate* if for every $1 \leq k \leq \ell$ and $J = \{j_1, \dots, j_k\} \in \{\ell\}_k$ the intersection $\bigcap_{j \in J} \partial B(y_j, r)$ is transversal (in particular this intersection is empty for $k > d$).

Random geometric complexes are nondegenerate with probability one. However, it could be that without the nondegeneracy assumption one could construct a geometric complex which is not homotopy equivalent to any nondegenerate one. This is not the case, as next Lemma shows.

Lemma 2.2. *The set of homotopy types of \mathbb{R}^d -geometric, connected, nondegenerate complexes coincides with \mathcal{G}^3 .*

Proof. Given a possibly degenerate $\mathcal{P} = \bigcup_{k=1}^{\ell} B(y_k, r)$, let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be the semialgebraic and continuous function defined by

$$f(x) = d(x, \{y_1, \dots, y_k\}) = \min_k \|y_k - x\|,$$

and observe that:

$$\bigcup_{k=1}^{\ell} B(y_k, r) = \{f \leq r\}.$$

We consider now the semialgebraic, monotone family $\{X(r + \epsilon) = \{f \leq r + \epsilon\}\}_{\epsilon \geq 0}$. By [2, Lemma 16.17] for $\epsilon > 0$ the inclusion $X(r) \hookrightarrow X(r + \epsilon)$ is a homotopy equivalence. It suffices therefore to show that for $\epsilon > 0$ small enough $X(r + \epsilon)$ is nondegenerate; this follows from the fact that given points $y_1, \dots, y_{\ell} \in \mathbb{R}^d$, for every $1 \leq k \leq d$ and $J = \{j_1, \dots, j_k\} \in \{\ell\}_k$ there are only finitely many $r > 0$ such that the intersection $\bigcap_{j \in J} \partial B(y_j, r)$ is nontransversal (and the number of possible multi-indices to consider is also finite). \square

The following Proposition plays an important role in all asymptotic stability arguments.

Proposition 2.3. *Let (M, g) be a compact Riemannian manifold of dimension d and $p \in M$. Let $\mathcal{P} \subset \mathbb{R}^d$ be a nondegenerate complex such that:*

$$\mathcal{P} = \bigcup_{j=1}^{\ell} B(y_j, r) \subset B(0, R')$$

for some points $y_1, \dots, y_{\ell} \in \mathbb{R}^d$ and $r, R' > 0$. Given $\alpha > 0$ set $R = \frac{\alpha R'}{r}$ and consider the sequence of maps:

$$\psi_n : \hat{B}(p, Rn^{-1/d}) \xrightarrow{\exp_p^{-1}} B_{T_p M}(0, Rn^{-1/d}) \xrightarrow{\frac{r}{\alpha} n^{1/d}} B_{T_p M}(0, R') \simeq B(0, R').$$

Denoting by φ_n the inverse of ψ_n , there exist $\varepsilon_0 > 0$ and $n_0 > 0$ such that if $\|\tilde{y}_k - y_k\| \leq \varepsilon_0$ for every $k = 1, \dots, \ell$ then for $n \geq n_0$ we have:

$$\bigcup_{k=1}^{\ell} \hat{B}(\varphi_n(\tilde{y}_k), \alpha n^{-1/d}) \simeq \bigcup_{k=1}^{\ell} B(y_k, r).$$

Proof. For $k \leq d$ and for every $J = \{j_1, \dots, j_k\} \in \{\ell\}_k$ either one of these possibilities can verify:

- (1) $\bigcap_{j \in J} B(y_j, r) \neq \emptyset$, in which case, by nondegeneracy, there exists ε_J and y_J such that $\|y_J - y_j\| < r - \varepsilon_J$ for all $j \in J$;
- (2) $\bigcap_{j \in J} B(y_j, r) = \emptyset$, in which case there is no y solving $\|y - y_j\| \leq r$ for all $j \in J$.

³Recall that we did not assume the nondegeneracy condition in the definition of \mathcal{G} .

Since the sequence of maps $d_n : B(0, R') \times B(0, R') \rightarrow \mathbb{R}$ defined by

$$\frac{rn^{1/d}}{\alpha} \cdot d_n(z_1, z_2) = d_M(\varphi_n(z_1), \varphi_n(z_2))$$

converges uniformly to $d_{\mathbb{R}^d}$, then for every $\delta > 0$ there exists $n_1 > 0$ such that for all pairs of points $z_1, z_2 \in B(0, R')$ and for all $n \geq n_1$ we have:

$$(2.1) \quad \left| \frac{rn^{1/d}}{\alpha} \cdot d_M(\varphi_n(z_1), \varphi_n(z_2)) - \|z_1 - z_2\| \right| \leq \delta.$$

For every index set J satisfying condition (1) above, choosing $\delta = \frac{\varepsilon_J r}{3\alpha}$ and setting $\varepsilon_J = \delta$, the previous inequality (2.1) implies that, if $\|\tilde{y}_k - y_k\| < \varepsilon_J$ for every $k = 1, \dots, \ell$, then for $n \geq n_J$:

$$d_M(\varphi_n(\tilde{y}_j), \varphi_n(y_J)) < \alpha n^{-1/d}.$$

This means that the combinatorics of the covers $\{B(y_j, r)\}_{j \in J}$ and $\{\hat{B}(\varphi_n(\tilde{y}_j), \alpha n^{-1/d})\}_{j \in J}$ are the same if $\|y_j - \tilde{y}_j\| < \varepsilon_J$ for $j \in J$ and $n \geq n_J$.

Let us consider now an index set J satisfying condition (2) above. We want to prove that there exists $\varepsilon_J > 0$ and n_J such that if $\|\tilde{y}_j - y_j\| < \varepsilon_J$ for all $j \in J$, then for $n \geq n_J$ the intersection $\bigcap_{j \in J} \hat{B}(\varphi_n(\tilde{y}_j), \alpha n^{-1/d})$ is still empty. We argue by contradiction and assume there exist a sequence of points $x_n \in \hat{B}(p, Rn^{-1/d})$ and for $j \in J$ points $y_{j,n} \in B(0, R')$ with $\|y_{i,n} - y_j\| \leq \frac{1}{n}$ such that for all $j \in J$ and all n large enough:

$$(2.2) \quad d_M(x_n, \varphi_n(y_{j,n})) < \alpha n^{-1/d}.$$

We call $y_n = \psi_n(x_n)$ and assume that (up to subsequences) it converges to some $\bar{y} \in B(0, R')$. Using again the uniform convergence of d_n to $d_{\mathbb{R}^d}$, the inequality (2.2) would give:

$$r > \lim_{n \rightarrow \infty} \frac{n^{1/d} r}{\alpha} \cdot d_M(x_n, \varphi_n(y_{j,n})) = \|\bar{y} - y_j\| \quad \forall j \in J$$

which gives the contradiction $\bar{y} \in \bigcap_{j \in J} B(y_j, r) = \emptyset$.

Set now $n_1 = \max_{J \in \{\ell\}, k \leq d} n_J$ and $\varepsilon_0 = \min_{J \in \{\ell\}, k \leq d} \varepsilon_J$. We have proved that, if $\|\tilde{y}_j - y_j\| < \varepsilon_0$ for all $j = 1, \dots, \ell$, then for all $n \geq n_1$ the two open covers $\{B(y_j, r)\}_{j \in J}$ and $\{\hat{B}(\varphi_n(\tilde{y}_j), \alpha n^{-1/d})\}_{j \in J}$ have the same combinatorics. In particular their Čech complex is the same. Moreover, Lemma 2.1 implies that for a possibly larger $n_0 \geq n_1$ all the balls $\hat{B}(x, \alpha n^{-1/d})$ are strictly convex in M ; consequently, by the Nerve Lemma, for n larger than such n_0 these two open covers are each one homotopy equivalent to their Čech complexes, hence they are themselves homotopy equivalent. \square

2.2. Measure theory. The following lemma will be used in the proof of Theorem 1.3. This lemma and its proof are essentially in [19, Thm. 4.2 (2)], but we provide a proof to make the paper more self-contained and to ensure that it is clear this result is purely measure-theoretic.

Lemma 2.4. *Let $\mu_\lambda = \sum a_{\lambda,k} \delta_k$ be a one-parameter family of random probability measures on \mathbb{N} , and let $\mu = \sum a_k \delta_k$ be a deterministic probability measure on \mathbb{N} . Assume that for every $k \in \mathbb{N}$ $a_{\lambda,k} \rightarrow a_k$ in probability as $\lambda \rightarrow \infty$. Then $\mu_\lambda \rightarrow \mu$ in probability, i.e., for every $\varepsilon > 0$ we have*

$$\lim_{\lambda \rightarrow \infty} \mathbb{P}\{d(\mu_\lambda, \mu) \geq \varepsilon\} = 0,$$

where d denotes the total variation distance.

Proof. Let $\delta > 0$ be arbitrary.

Since μ is a probability measure on \mathbb{N} , there exists K such that

$$(2.3) \quad \sum_{k \geq K} a_k < \frac{\varepsilon}{4}.$$

We have

$$\mathbb{P} \left\{ |a_{\lambda,k} - a_k| > \frac{\varepsilon}{4K} \right\} < \frac{\delta}{2K},$$

which implies (by a union bound)

$$(2.4) \quad \mathbb{P} \left\{ \sum_{k < K} |a_{\lambda,k} - a_k| > \frac{\varepsilon}{4} \right\} < \frac{\delta}{2},$$

and also (by the triangle inequality)

$$(2.5) \quad \mathbb{P} \left\{ \left| \sum_{k < K} a_{\lambda,k} - \sum_{k < K} a_k \right| > \frac{\varepsilon}{4} \right\} < \frac{\delta}{2},$$

for $\lambda \geq \lambda_0$.

