

ORIGINAL ARTICLE

Open Access



A remark on the convergence of Betti numbers in the thermodynamic regime

Khanh Duy Trinh

Abstract

The convergence of the expectations of Betti numbers of Čech complexes built on binomial point processes in the thermodynamic regime is established.

Keywords: Čech complex, Betti number, Binomial point process, Thermodynamic regime

AMS Subject Classification: Primary 55N05, 60F99

1 Terminologies and main results

Definition 1.1 (Čech complex) Let $\mathfrak{X} = \{x_1, x_2, \dots, x_n\}$ be a collection of points in \mathbb{R}^d . The Čech complex $\mathcal{C}(\mathfrak{X}, r)$, for $r > 0$, is constructed as follows.

- (i) The 0-simplices (vertices) are the points in \mathfrak{X} .
- (ii) A k -simplex $[x_{i_0}, \dots, x_{i_k}]$ is in $\mathcal{C}(\mathfrak{X}, r)$ if

$$\bigcap_{j=0}^k B_r(x_{i_j}) \neq \emptyset.$$

Here $B_r(x) = \{y \in \mathbb{R}^d : \|y - x\| \leq r\}$ denotes a ball of radius r and center x , and $\|x\|$ is the Euclidean norm of x . The Čech complex can be also constructed from an infinite collection of points.

Let X_1, X_2, \dots , be a sequence of i.i.d. (independent identically distributed) \mathbb{R}^d -valued random variables with common probability density function $f(x)$. Define the induced binomial point processes as $\mathfrak{X}_n = \{X_1, \dots, X_n\}$. The object here is the Čech complex $\mathcal{C}(\mathfrak{X}_n, r_n)$ built on \mathfrak{X}_n , where the radius r_n also varies with n . Denote by $\beta_k(\mathcal{K})$ the k th Betti number, or the rank of the k th homology group, of a simplicial complex \mathcal{K} . The limiting behaviour of Betti numbers $\beta_k(\mathcal{C}(\mathfrak{X}_n, r_n))$ in various regimes has been studied recently by many authors. See [1] for a brief survey. The aim of this paper is to refine a limit theorem in the thermodynamic regime, a regime that $n^{1/d} r_n \rightarrow r \in (0, \infty)$.

In the thermodynamic regime, the expectations of the k th Betti numbers, for $1 \leq k \leq d - 1$, grow linearly in n ,

that is, $c_1 n \leq \mathbb{E}[\beta_k(\mathcal{C}(\mathfrak{X}_n, r_n))] \leq c_2 n$ as $n \rightarrow \infty$. After centralizing, the strong law of large numbers holds,

$$\frac{1}{n} (\beta_k(\mathcal{C}(\mathfrak{X}_n, r_n)) - \mathbb{E}[\beta_k(\mathcal{C}(\mathfrak{X}_n, r_n))]) \rightarrow 0 \text{ almost surely as } n \rightarrow \infty, \quad (1)$$

provided that the density function f has compact, convex support and that on the support of f , it is bounded both below and above [9, Theorem 4.6]. A remaining problem is to describe the exact limiting behaviour of the expected values of the Betti numbers. This paper gives a solution to that problem. Note that the 0th Betti number which counts connected components in a random geometric graph was completely described [5, Chapter 13]. Note also that the k th Betti number of the Čech complex built on a finite set of points in \mathbb{R}^d is vanishing, if $k \geq d$. These facts explain why we only need to consider the case $1 \leq k \leq d - 1$.

Betti numbers are tightly related to the number of j -simplices in $\mathcal{C}(\mathfrak{X}, r)$, denoted by $S_j(\mathcal{C}(\mathfrak{X}, r))$ or simply by $S_j(\mathfrak{X}, r)$, which can be expressed as

$$S_j(\mathfrak{X}, r) = \frac{1}{j+1} \sum_{x \in \mathfrak{X}} \xi(x, \mathfrak{X}),$$

where $\xi(x, \mathfrak{X})$ is the number of j -simplices containing x . Note that $\xi(x, \mathfrak{X})$ is a local function in the sense that it depends only on points near x . Then in the thermodynamic regime, the weak and strong laws of large numbers for $S_j(\mathcal{C}(\mathfrak{X}_n, r_n))$ hold as a consequence of general results in [6, 7],

$$\frac{S_j(\mathfrak{X}_n, r_n)}{n} \rightarrow \hat{S}_j(r) \text{ almost surely as } n \rightarrow \infty.$$

Correspondence: trinh@imi.kyushu-u.ac.jp
Institute of Mathematics for Industry, Kyushu University, 819-0395 Fukuoka, Japan

The limit $\hat{S}_j(r)$ can be explicitly calculated. However, Betti numbers do not have expression like the above form, and hence those general results can not be applied.

To establish a limit theorem for Betti numbers, we exploit the following two properties. The first one is the nearly additive property of Betti numbers that was used in [9] to study Betti numbers of Čech complexes built on stationary point processes. The second one is the property that binomial point processes behave locally like a homogeneous Poisson point process. The latter property is also a key tool to establish the law of large numbers for local geometric functionals [6, 7].