The estimate (2.5) implies an estimate for the tails:

$$(2.6) \quad \mathbb{P} \left\{ \left| \sum_{k \geq K} a_{\lambda,k} - \sum_{k \geq K} a_k \right| > \frac{\varepsilon}{4} \right\} < \frac{\delta}{2},$$

since

$$\left| \sum_{k < K} a_{\lambda,k} - \sum_{k < K} a_k \right| = \left| \sum_{k \geq K} a_{\lambda,k} - \sum_{k \geq K} a_k \right|,$$

which follows from μ_λ and μ being probability measures.

For any $\lambda > \lambda_0$, we then have

$$(2.7) \quad \mathbb{P} \left\{ \sum_{k \geq K} a_{\lambda,k} > \frac{\varepsilon}{2} \right\} < \frac{\delta}{2}.$$

Indeed, if

$$\sum_{k \geq K} a_{\lambda,k} > \frac{\varepsilon}{2}$$

then equation (2.3) gives

$$\left| \sum_{k < K} a_{\lambda,k} - \sum_{k < K} a_k \right| > \frac{\varepsilon}{4},$$

and (2.7) then follows from (2.6).

In order to estimate the total variation distance between μ_λ and μ , let $A \subset \mathbb{N}$ be arbitrary. We have:

$$\begin{aligned} \left| \sum_{k \in A} a_{\lambda,k} - \sum_{k \in A} a_k \right| &= \left| \sum_{k \in A, k < K} (a_{\lambda,k} - a_k) + \sum_{k \in A, k \geq K} a_{\lambda,k} - \sum_{k \in A, k \geq K} a_k \right| \\ &\leq \sum_{k \in A, k < K} |a_{\lambda,k} - a_k| + \sum_{k \in A, k \geq K} a_{\lambda,k} + \sum_{k \in A, k \geq K} a_k \\ &\leq \sum_{k < K} |a_{\lambda,k} - a_k| + \sum_{k \geq K} a_{\lambda,k} + \sum_{k \geq K} a_k \\ &\leq \sum_{k < K} |a_{\lambda,k} - a_k| + \sum_{k \geq K} a_{\lambda,k} + \frac{\varepsilon}{4}. \end{aligned}$$

Using a union bound, this implies

$$\mathbb{P} \left\{ \left| \sum_{k \in A} a_{\lambda, k} - \sum_{k \in A} a_k \right| > \varepsilon \right\} \leq \mathbb{P} \left\{ \sum_{k < K} |a_{\lambda, k} - a_k| > \frac{\varepsilon}{4} \right\} + \mathbb{P} \left\{ \sum_{k \geq K} a_{\lambda, k} > \frac{\varepsilon}{2} \right\},$$

which is less than δ by (2.4) and (2.7).

This implies that for every $\delta > 0$ we have, for all λ sufficiently large,

$$\mathbb{P} \left\{ \sup_{A \subset \mathbb{N}} |\mu_\lambda(A) - \mu(A)| \geq \varepsilon \right\} \leq \delta,$$

i.e. we have shown

$$\lim_{\lambda \rightarrow \infty} \mathbb{P} \{d(\mu_\lambda, \mu) \geq \varepsilon\} = 0.$$

□

2.3. The ergodic theorem. The proof of Proposition 3.1 uses the following special case of the d -dimensional ergodic theorem. We follow [14, Ch. 2] and [8, Sec. 12.2]).

Theorem 2.5 (Ergodic Theorem). *Let $(\Omega, \mathcal{F}, \rho)$ be a probability space, and let $T_x, x \in \mathbb{R}^d$ be an \mathbb{R}^d -action on Ω . Let $f \in L^1(\rho)$, and suppose further that the action of T_x on Ω is ergodic. Then we have*

$$\frac{1}{\text{Vol}(B_R)} \int_{B_R} f(T_x(\omega)) dx \rightarrow \mathbb{E} f(\omega) \text{ a.s. and in } L^1$$

as $R \rightarrow \infty$.

Let us explain the terminology appearing in the statement of this theorem. An \mathbb{R}^d -action $T_x, x \in \mathbb{R}^d$ is a group of invertible, commuting, measure-preserving transformations acting measurably on a probability space $(\Omega, \mathcal{F}, \rho)$ and indexed by \mathbb{R}^d . An \mathbb{R}^d -action T_x is said to be *ergodic* if any invariant event has probability either zero or one.

For our application of the ergodic theorem (see the proof of Proposition 3.1 below), the \mathbb{R}^d -action T_x will simply be translation by x acting on the Poisson process (this case is known to be ergodic [14, Prop. 2.6]).

2.4. Component counting function and the integral geometry sandwiches.

Definition 1 (Component counting function). *Let $Y_1, Y_2 \subset X$ and Z be topological spaces (in the case of our interest they will be homotopy equivalent to finite simplicial complexes). We denote by $\mathcal{N}(Y_1, Y_2; [Z])$ the number of connected components of Y_1 entirely contained in the interior of Y_2 and which have the same homotopy type as Z . Similarly, we denote by $\mathcal{N}^*(Y_1, Y_2; [Z])$ the number of connected components of Y_1 which intersect Y_2 and which have the same homotopy type as Z .*

Theorem 2.6 (Integral Geometry Sandwich). *Let \mathcal{P} be a generic geometric complex in \mathbb{R}^d and fix $\gamma \in \mathcal{G}$. Then for $0 < r < R$*

$$\int_{B_{R-r}} \frac{\mathcal{N}(\mathcal{P}, B(x, r); \gamma)}{\text{Vol}(B_r)} dx \leq \mathcal{N}(\mathcal{P}, B_R; \gamma) \leq \int_{B_{R+r}} \frac{\mathcal{N}^*(\mathcal{P}, B(x, r); \gamma)}{\text{Vol}(B_r)} dx.$$

Theorem 2.7 (Integral Geometry Sandwich on a Riemannian manifold). *Let \mathcal{U} be a generic geometric complex on M and fix $\gamma \in \mathcal{G}$. Then for any $\varepsilon > 0$ there exists $\eta > 0$ such that for every $r < \eta$*

$$(1 - \varepsilon) \int_M \frac{\mathcal{N}(\mathcal{U}, \hat{B}(x, r); \gamma)}{\text{Vol}(B_r)} dx \leq \mathcal{N}(\mathcal{U}, M; \gamma) \leq (1 + \varepsilon) \int_M \frac{\mathcal{N}^*(\mathcal{U}, \hat{B}(x, r); \gamma)}{\text{Vol}(B_r)} dx,$$

where B_r still denotes the Euclidean ball of radius r .

Proofs of Theorems 2.6 and 2.7. These results follow from the same proof as in [19]. \square

Remark 6. Similar statements hold true if we take the sum over *all* components, ignoring their type (an observation used throughout the paper). More precisely, denoting by $\mathcal{N}(Y_1, Y_2)$ the number of components of Y_1 entirely contained in the interior of Y_2 and by $\mathcal{N}^*(Y_1, Y_2)$ the number of components of Y_1 that intersect Y_2 , we have the following inequality:

$$\int_{B_{R-r}} \frac{\mathcal{N}(\mathcal{P}, B(x, r))}{\text{Vol}(B_r)} dx \leq \mathcal{N}(\mathcal{P}, B_R) \leq \int_{B_{R+r}} \frac{\mathcal{N}^*(\mathcal{P}, B(x, r))}{\text{Vol}(B_r)} dx$$

and, in the Riemannian framework:

$$(1 - \varepsilon) \int_M \frac{\mathcal{N}(\mathcal{U}, \hat{B}(x, r))}{\text{Vol}(B_r)} dx \leq \mathcal{N}(\mathcal{U}, M) \leq (1 + \varepsilon) \int_M \frac{\mathcal{N}^*(\mathcal{U}, \hat{B}(x, r))}{\text{Vol}(B_r)} dx.$$

Since both \mathcal{P} and \mathcal{U} have only finitely many components, these inequalities follow by simply summing up the two inequalities from the previous theorems over all components type (the sums are over finitely many elements). In fact the integral geometry sandwiches as proved in [19] are adaptations of the original construction from [15], where components were counted without regard to topological type.

3. LIMIT LAW FOR THE EUCLIDEAN CASE

In this section we prove Theorem 1.3. The main step is provided by the following proposition. We continue to use the above notation for the component counting function (see Definition 1).

Proposition 3.1. *For every homotopy type $\gamma \in \mathcal{G}$ there exists a constant c_γ such that the random variable*

$$c_{R,\gamma} = \frac{\mathcal{N}(\mathcal{P}, B(0, R); \gamma)}{\text{Vol}(B(0, R))}$$

converges to a constant c_γ in L^1 as $R \rightarrow \infty$. The same is true for the random variable

$$c_R = \frac{\mathcal{N}(\mathcal{P}, B(0, R))}{\text{Vol}(B(0, R))},$$

(i.e. when we consider all components, with no restriction on their types): as $R \rightarrow \infty$, it converges to a constant c in L^1 .

The next proposition is proved in Section 6.

Proposition 3.2. *The constants c_γ defined in Proposition 3.1 are positive for all $\gamma \in \mathcal{G}$.*

Proof of Proposition 3.1. The proof follows the argument from [19, Theorem 3.3], with some needed modifications.

We will use the shortened notation $\mathcal{N}_R = \mathcal{N}(\mathcal{P}, B(0, R); \gamma)$, $\mathcal{N}(x, r) = \mathcal{N}(\mathcal{P}, B(x, r); \gamma)$ and $\mathcal{N}^*(x, r) = \mathcal{N}^*(\mathcal{P}, B(x, r); \gamma)$ (γ will be fixed for the rest of the proof and we omit dependence on it in the notation). Using Theorem 2.6 we can write, for $0 < \alpha < r < R$:

$$(3.1) \quad \left(1 - \frac{r}{R}\right)^d \frac{1}{\text{Vol}(B_{R-r})} \int_{B_{R-r}} \frac{\mathcal{N}(x, r)}{\text{Vol}(B_r)} dx \leq \frac{\mathcal{N}_R}{\text{Vol}(B_R)} \leq \left(1 + \frac{r}{R}\right)^d \frac{1}{\text{Vol}(B_{R+r})} \int_{B_{R+r}} \frac{\mathcal{N}^*(x, r)}{\text{Vol}(B_r)} dx.$$

Denoting by $A(x, r, \alpha)$ the annulus $\{z \in \mathbb{R}^d : r - \alpha \leq \|x - z\| \leq r + \alpha\}$, we can estimate the integral on the r.h.s. of (3.1) with:

$$\int_{B_{R+r}} \frac{\mathcal{N}^*(x, r)}{\text{Vol}(B_r)} dx \leq \int_{B_{R+r}} \frac{\mathcal{N}(x, r)}{\text{Vol}(B_r)} + \frac{\#P \cap A(x, r, \alpha)}{\text{Vol}(B_r)} dx.$$

In fact, if a component of \mathcal{P} is not entirely contained in the interior of $B(x, r)$, then it touches the boundary of $B(x, r)$ and hence this component must contain a point $p \in P \cap A(x, r, \alpha)$.

We now take $R \rightarrow \infty$ (with r fixed) and use the Ergodic Theorem in order to assert that

$$(3.2) \quad \frac{1}{\text{Vol}(B_{R-r})} \int_{B_{R-r}} \frac{\mathcal{N}(x, r)}{\text{Vol}(B_r)} dx \rightarrow \lambda(r) \quad \text{a.s. and in } L^1,$$

as $R \rightarrow \infty$ where $\lambda(r) := \mathbb{E} \frac{\mathcal{N}(x, r)}{\text{Vol}(B_r)}$ is a constant.

In order to apply the ergodic theorem (Theorem 2.5 stated above) we introduce the function f defined as

$$f(P) = \frac{\mathcal{N}(0, r)}{\text{Vol}(B(0, r))},$$

where r is fixed, and the dependence of f on the Poisson process P is through \mathcal{P} which we recall is the α -neighborhood of P . We also let T_x denote translation by x acting on the Poisson process P . This action is ergodic as noted above in Section 2. We also have $f \in L^1$, since

$$\mathbb{E} \left[\frac{\mathcal{N}(\mathcal{P}, B(0, r), \gamma)}{\text{Vol}(B(0, r))} \right] \leq \frac{\mathbb{E} \#P \cap B(0, r)}{\text{Vol}(B(0, r))} = 1 < \infty,$$

so that the ergodic theorem may be applied.

Furthermore, we notice that

$$f(T_x(P)) = \frac{\mathcal{N}(x, r)}{\text{Vol}(B(0, r))},$$

i.e., shifting P has the same effect as recentering the ball $B(0, r)$ to $B(x, r)$. Thus, the result of applying the ergodic theorem to this choice of f gives precisely the convergence statement in (3.2).

Note that the same convergence statement in (3.2) holds for $\frac{1}{\text{Vol}(B_{R+r})} \int_{B_{R+r}} \frac{\mathcal{N}(x, r)}{\text{Vol}(B_r)} dx$, namely,

$$(3.3) \quad \frac{1}{\text{Vol}(B_{R+r})} \int_{B_{R+r}} \frac{\mathcal{N}(x, r)}{\text{Vol}(B_r)} dx \rightarrow \lambda(r) \quad \text{a.s. and in } L^1,$$

as $R \rightarrow \infty$ where $\lambda(r)$ is the same constant as in (3.2).

We also have

$$(3.4) \quad \frac{1}{\text{Vol}(B_{R+r})} \int_{B_{R+r}} \frac{\#P \cap A(x, r, \alpha)}{\text{Vol}(B_r)} dx \rightarrow a(r) \quad \text{a.s. and in } L^1$$

as $R \rightarrow \infty$ where $a(r) := \mathbb{E} \frac{\#P \cap A(x, r, \alpha)}{\text{Vol}(B_r)} = \frac{\text{Vol}(A(x, r, \alpha))}{\text{Vol}(B_r)}$. This follows from the ergodic theorem as well (although it seems more natural to view it as a consequence of the law of large numbers).

Let $\varepsilon > 0$ be arbitrary. Since $a(r) = O(r^{-1})$ we can choose r sufficiently large that $a(r) < \varepsilon$. We then choose $R \gg r \gg 0$ sufficiently large so that $(1 + r/R)^d < 1 + \varepsilon$, $(1 - r/R)^d > 1 - \varepsilon$. Using the above convergence statements (3.2), (3.3), and (3.4) (while making R even larger if necessary) we have,

$$(3.5) \quad (1 - \varepsilon)(\lambda(r) - \varepsilon) \leq \frac{\mathcal{N}_R}{\text{Vol}(B_R)} \leq (1 + \varepsilon)(\lambda(r) + a(r) + \varepsilon) \quad \text{in expectation.}$$

Using $a(r) < \varepsilon$ and also that $\lambda(r) \leq 1$ for all r , which follows from

$$\mathbb{E} \mathcal{N}(x, r) \leq \mathbb{E} \#P \cap B(0, r) = \text{Vol}(B(0, r)),$$

(3.5) implies

$$\mathbb{E} \left| \frac{\mathcal{N}_R}{\text{Vol}(B_R)} - \lambda(r) \right| < \varepsilon(1 + \lambda(r)) + (1 + \varepsilon)a(r) < 2\varepsilon + (1 + \varepsilon)\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this implies the existence of a constant c_γ such that $\lambda(r) \rightarrow c_\gamma$ in L^1 as $r \rightarrow \infty$ and $\mathcal{N}_R/\text{Vol}(B_R) \rightarrow c_\gamma$ in L^1 as $R \rightarrow \infty$.

The proof of the second statement in the proposition concerning the existence of a limit $c_R \rightarrow c$ for all components (with no restriction on homotopy type) follows from the same argument while replacing the integral geometry sandwich with its more basic version (see Remark 6 in Section 2). \square

3.1. Proof of Theorem 1.3. We write the measure μ_R as:

$$\begin{aligned} \mu_R &= \frac{1}{\mathcal{N}(\mathcal{P}, B(0, R))} \sum_{\gamma \in \mathcal{G}} \mathcal{N}(\mathcal{P}, B(0, R); \gamma) \delta_\gamma \\ &= \frac{\text{Vol}(B(0, R))}{\mathcal{N}(\mathcal{P}, B(0, R))} \sum_{\gamma \in \mathcal{G}} \frac{\mathcal{N}(\mathcal{P}, B(0, R); \gamma)}{\text{Vol}(B(0, R))} \delta_\gamma \\ &= \sum_{\gamma \in \mathcal{G}} \frac{c_{R, \gamma}}{c_R} \delta_\gamma. \end{aligned}$$

By the convergence statements in Proposition 3.1, and since $c > 0$ (which follows from Proposition 3.2 since $c \geq c_\gamma$), we have $a_{R, \gamma} = \frac{c_{R, \gamma}}{c_R}$ converges in L^1 to a constant $a_\gamma = \frac{c_\gamma}{c}$ as $R \rightarrow \infty$.

The positivity of the coefficients a_γ follows from the positivity of c_γ stated in Proposition 3.2.

Next we prove that the measure

$$\mu = \sum_{\gamma \in \mathcal{G}} a_\gamma \delta_\gamma$$

is indeed a probability measure (see Proposition 3.4 below). The main obstacle, addressed in the following lemma (cf. [19, Sec. 4]), is in preventing the mass in the sequence of measures μ_R from escaping to infinity.

Lemma 3.3 (Topology does not leak to infinity). *For every $\delta > 0$ there exists a finite set $g \subset \mathcal{G}$ and $R_0 > 0$ such that for all $R \geq R_0$*

$$\mathbb{E} \sum_{\gamma \in g^c} c_{R, \gamma} < \frac{\delta}{4}.$$

Proof of Lemma 3.3. First we observe that

$$(3.6) \quad \sum_{\gamma \in \mathcal{G}} \frac{\mathbb{E} \mathcal{N}(\mathcal{P}, B(0, r); \gamma)}{\text{Vol}(B(0, r))} < a_0 < \infty,$$

where a_0 is independent of r . Indeed,

$$\sum_{\gamma \in \mathcal{G}} \frac{\mathbb{E} \mathcal{N}(\mathcal{P}, B(0, r); \gamma)}{\text{Vol}(B(0, r))} = \frac{\mathbb{E} \mathcal{N}(\mathcal{P}, B(0, r))}{\text{Vol}(B(0, r))} \leq \frac{\mathbb{E} |\{P \cap B(0, r)\}|}{\text{Vol}(B(0, r))},$$

which is a constant independent of r (the average number of points of a Poisson process in a given region is proportional to the volume of the region).

Let $A \subset \mathcal{G}$ be arbitrary. Then, using the Integral Geometry Sandwich, we obtain:

$$\begin{aligned} \int_{\Omega} \sum_{\gamma \in A} c_{R, \gamma}(\omega) d\omega &\leq \left(1 + \frac{r}{R}\right)^d \frac{1}{\text{Vol}(B_{R+r})} \mathbb{E} \left\{ \sum_{\gamma \in A} \int_{B_{R+r}} \frac{N^*(\mathcal{P}, B(x, r); \gamma)}{\text{Vol}(B_r)} dx \right\} \\ &\leq \left(1 + \frac{r}{R}\right)^d \left(\sum_{\gamma \in A} \mathbb{E} \frac{N(\mathcal{P}, B(x, r); \gamma)}{\text{Vol}(B_r)} + O(r^{-1}) \right). \end{aligned}$$

Let $\delta > 0$ be arbitrary, and choose r sufficiently large that the above $O(r^{-1})$ error term is smaller than $\delta/16$.

By the convergence (3.6) there exists a finite set $g \subset \mathcal{G}$ such that

$$\sum_{\gamma \in g^c} \frac{\mathbb{E}N(\mathcal{P}, B(0, r); \gamma)}{\text{Vol}(B(0, r))} < \frac{\delta}{16}.$$

Choosing R_0 large enough that $\left(1 + \frac{r}{R}\right)^d < 2$ we then have for all $R \geq R_0$

$$\mathbb{E} \sum_{\gamma \in g^c} c_{R, \gamma} < 2 \left(\frac{\delta}{16} + \frac{\delta}{16} \right) = \frac{\delta}{4},$$

as desired, and this completes the proof of the lemma. \square

Proposition 3.4. *The measure*

$$\mu = \sum_{\gamma \in \mathcal{G}} a_\gamma \delta_\gamma,$$

with $a_\gamma = \frac{c_\gamma}{c}$, is a probability measure.

Proof of Proposition 3.4. We need to show that $\sum_{\gamma \in \mathcal{G}} a_\gamma = 1$, or equivalently,

$$(3.7) \quad \sum_{\gamma \in \mathcal{G}} c_\gamma = c.$$

Let $\delta > 0$ and take $g \subset \mathcal{G}$ to be the set guaranteed by Lemma 3.3.

We want to show that

$$(3.8) \quad \left| \sum_{\gamma \in \mathcal{G}} c_\gamma - c \right| \leq \delta,$$

which will then immediately establish (3.7) since $\delta > 0$ is arbitrary.

For fixed ω in the sample space Ω observe that by Fatou's lemma

$$\sum_{\gamma \in g^c} c_\gamma \leq \liminf_{R \rightarrow \infty} \sum_{\gamma \in g^c} c_{R, \gamma}(\omega),$$

and applying Fatou's lemma again followed by Tonelli's theorem, we have

$$\begin{aligned} \int_{\Omega} \sum_{\gamma \in g^c} c_\gamma d\omega &\leq \int_{\Omega} \liminf_{R \rightarrow \infty} \sum_{\gamma \in g^c} c_{R, \gamma}(\omega) d\omega \\ &\leq \liminf_{R \rightarrow \infty} \int_{\Omega} \sum_{\gamma \in g^c} c_{R, \gamma} d\omega \\ &= \liminf_{R \rightarrow \infty} \sum_{\gamma \in g^c} \int_{\Omega} c_{R, \gamma}(\omega) d\omega. \end{aligned}$$

Combining this with Lemma 3.3 we obtain

$$(3.9) \quad \int_{\Omega} \sum_{\gamma \in g^c} c_\gamma d\omega \leq \frac{\delta}{4}.$$

We proceed to estimate $\left| \sum_{\gamma \in \mathcal{G}} c_\gamma - c \right|$:

$$\begin{aligned}
\left| \sum_{\gamma \in \mathcal{G}} c_\gamma - c \right| &= \int_{\Omega} \left| \sum_{\gamma \in \mathcal{G}} c_\gamma - c \right| d\omega \\
&= \int_{\Omega} \left| \sum_{\gamma \in \mathcal{G}} c_\gamma - c_R + c_R - c \right| d\omega \\
&= \int_{\Omega} \left| \sum_{\gamma \in \mathcal{G}} c_\gamma - \sum_{\gamma \in \mathcal{G}} c_{R,\gamma} + c_R - c \right| d\omega \\
&\leq \int_{\Omega} \left| \sum_{\gamma \in \mathcal{G}} c_\gamma - \sum_{\gamma \in \mathcal{G}} c_{R,\gamma} \right| d\omega + \int_{\Omega} |c_R - c| d\omega \\
&\leq \int_{\Omega} \left| \sum_{\gamma \in g} c_\gamma - \sum_{\gamma \in g} c_{R,\gamma} \right| d\omega + \int_{\Omega} \left| \sum_{\gamma \in g^c} c_\gamma - \sum_{\gamma \in g^c} c_{R,\gamma} \right| d\omega + \int_{\Omega} |c_R - c| d\omega \\
&\leq \int_{\Omega} \sum_{\gamma \in g} |c_\gamma - c_{R,\gamma}| d\omega + \int_{\Omega} \sum_{\gamma \in g^c} c_\gamma d\omega + \int_{\Omega} \sum_{\gamma \in g^c} c_{R,\gamma} d\omega + \int_{\Omega} |c_R - c| d\omega \\
&\leq \frac{\delta}{4} + \frac{\delta}{4} + \frac{\delta}{4} + \frac{\delta}{4}.
\end{aligned}$$

In the last line, we have estimated the first and last terms by $\delta/4$ by choosing R sufficiently large and using the L^1 convergence $c_{R,\gamma} \rightarrow c_\gamma$ (and the fact that g is a finite set) and the L^1 convergence $c_R \rightarrow c$; we have estimated the second term by $\delta/4$ using (3.9); and we have estimated the third term by $\delta/4$ by choosing R sufficiently large to apply Lemma 3.3. This establishes (3.8) and concludes the proof of the proposition. \square

Having established by Proposition 3.4 that μ is a probability measure, the convergence in probability $\mu_R \rightarrow \mu$ now follows from the coefficient-wise convergence $a_{R,\gamma} \rightarrow a_\gamma$ along with the purely measure-theoretic result Lemma 2.4.

The statement that the support of μ is all of \mathcal{G} follows from the positivity of the coefficients a_γ . This concludes the proof of Theorem 1.3.

4. SEMI-LOCAL COUNTS IN THE RIEMANNIAN CASE

In this section, we study the components of \mathcal{U}_n contained in a neighborhood $\hat{B}(p, Rn^{-1/d})$ of a point $p \in M$ by relating this case to the Euclidean case. As a preliminary step, we consider the following diffeomorphism (see also Proposition 2.3):

$$(4.1) \quad \psi_n : \hat{B}(p, Rn^{-1/d}) \xrightarrow{\exp_p^{-1}} B_{T_p M}(0, Rn^{-1/d}) \xrightarrow{n^{1/d}} B_{T_p M}(0, R) \simeq B(0, R).$$

Through ψ_n , the stochastic point process $U_n \cap \hat{B}(p, Rn^{-1/d})$ induces a stochastic point process on $B(0, R)$ which converges in distribution to the uniform Poisson process on $B(0, R)$ [18, Sec. 3.5]. By Skorokhod's representation theorem [3, Ch. 1, Sec. 6], there exists a representation, or "coupling", of these point processes defined on a common probability space such that the convergence of these stochastic processes is almost sure.

Theorem 4.1. *Let $p \in M$. For every $\delta > 0$ and for $R > 0$ sufficiently large there exists n_0 such that (using the coupling given by Skorokhod's theorem mentioned above) for every $\gamma \in \mathcal{G}$ and for $n \geq n_0$:*

$$(4.2) \quad \mathbb{P} \left\{ \mathcal{N}(\mathcal{P}, B(0, R); \gamma) = \mathcal{N}(\mathcal{U}_n, \hat{B}(p, Rn^{-1/d}); \gamma) \right\} \geq 1 - \delta.$$

Proof. Let φ_n denote the inverse of the map ψ_n defined in (4.1). For the proof of (4.2) we will need to establish the following three facts:

- (1) There exists $\ell_0 > 0$ and $n_1 > 0$ such that with probability at least $1 - \delta/3$ we have:

$$(4.3) \quad \#(U_n \cap \hat{B}(p, Rn^{-1/d})) = \#(P \cap B(0, R)) \leq \ell_0$$

(i.e. with positive probability for large n , depending on δ , both point processes have the same number of points and this number is bounded by some constant ℓ_0 , which also depends on δ).

- (2) There exists⁴ $W \subset \coprod_{\ell \leq \ell_0} B(0, R)^\ell$, $r > 0$ and $n_2 \geq n_1 > 0$ such that $\mathbb{P}(W) \geq 1 - \delta/3$ and for every $x = (y_1, \dots, y_\ell) \in W$ if $\tilde{x} = (\tilde{y}_1, \dots, \tilde{y}_\ell)$ is such that $\|x - \tilde{x}\| < r$ and $n \geq n_2$ then:

$$(4.4) \quad \bigcup_{k=1}^{\ell} B(y_k, \alpha) \simeq \bigcup_{k=1}^{\ell} \hat{B}(\varphi_n(\tilde{y}_k), \alpha n^{-1/d}),$$

(i.e. the two spaces are homotopy equivalent), and for every connected component of $\bigcup_{k=1}^{\ell} B(y_k, \alpha)$ this component intersects $\partial B(0, R)$ if and only if the corresponding component of $\bigcup_{k=1}^{\ell} \hat{B}(\varphi_n(\tilde{y}_k), \alpha n^{-1/d})$ intersects $\partial \hat{B}(p, Rn^{-1/d})$.

Let us explain this condition. A point $x = (y_1, \dots, y_\ell)$ in W corresponds to the spatial Poisson event: we sample ℓ points and each of these points is sampled uniformly from $B(0, R)$. With probability at least $1 - \delta/3$, by the previous point, the number of samples of the spatial Poisson distribution is at most ℓ_0 . Each such Poisson sample in \mathbb{R}^d gives rise to a geometric complex in \mathbb{R}^d , and this complex is nondegenerate with probability one. Given such a nondegenerate complex $\bigcup_{k=1}^{\ell} B(y_k, \alpha)$ in \mathbb{R}^d , we can perturb “a little” (how little is quantified by “ $r > 0$ ”) the point $x = (y_1, \dots, y_\ell)$ to a point $\tilde{x} = (\tilde{y}_1, \dots, \tilde{y}_\ell)$ inside W and still get a nondegenerate complex which has the same homotopy type of the original one. Now, to each point $\tilde{x} \in W$ there also corresponds a complex $\bigcup_{k=1}^{\ell} \hat{B}(\varphi_n(\tilde{y}_k), \alpha n^{-1/d})$ in the manifold M , through the map φ_n . When n is “large enough”, it is natural to expect that the two complexes $\bigcup_{k=1}^{\ell} B(y_k, \alpha)$ and $\bigcup_{k=1}^{\ell} \hat{B}(\varphi_n(\tilde{y}_k), \alpha n^{-1/d})$ have the same homotopy type. Point (2) says that, given ℓ_0 , we can find W with probability at least $1 - \delta/3$, $r > 0$ small enough and $n > 0$ large enough such that this is true. (We also added the requirement that the intersection with the boundary of the big containing ball is the same, but the essence of point (2) is in the requirement that the two complexes should have the same homotopy type.)

- (3) Assuming point (1), denoting by $\{x_1, \dots, x_\ell\} = P_R \cap B(0, R)$, and by $\{\tilde{x}_1, \dots, \tilde{x}_\ell\} = \psi_n(U_n \cap \hat{B}(p, Rn^{-1/d}))$, there exists $n_3 \geq n_1 > 0$ such that for every $n \geq n_3$:

$$\mathbb{P}\{\forall \ell \leq \ell_0, \forall k = 1, \dots, \ell, \|x_k - \tilde{x}_k\| \leq r\} \geq 1 - \delta/3.$$

Assuming these three facts, (4.2) follows arguing as follows. With probability at least $1 - \delta$ for $n \geq n_0 = \max\{n_1, n_2, n_3\}$ all the conditions from (1), (2) and (3) verify and the two random sets

$$\bigcup_{p \in \mathcal{P}_R} B(p, \alpha) \quad \text{and} \quad \bigcup_{p_k \in U_n \cap \hat{B}(p, Rn^{-1/d})} \hat{B}(p_k, \alpha n^{-1/d})$$

are homotopy equivalent and by the second part of point (2) also the unions of all the components entirely contained in $B(0, R)$ (respectively $\hat{B}(p, Rn^{-1/d})$) are homotopy equivalent. In particular the number of components of a given homotopy type γ is the same for both sets with probability at least $1 - \delta$.

It remains to prove (1), (2) and (3).

⁴The symbol $\coprod A_j$ denotes the disjoint union of the sets A_j .

Point (1) follows from the fact that (working with the representation provided by Skorokhod's theorem), the point process $\psi_n(U_n \cap \hat{B}(p, Rn^{-1/d}))$ converges almost surely to the Poisson point process on $B(0, R)$. In particular the sequence of random variables $\{\#(U_n \cap \hat{B}(p, Rn^{-1/d}))\}$ converges almost surely to $\#(P_R \cap B(0, R))$ and (4.3) follows from the fact that almost sure convergence implies convergence in probability.

For point (2) we argue as follows. Given ℓ_0 we consider the compact semialgebraic set:

$$X = \coprod_{\ell \leq \ell_0} B(0, R)^\ell.$$

This set is endowed with the measure $d\rho$:

$$d\rho = \sum_{\ell \leq \ell_0} \frac{\text{vol}(B(0, R)^\ell)}{\ell!} \chi_{B(0, R)^\ell} d\lambda_{B(0, R)^\ell}$$

where $d\lambda$ denotes the Lebesgue measure (this is the measure induced from the Poisson distribution).

Let now $Z \subset X$ be the set of points $x = (y_1, \dots, y_\ell)$ such that either the intersection $\bigcap_{j \in J_1} \partial B(y_j, \alpha)$ or the intersection $\partial B(0, R) \bigcap_{j \in J_2} \partial B(y_j, \alpha)$ is non-transversal for some index sets $J_1, J_2 \in \binom{[d]}{\ell}$ (note that the generic intersection of more than d spheres will be empty). This set Z is also a semialgebraic set, and it has measure zero: it cannot contain any open set, because the nondegeneracy condition is open and dense.

Let $U(Z)$ be an open neighborhood of Z such that $\rho(U(Z)^c) \geq 1 - \delta/3$ (for example one can take $U(Z) = \coprod_{\ell \leq \ell_0} \{d(\cdot, Z) < \varepsilon\}$ for $\varepsilon > 0$ small enough). We set $W = U(Z)^c$ (note that $\mathbb{P}(W) \geq 1 - \delta/3$).

We will first argue that for every $x \in U(Z)^c \subset W$ we can find $r_2(x) > 0$ and $n(x) > 0$ such that the two complexes (4.4) have the same homotopy type (and the same combinatorics of intersection with the boundary of the big containing ball) whenever $\|\tilde{x} - x\| < r_2(x)$ and $n > n(x)$. Then we will use the compactness of $U(Z)^c$ in order to find uniform $r > 0$ and $n > 0$.

Pick therefore $x = (y_1, \dots, y_\ell) \in U(Z)^c$. The property of transversal intersection implies that for every index set $J_1 \in \binom{[d]}{\ell}$ such that the intersection $\bigcap_{j \in J_1} B(y_j, \alpha)$ is nonempty, this intersection contains a nonempty open set, and there exists a point $\sigma_{J_1}(x)$ such that for every $j \in J_1$ we have $\|y_j - \sigma_{J_1}(x)\| < \alpha$. Similarly for every $J_2 \in \binom{[d]}{\ell}$ whenever an intersection $\partial B(0, R) \bigcap_{j \in J_2} \partial B(y_j, \alpha)$ is transversal and nonempty, there exists a point $\sigma_{J_2}(x)$ such that $\|\sigma_{J_2}(x)\| > R$ and for every $j \in J_2$ we have $\|y_j - \sigma_{J_2}(x)\| < \alpha$. Because these are open properties, there exists $r_1(x), r_2(x) > 0$ such that for every $w = (w_1, \dots, w_\ell)$ and $z = (z_1, \dots, z_\ell)$ with $\|w_j - y_j\| \leq r_1(x)$ and $\|z_j - w_j\| < r_2(x)$ for all $j = 1, \dots, \ell$, we have:

$$\forall j \in J_1 : \|z_j - \sigma_{J_1}(x)\| < \alpha \quad \text{and} \quad \forall j \in J_2 : \|z_j - \sigma_{J_2}(x)\| < \alpha.$$

Moreover since the property of having non-empty intersection is also stable under small perturbations, we can assume that $r_1(x), r_2(x)$ are small enough to guarantee also that:

$$\bigcap_{j \in J_3} B(z_j, \alpha) = \emptyset \iff \bigcap_{j \in J_3} B(x_j, \alpha) = \emptyset.$$

Observe now that the sequence of functions $d_n : B(0, R) \times B(0, R) \rightarrow \mathbb{R}$ defined by:

$$d_n(x_1, x_2) = d_M(\varphi_n(x_1), \varphi_n(x_2))n^{1/d}$$

converges uniformly to the Euclidean distance in \mathbb{R}^d . In particular there exists $n(x) > 0$ such that for every $n \geq n(x)$, for every $w = (w_1, \dots, w_\ell)$ and $z = (z_1, \dots, z_\ell)$ with $\|w_j - y_j\| \leq r_1(x)$ and $\|z_j - w_j\| < r_2(x)$, for all $j = 1, \dots, \ell$ and for $i = 1, 2$ we have:

$$d_M(\varphi_n(z_j), \varphi_n(\sigma_{J_i}(x))) < \alpha n^{-1/d}.$$

Moreover, for a possibly larger $n(x)$, we also have that

$$\bigcap_{j \in J_3} \hat{B}(\varphi_n(z_j), \alpha n^{-1/d}) = \emptyset \iff \bigcap_{j \in J_3} B(x_j, \alpha) = \emptyset.$$

Choosing $n(x)$ to be even larger, so that balls of radius smaller than $\alpha n^{-1/d}$ in M are geodesically convex, these conditions imply that the combinatorics of the covers

$$\{B(x_j, \alpha)\}_{j=1}^\ell \quad \text{and} \quad \{\hat{B}(\varphi_n(z_j), \alpha n^{-1/d})\}_{j=1}^\ell$$

are the same and, by Lemma 2.1, the two sets $\bigcup_{j \leq \ell} B(x_j, \alpha)$ and $\bigcup_{j \leq \ell} \hat{B}(\varphi_n(z_j), \alpha n^{-1/d})$ are homotopy equivalent. Also, the above condition on $\sigma_{J_2}(x)$ implies that a component of $\bigcup_{j \leq \ell} B(x_j, \alpha)$ intersects $\partial B(0, R)$ if and only if the corresponding component of $\bigcup_{j \leq \ell} \hat{B}(\varphi_n(z_j), \alpha n^{-1/d})$ intersects $\hat{B}(p, Rn^{-1/d})$.

Finally, we cover now $W = X \setminus U(Z)$ with the family of open sets $\bigcup_{x \in W} B(x, r_1(x))$ and find, by compactness of W , finitely many points x_1, \dots, x_L such that the union of the balls $B(x_k, r_1(x_k))$ with $k = 1, \dots, L$ covers W . With the choice $n_2 = \max\{n(x_k), k = 1, \dots, L\}$ and $r = \min\{r_2(x_k), k = 1, \dots, L\}$ property (2) is true.

Concerning point (3), we observe that again this follows from the fact that the point process $\psi_n(U_n \cap \hat{B}(p, Rn^{-1/d}))$ converges almost surely (hence in probability) to the Poisson point process on $B(0, R)$. \square

Corollary 4.2. *For each $\gamma \in \mathcal{G}$, $\alpha > 0$, $x \in M$, and $\varepsilon > 0$, we have*

$$\lim_{R \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{\mathcal{N}(\mathcal{U}_n, \hat{B}(x, Rn^{-1/d}); \gamma)}{\text{Vol}(B_R)} - c_\gamma \right| > \varepsilon \right\} = 0.$$

Proof. This follows from Theorem 4.1 combined with Proposition 3.1. Indeed, let $\varepsilon > 0$ and $\delta > 0$ be arbitrary. By Proposition 3.1 there exists R_0 such that for $R > R_0$ we have

$$\mathbb{P} \left\{ \left| \frac{\mathcal{N}(\mathcal{P}, B(x, R); \gamma)}{\text{Vol}(B_R)} - c_\gamma \right| > \varepsilon \right\} < \delta.$$

Fix any such $R > R_0$. The event

$$\left| \frac{\mathcal{N}(\mathcal{U}_n, \hat{B}(x, Rn^{-1/d}); \gamma)}{\text{Vol}(B_R)} - c_\gamma \right| > \varepsilon$$

is contained in the union of the event E_1 that

$$\left| \frac{\mathcal{N}(\mathcal{P}, B(x, R); \gamma)}{\text{Vol}(B_R)} - c_\gamma \right| > \varepsilon$$

and another event E_δ , which is the event that $\mathcal{N}(\mathcal{P}, B(x, R); \gamma) \neq \mathcal{N}(\mathcal{U}_n, \hat{B}(x, Rn^{-1/d}); \gamma)$. Thus,

$$(4.5) \quad \mathbb{P} \left\{ \left| \frac{\mathcal{N}(\mathcal{U}_n, \hat{B}(x, Rn^{-1/d}); \gamma)}{\text{Vol}(B_R)} - c_\gamma \right| > \varepsilon \right\} \leq \mathbb{P}\{E_1\} + \mathbb{P}\{E_\delta\} < \delta + \mathbb{P}\{E_\delta\}.$$

By Theorem 4.1, there exists n_0 such that for all $n \geq n_0$ we have $\mathbb{P}\{E_\delta\} \leq \delta$.

Thus, applying this to (4.5) we obtain

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{\mathcal{N}(\mathcal{U}_n, \hat{B}(x, Rn^{-1/d}); \gamma)}{\text{Vol}(B_R)} - c_\gamma \right| > \varepsilon \right\} < 2\delta.$$

Since $\delta > 0$ was arbitrary, this completes the proof of Corollary 4.2. \square

5. THE GLOBAL COUNT FOR THE RIEMMANIAN CASE: PROOF OF THEOREM 1.1

In this section we establish the limit law in the manifold setting (cf. [19, Sec. 7]). As in the Euclidean case, the main step is to prove coefficient-wise convergence, which is stated in the following theorem.

Theorem 5.1. *For every $\gamma \in \mathcal{G}$, the random variable*

$$c_{n,\gamma} = \frac{\mathcal{N}(\mathcal{U}_n, M; \gamma)}{n}$$

converges in L^1 to the constant $c_\gamma = c_\gamma(\alpha)$ (the same constant as in Proposition 3.1). The same statement is true for the random variable

$$c_n = \frac{\mathcal{N}(\mathcal{U}_n, M)}{n}$$

(i.e. when we consider all components, with no restriction on their type): as $n \rightarrow \infty$, it converges in L^1 to the constant $c = \sum_{\gamma \in \mathcal{G}} c_\gamma$.

5.1. Proof of Theorem 1.1 assuming Theorem 5.1. Since convergence in L^1 implies convergence in probability, Theorem 5.1 ensures that the random variable $c_{n,\gamma} = \frac{\mathcal{N}(\mathcal{U}_n, M; \gamma)}{n}$ converges in probability to the constant c_γ ; similarly the random variable $c_n = \frac{\mathcal{N}(\mathcal{U}_n, M)}{n}$ converges in L^1 (hence in probability) to $c > 0$. The proof now proceeds similarly to the proof of Theorem 1.3. We write the measure $\hat{\mu}_n$ as:

$$\begin{aligned} \hat{\mu}_n &= \frac{1}{b_0(\tilde{C}(\mathcal{U}_n))} \sum_{\gamma \in \hat{\mathcal{G}}} \mathcal{N}(\mathcal{U}_n, M; \gamma) \delta_\gamma \\ &= \frac{1}{b_0(\tilde{C}(\mathcal{U}_n))} \left(\sum_{\gamma \in \mathcal{G}} \mathcal{N}(\mathcal{U}_n, M; \gamma) \delta_\gamma + \sum_{\gamma \in \hat{\mathcal{G}} \setminus \mathcal{G}} \mathcal{N}(\mathcal{U}_n, M; \gamma) \delta_\gamma \right) \\ &= \frac{1}{\mathcal{N}(\mathcal{U}_n, M)} \sum_{\gamma \in \mathcal{G}} \mathcal{N}(\mathcal{U}_n, M; \gamma) \delta_\gamma + \frac{1}{\mathcal{N}(\mathcal{U}_n, M)} \sum_{\gamma \in \hat{\mathcal{G}} \setminus \mathcal{G}} \mathcal{N}(\mathcal{U}_n, M; \gamma) \delta_\gamma \\ &= \frac{n}{\mathcal{N}(\mathcal{U}_n, M)} \sum_{\gamma \in \mathcal{G}} \frac{\mathcal{N}(\mathcal{U}_n, M; \gamma)}{n} \delta_\gamma + \frac{n}{\mathcal{N}(\mathcal{U}_n, M)} \sum_{\gamma \in \hat{\mathcal{G}} \setminus \mathcal{G}} \frac{\mathcal{N}(\mathcal{U}_n, M; \gamma)}{n} \delta_\gamma \\ (5.1) \quad &= \sum_{\gamma \in \mathcal{G}} \frac{c_{n,\gamma}}{c_n} \delta_\gamma + \sum_{\gamma \in \hat{\mathcal{G}} \setminus \mathcal{G}} \frac{c_{n,\gamma}}{c_n} \delta_\gamma. \end{aligned}$$

We have $a_{n,\gamma} = \frac{c_{n,\gamma}}{c_n}$ converges in L^1 to the constant $a_\gamma = \frac{c_\gamma}{c}$. Recalling that the measure

$$\mu = \sum_{\gamma \in \mathcal{G}} a_\gamma \delta_\gamma$$

is a probability measure (see Proposition 3.4), we can apply Lemma 2.4 to conclude that the measure on the left in (5.1) converges in probability to μ .

Since μ is a probability measure, this implies that

$$\sum_{\gamma \in \mathcal{G}} \frac{c_{n,\gamma}}{c_n}$$

converges to 1 in probability. For any $\gamma_0 \in \hat{\mathcal{G}} \setminus \mathcal{G}$ this implies that the coefficient $\frac{c_{n,\gamma_0}}{c_n}$ appearing in the measure on the right in (5.1) converges to zero in probability, since

$$0 \leq \frac{c_{n,\gamma_0}}{c_n} \leq 1 - \sum_{\gamma \in \mathcal{G}} \frac{c_{n,\gamma}}{c_n}.$$

Thus, the measure $\hat{\mu}_n$ converges in probability to μ by another application of Lemma 2.4.

5.2. Proof of Theorem 5.1. Note: Since $\alpha > 0$ and $\gamma \in \mathcal{G}$ are fixed, we will simply use

$$(5.2) \quad \mathcal{N}_n := \mathcal{N}(\mathcal{U}_n, M; \gamma)$$

to denote the number of components of \mathcal{U}_n in M of type γ . We will use

$$\mathcal{N}_n^*(x, r) := \mathcal{N}^*(\mathcal{U}_n, \hat{B}(x, r); \gamma)$$

to denote the number of such components intersecting the geodesic ball $\hat{B}(x, r)$ of radius r centered at x and

$$\mathcal{N}_n(x, r) := \mathcal{N}(\mathcal{U}_n, \hat{B}(x, r); \gamma)$$

to denote the number of components completely contained in $\hat{B}(x, r)$.

Thus, our goal, stated in the abbreviated notation (5.2), is to prove

$$(5.3) \quad \mathbb{E} \left[\left| \frac{\mathcal{N}_n}{n} - c_\gamma \right| \right] \rightarrow 0.$$

Using the integral geometry sandwich from Theorem 2.7 we have

$$(5.4) \quad (1 - \varepsilon) \int_M \frac{\mathcal{N}_n(x, Rn^{-1/d})}{\text{Vol}(B_R)} dx \leq \frac{\mathcal{N}_n}{n} \leq (1 + \varepsilon) \int_M \frac{\mathcal{N}_n^*(x, Rn^{-1/d})}{\text{Vol}(B_R)} dx.$$

Letting I_1 denote the integral on the left side and I_2 the one on the right side, we subtract I_1 from each part of (5.4) and write

$$(5.5) \quad -\varepsilon I_1 \leq \frac{\mathcal{N}_n}{n} - I_1 \leq \varepsilon I_1 + (1 + \varepsilon)(I_2 - I_1).$$

In order to estimate $I_2 - I_1$ we note that the number of connected components of \mathcal{U}_n that intersect, but are not completely contained in, the geodesic ball $\hat{B}(x, Rn^{-1/d})$ is bounded above by the number of points that fall within distance $\alpha n^{-1/d}$ to the boundary $\partial \hat{B}(x, Rn^{-1/d})$. This $\alpha n^{-1/d}$ -neighborhood of $\partial \hat{B}(x, Rn^{-1/d})$ is the same as the geodesic annulus centered at x with inner radius $(R - \alpha)n^{-1/d}$ and outer radius $(R + \alpha)n^{-1/d}$. The average number of points in this annulus equals its volume which can be estimated (uniformly over $x \in M$) by that of the Euclidean annulus, and this gives

$$\mathbb{E}|I_2 - I_1| = O(R^{-1}).$$

which together with (5.5) implies

$$(5.6) \quad \mathbb{E} \left| \frac{\mathcal{N}_n}{n} - I_1 \right| = O(\varepsilon) + O(R^{-1}),$$

where we have also used $I_1 \leq \frac{1}{1-\varepsilon}$ which follows from the first inequality in (5.4) along with the simple estimate $\mathcal{N}_n \leq n$.

By (5.6) we obtain

$$\begin{aligned} \mathbb{E} \left[\left| \frac{\mathcal{N}_n}{n} - c_\gamma \right| \right] &= \mathbb{E} \left[\left| \frac{\mathcal{N}_n}{n} - I_1 + I_1 - c_\gamma \right| \right] \\ &\leq \mathbb{E} [|I_1 - c_\gamma|] + O(\varepsilon) + O(R^{-1}). \\ &= \mathbb{E} \left[\left| \int_M \frac{\mathcal{N}_n(x, Rn^{-1/d})}{\text{Vol}(B_R)} - c_\gamma dx \right| \right] + O(\varepsilon) + O(R^{-1}). \end{aligned}$$

Thus, in order to prove the theorem it suffices to show that the above term $\mathbb{E}[|I_1 - c_\gamma|]$ can be made arbitrarily small for all sufficiently large n .

Define the “bad” event

$$\Omega_{x,R,n} := \left\{ \left| \frac{\mathcal{N}_n(x, Rn^{-1/d})}{\text{Vol}(B_R)} - c_\gamma \right| > \varepsilon \right\}.$$

Claim: There exists a sequence $R_j \rightarrow \infty$ such that for every $\delta > 0$ there exists $M_\delta \subset M$ with $\text{Vol}(M_\delta) > 1 - \delta$ such that

$$(5.7) \quad \lim_{R_j \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{x \in M_\delta} \mathbb{P}(\Omega_{x,R_j,n}) = 0.$$

The proof of this claim closely follows [19] and uses Egorov’s theorem as well as the idea from the proof of Egorov’s theorem. We start by recalling the point-wise limit stated in Corollary 4.2. For each $x \in M$, we have

$$\lim_{R \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\{\Omega_{x,R,n}\} = 0.$$

Let us restrict to $R \in \mathbb{N}$. Apply Egorov’s theorem to obtain $M'_\delta \subset M$ with $\text{Vol}(M'_\delta) > 1 - \frac{\delta}{2}$ such that

$$(5.8) \quad \lim_{R \rightarrow \infty} \sup_{x \in M'_\delta} \limsup_{n \rightarrow \infty} \mathbb{P}\{\Omega_{x,R,n}\} = 0.$$

Next we use an additional Egorov-type argument in order to obtain the statement in the claim (where we will obtain the set M_δ by slightly shrinking M'_δ). For each fixed integer $j > 0$, we can find by (5.8) an $R_j \in \mathbb{N}$ sufficiently large so that

$$(5.9) \quad \sup_{x \in M'_\delta} \limsup_{n \rightarrow \infty} \mathbb{P}\{\Omega_{x,R_j,n}\} < \frac{1}{j}.$$

Letting $F_m(j)$ denote the monotone decreasing (with m) sequence of sets

$$F_m(j) = \bigcup_{k \geq m} \left\{ x \in M'_\delta : \mathbb{P}(\Omega_{x,R_j,k}) > \frac{2}{j} \right\},$$

we see from (5.9) that

$$\bigcap_{m \geq 1} F_m(j) = \emptyset.$$

Thus, there exists $m = m(j)$ such that $\text{Vol}(F_m(j)) < \frac{\delta}{2j+1}$. We take

$$M_\delta = M'_\delta \setminus \left(\bigcup_{j \geq 1} F_{m(j)}(j) \right),$$

which satisfies $\text{Vol}(M_\delta) > \text{Vol}(M'_\delta) - \frac{\delta}{2} > 1 - \delta$. It follows from the definition of $F_m(j)$ that

$$\limsup_{n \rightarrow \infty} \sup_{x \in M_\delta} \mathbb{P}\{\Omega_{x,R_j,n}\} \leq \frac{2}{j},$$

and we see that (5.7) is satisfied.

Denoting the whole probability space as Ω , we separate the integration (defining the expectation) over the two sets $\Omega_{x,R_j,n}$ and $\Omega \setminus \Omega_{x,R_j,n}$.

$$(5.10) \quad \mathbb{E}[|I_1 - c_\gamma|] = \int_{\Omega \setminus \Omega_{x,R_j,n}} |I_1 - c_\gamma| d\omega + \int_{\Omega_{x,R_j,n}} |I_1 - c_\gamma| d\omega.$$

We use the definition of $\Omega_{x,R_j,n}$ to estimate the first integral in (5.10):

$$(5.11) \quad \int_{\Omega \setminus \Omega_{x,R_j,n}} |I_1 - c_\gamma| d\omega \leq \int_{\Omega \setminus \Omega_{x,R_j,n}} \int_M \left| \frac{\mathcal{N}_n(x, Rn^{-1/d})}{\text{Vol}(B_R)} - c_\gamma \right| dx d\omega \leq \varepsilon.$$

For the second integral in (5.10), we use the estimate

$$(5.12) \quad \frac{\mathcal{N}_n(x, Rn^{-1/d})}{\text{Vol}(B_R)} \leq (1 + \varepsilon)n\xi^{-1} = O(1),$$

where $\xi > 0$ is the minimum (over $x \in M$) volume of a geodesic ball of radius $\alpha n^{-1/d}$, which is uniformly (over $x \in M$) comparable to the volume of the Euclidean ball of the same radius, and hence ξ is bounded below by a constant times n^{-1} . The estimate (5.12) is based on the fact that each component has volume trivially at least a constant times α^d/n , and the fact that the minimal volume of a component times the number of components cannot exceed the volume of the region where they are contained (while fixing attention on components of type γ as we are throughout the proof). Applying (5.12), we obtain

$$(5.13) \quad \begin{aligned} \int_{\Omega_{x,R_j,n}} |I_1 - c_\gamma| d\omega &\leq \int_M \int_{\Omega_{x,R_j,n}} \left| \frac{\mathcal{N}_n(x, Rn^{-1/d})}{\text{Vol}(B_R)} - c_\gamma \right| d\omega dx \\ &\leq (O(1) + c_\gamma) \cdot \int_M \mathbb{P}(\Omega_{x,R_j,n}) dx \\ &= O(1) \cdot \int_M \mathbb{P}(\Omega_{x,R_j,n}) dx. \end{aligned}$$

Next, we split this last integration over M_δ and $M \setminus M_\delta$:

$$(5.14) \quad \begin{aligned} \int_M \mathbb{P}(\Omega_{x,R_j,n}) dx &= \int_{M_\delta} \mathbb{P}(\Omega_{x,R_j,n}) dx + \int_{M \setminus M_\delta} \mathbb{P}(\Omega_{x,R_j,n}) dx \\ &\leq \sup_{x \in M_\delta} \mathbb{P}(\Omega_{x,R_j,n}) + \delta. \end{aligned}$$

Bringing the estimates (5.10), (5.11), (5.13), (5.14) together, we have

$$\mathbb{E} [|I_1 - c_\gamma|] \leq \varepsilon + O(1) \left(\delta + \sup_{x \in M_\delta} \mathbb{P}(\Omega_{x,R_j,n}) \right),$$

which can be made arbitrarily small using (5.7). This establishes (5.3) and completes the proof of the first part of Theorem 5.1. The proof of the second part concerning the count for all components (without restriction on homotopy type) follows from the same proof while replacing the integral geometry sandwich with its more basic version (see Remark 6 in Section 2).

6. POSITIVITY OF ALL COEFFICIENTS

In this section, we prove Propositions 3.2 and 1.2.

6.1. **Proof of Proposition 3.2.** Recall that, by Proposition 3.1, for every $\gamma \in \mathcal{G}$ we have:

$$c_\gamma = \lim_{R \rightarrow \infty} \mathbb{E} \left(\frac{\mathcal{N}(\mathcal{P}, B(0, R); \gamma)}{\text{Vol}(B(0, R))} \right).$$

The desired lower bound will come from adding up certain local contributions provided by the following lemma

Lemma 6.1. *Let $\mathcal{P}_0 \subset \mathbb{R}^d$ be a finite geometric complex. Fix $\alpha > 0$. There exist $r, a > 0$ (depending on \mathcal{P}_0 and α) such that for any $p \in \mathbb{R}^d$*

$$\mathbb{P} \{ \mathcal{P} \cap B(p, r) \simeq \mathcal{P}_0 \} > a.$$

Proof of Lemma 6.1. By the translation invariance of the Poisson point process, it is enough to prove the statement while taking $p = 0$. We can assume, by Remark 2.2 above, that \mathcal{P}_0 is nondegenerate, and we write

$$\mathcal{P}_0 = \bigcup_{k=1}^{\ell} B(y_k, \alpha),$$

where $y_1, \dots, y_\ell \in \mathbb{R}^d$, and we have taken the radius to be α , since we can dilate the entire set \mathcal{P}_0 if necessary (which does not change the homotopy type). Choose r such that $\mathcal{P}_0 \subset B(0, r - \alpha)$. By nondegeneracy, there exists $\varepsilon > 0$ such that if $\|\tilde{y}_k - y_k\| \leq \varepsilon$ then the two complexes $\bigcup_{k=1}^{\ell} B(y_k, \alpha)$ and $\bigcup_{k=1}^{\ell} B(\tilde{y}_k, \alpha)$ are homotopy equivalent.

We are thus interested in the event E that for $j = 1, 2, \dots, \ell$ each ball $B(y_j, \varepsilon)$ contains exactly one point from the random set $P = \{p_1, p_2, \dots\}$ and that these ℓ points are the only points of P in the ball $B(0, r + \alpha)$. If E occurs then we have $\mathcal{P} \cap B(0, r) \cong \mathcal{P}_0$ as desired. The positivity of the probability of the event E can be seen by noticing that this probability is a product of finitely many positive probabilities. Indeed, E is an intersection of the events that $B(y_j, \varepsilon)$ contains a single point from P for $j = 1, 2, \dots, \ell$ along with the event that there are no points in the set $B(0, r + \alpha) \setminus \bigcup_{i=1}^{\ell} B(y_i, \varepsilon)$. These events are independent by a basic property of the Poisson point process, and each of them has positive probability (the probabilities can be specified explicitly in terms of the volumes of the sets involved). \square

Let r, a be given by Lemma 6.1 for the choice of $[\mathcal{P}_0] = \gamma$. For an appropriate $\beta > 0$ we can fit $k \geq \beta \text{Vol}(B(0, R))$ many disjoint Riemannian balls

$$B_1 = B(p_1, r), \dots, B_k = B(p_k, r)$$

in $B(0, R)$. Observing that

$$\mathcal{N}(\mathcal{P}, B(0, R); \gamma) \geq \sum_{j=1}^k \mathcal{N}(\mathcal{P}, B_j; \gamma),$$

we have (also using linearity of expectation)

$$\mathbb{E} \left(\frac{\mathcal{N}(\mathcal{P}, B(0, R); \gamma)}{\text{Vol}(B(0, R))} \right) \geq \beta a > 0,$$

and the positivity of c_γ follows.

6.2. Proof of Proposition 1.2. The positivity of coefficients a_γ in Theorem 1.1 is already established in Proposition 3.2 proved above. In this section, we prove the related Proposition 1.2 which provides a more direct analysis in the manifold setting.

Proof. Let $y_1, \dots, y_\ell \in \mathbb{R}^d$ and $r > 0$ such that

$$\bigcup_{k=1}^{\ell} B(y_k, r) = \mathcal{P}$$

with $\bigcup_{k=1}^{\ell} B(y_k, r)$ a nondegenerate complex (it is not restrictive to consider nondegenerate complexes by Remark 2.2 above). Let now $R' > 0$ such that $B(0, R')$ contains $\bigcup_{k=1}^{\ell} B(y_k, r)$ and set $R = \frac{R'\alpha}{r}$. Consider also the sequence of maps:

$$\psi_n : \hat{B}(p, Rn^{-1/d}) \xrightarrow{\exp_p^{-1}} B_{T_p M}(0, Rn^{-1/d}) \xrightarrow{\frac{r}{\alpha}n^{1/d}} B_{T_p M}(0, R') \simeq B(0, R').$$

Proposition 2.3 implies that there exists $\epsilon_0 > 0$ and n_0 such that if $\|\tilde{y}_k - y_k\| \leq \epsilon_0$ then for $n \geq n_0$ the two complexes $\bigcup_{k=1}^{\ell} B(y_k, r)$ and $\bigcup_{k=1}^{\ell} \hat{B}(\varphi_n(\tilde{y}_k), \alpha n^{-1/d})$ are homotopy equivalent.

We are interested in the the event:

$$E_n = \left\{ \exists I_\ell \in \binom{[n]}{\ell} : \forall j \in I_\ell \quad p_j \in \psi_n^{-1}(B(y_j, \epsilon)), \text{ and } \forall j \notin I_\ell \quad p_j \in \hat{B}(p, (R + \alpha)n^{-1/d})^c \right\}.$$

Observe that if E_n verifies, then $\mathcal{U}_n \cap \hat{B}(p, Rn^{-1/d}) \simeq \mathcal{P}$: in fact, since there is no other point in $\hat{B}(p, (R + \alpha)n^{-1/d})$ other than $\{p_j\}_{j \in I_\ell}$, then the complex \mathcal{U}_n is the disjoint union of the two complexes $\mathcal{U}_n \cap \hat{B}(p, Rn^{-1/d})$ and $\mathcal{U}_n \cap \hat{B}(p, (R + \alpha)n^{-1/d})^c$; the complex $\mathcal{U}_n \cap \hat{B}(p, Rn^{-1/d}) \simeq \mathcal{P}$ by Proposition 2.3.

It is therefore enough to estimate from below the probability of E_n . Note that for every measurable subset $B \subset B(0, R')$ there exists a constant $c_B > 0$ such that $\text{Vol}(\psi_n^{-1}(B)) \geq \frac{c_B}{n}$. In particular, using the independence of the points in U_n , we can estimate

$$\begin{aligned} \mathbb{P}(E_n) &= \binom{n}{\ell} \mathbb{P} \left\{ \forall j \leq \ell : p_j \in \psi_n^{-1}(B(y_j, \epsilon)) \quad \text{and} \quad \forall j \geq \ell + 1 : p_j \in \hat{B}(p, (R + \alpha)n^{-1/d})^c \right\} \\ &= \binom{n}{\ell} \left(\prod_{j=1}^{\ell} \text{vol}(\psi_n^{-1}(B(y_j, \epsilon))) \right) \left(\text{vol}(\hat{B}(p, (R + \alpha)n^{-1/d})^c) \right)^{n-\ell} \\ &\geq \binom{n}{\ell} \left(\frac{c_1}{n} \right)^\ell \left(1 - \frac{c_2}{n} \right)^{n-\ell} \xrightarrow{n \rightarrow \infty} \frac{c_1^\ell}{\ell!} (1 - c_2)^{-\ell} e^{-c_2}. \end{aligned}$$

In particular there exists $c > 0$ such that:

$$\mathbb{P} \left\{ \mathcal{U}_n \cap \hat{B}(p, Rn^{-1/d}) \simeq \mathcal{P} \right\} \geq \mathbb{P}(E_n) > c,$$

and this concludes the proof. \square

REFERENCES

- [1] V. I. Arnold. *Arnold's problems*. Springer-Verlag, Berlin; PHASIS, Moscow, 2004. Translated and revised edition of the 2000 Russian original, With a preface by V. Philippov, A. Yakivchik and M. Peters.
- [2] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry*, volume 10 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, second edition, 2006.
- [3] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [4] O. Bobrowski and M. Kahle. Topology of random geometric complexes: a survey. *J. Appl. Comput. Topol.*, 1(3):331–364, Jun 2018.

- [5] O. Bobrowski and S. Mukherjee. The topology of probability distributions on manifolds. *Probab. Theory Related Fields*, 161(3-4):651–686, 2015.
- [6] O. Bobrowski and S. Weinberger. On the vanishing of homology in random Čech complexes. *Random Structures Algorithms*, 51(1):14–51, 2017.
- [7] J. Cheeger and D. G. Ebin. *Comparison theorems in Riemannian geometry*. AMS Chelsea Publishing, Providence, RI, 2008. Revised reprint of the 1975 original.
- [8] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II. Probability and its Applications* (New York). Springer, New York, second edition, 2008. General theory and structure.
- [9] M. P. a. do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.
- [10] K. A. Dowling and E. Lundberg. Homotopy types of random cubical complexes. *Preprint at arXiv:1910.12803*, 2019.
- [11] A. Goel, K. D. Trinh, and K. Tsunoda. Strong law of large numbers for Betti numbers in the thermodynamic regime. *J. Stat. Phys.*, 174(4):865–892, Feb 2019.
- [12] Y. Hiraoka, T. Shirai, and K. D. Trinh. Limit theorems for persistence diagrams. *Ann. Appl. Probab.*, 28(5):2740–2780, 10 2018.
- [13] M. Kahle. Random geometric complexes. *Discrete Comput. Geom.*, 45(3):553–573, 2011.
- [14] R. Meester and R. Roy. *Continuum percolation*, volume 119 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.
- [15] F. Nazarov and M. Sodin. Asymptotic laws for the spatial distribution and the number of connected components of zero sets of Gaussian random functions. *Zh. Mat. Fiz. Anal. Geom.*, 12(3):205–278, 2016.
- [16] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, 2008.
- [17] M. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.
- [18] S. I. Resnick. *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. Reprint of the 1987 original.
- [19] P. Sarnak and I. Wigman. Topologies of nodal sets of random band-limited functions. *Comm. Pure Appl. Math.*, 72(2):275–342, 2019.
- [20] D. Yogeshwaran, E. Subag, and R. J. Adler. Random geometric complexes in the thermodynamic regime. *Probab. Theory Related Fields*, 167(1-2):107–142, 2017.

Antonio Auffinger

Northwestern University, Department of Mathematics

Evanston, IL 60208

Antonio Lerario

SISSA, Mathematics

Trieste, Italy

Erik Lundberg

Florida Atlantic University, Department of Mathematical Sciences

Boca Raton, FL 33431

email: elundber@fau.edu