Now let us get into more detail to state the main result of the paper. We begin with the definition of a homogeneous Poisson point process. Let \mathbf{N} be the set of all counting measures on \mathbb{R}^d which are finite on any bounded Borel set and for which the measure of a point is at most 1. Define \mathcal{N} as the σ -algebra generated by sets of the form

$$\{\mu \in \mathbf{N} : \mu(A) = k\},$$

where A is a bounded Borel set and k is an integer. Then a point process Φ is a measurable mapping from some probability space into $(\mathbf{N}, \mathcal{N})$. For a Borel set A , let $\Phi(A)$ denote the number of points in A . By definition of the σ -algebra \mathcal{N} , $\Phi(A)$ becomes a usual random variable. A homogeneous Poisson point process is defined as follows. For some basic properties of point processes, see [2], for example.

Definition 1.2 (Homogeneous Poisson point process) *The point process \mathcal{P} is said to be a Poisson point process with density $\lambda > 0$ if*

- (i) *for disjoint Borel sets A_1, \dots, A_k , the random variables $\mathcal{P}(A_1), \dots, \mathcal{P}(A_k)$ are independent;*
- (ii) *for any bounded Borel set A , the number of points in A has Poisson distribution with parameter $|A|$, $\mathcal{P}(A) \sim \text{Pois}(\lambda|A|)$, that is,*

$$\mathbb{P}(\mathcal{P}(A) = k) = e^{-\lambda|A|} \frac{\lambda^k |A|^k}{k!}, \quad k = 0, 1, \dots,$$

where $|A|$ denotes the Lebesgue measure of A .

For homogeneous Poisson point processes, the following law of large numbers for Betti numbers was established in [9]. Let $\mathcal{P}(\lambda)$ be a homogeneous Poisson point process on \mathbb{R}^d with density $\lambda > 0$. Denote by $\mathcal{P}_A(\lambda)$ the restriction of $\mathcal{P}(\lambda)$ on a Borel set A . For simplicity, we write $\mathcal{P}_L(\lambda)$ instead of $\mathcal{P}_{W_L}(\lambda)$ when W_L is a window of the form $W_L = [-\frac{L^{1/d}}{2}, \frac{L^{1/d}}{2}]^d$. Then for $1 \leq k \leq d - 1$, there is a constant $\hat{\beta}_k(\lambda, r)$ such that [9, Theorem 3.5],

$$\frac{\beta_k(\mathcal{C}(\mathcal{P}_L(\lambda), r))}{L} \rightarrow \hat{\beta}_k(\lambda, r) \text{ almost surely as } L \rightarrow \infty.$$

The Poisson point process $\mathcal{P}(0)$ is understood as a point process with no point. Thus we set $\hat{\beta}_k(0, r) = 0$ for all $r > 0$. Now we can state our main result.

Theorem 1.3 *Assume that the common probability density function $f(x)$ has compact support, is bounded and Riemann integrable. Then as $n \rightarrow \infty$ with $n^{1/d}r_n \rightarrow r \in (0, \infty)$,*

$$\frac{\mathbb{E}[\beta_k(\mathcal{C}(\mathcal{X}_n, r_n))]}{n} \rightarrow \int_{\mathbb{R}^d} \hat{\beta}_k(f(x), r) dx.$$

Consequently, together with (1), we have the following law of large numbers.

Corollary 1.4 *Assume that the support of f is compact and convex and that*

$$0 < \inf_{x \in \text{supp}(f)} f(x) \leq \sup_{x \in \text{supp}(f)} f(x) < \infty.$$

Assume further that f is Riemann integrable. Then for $1 \leq k \leq d - 1$,

$$\frac{\beta_k(\mathcal{C}(\mathcal{X}_n, r_n))}{n} \rightarrow \int_{\mathbb{R}^d} \hat{\beta}_k(f(x), r) dx \text{ almost surely as } n \rightarrow \infty. \tag{2}$$

It is noted that the method here can be applied to show the convergence of persistence diagrams of Čech complexes built on binomial point processes. The convergence of Betti numbers and persistence diagrams related to i.i.d. sampling were observed in [4] by numerical simulation. Here we give a rigorous mathematical proof of the convergences.

For the proof, we need a Poissonized version of the binomial processes. Let N_n be a random variable which is independent of $\{X_n\}_{n \geq 1}$ and has Poisson distribution with parameter n . Let

$$\bar{\mathcal{P}}_n = \{X_1, X_2, \dots, X_{N_n}\}.$$

Then $\bar{\mathcal{P}}_n$ becomes a non-homogeneous Poisson point process with intensity function $nf(x)$. Here a non-homogeneous Poisson point process is defined as follows.

Definition 1.5 (Non-homogeneous Poisson point process) *Let $f(x) \geq 0$ be a locally integrable function on \mathbb{R}^d . The point process \mathcal{P} is said to be a (non-homogeneous) Poisson point process with intensity function $f(x)$ if*

- (i) *for disjoint Borel sets A_1, \dots, A_k , the random variables $\mathcal{P}(A_1), \dots, \mathcal{P}(A_k)$ are independent;*
- (ii) *for any bounded Borel set A , $\mathcal{P}(A) \sim \text{Pois}(\int_A f(x) dx)$.*

As proved later, Theorem 1.3 is equivalent to the following result.

Theorem 1.6 Assume that the common probability density function $f(x)$ has compact support, is bounded and Riemann integrable. Then as $n \rightarrow \infty$ with $n^{1/d}r_n \rightarrow r \in (0, \infty)$,

$$\frac{\mathbb{E}[\beta_k(\mathcal{C}(\bar{\mathcal{P}}_n, r_n))]}{n} \rightarrow \int_{\mathbb{R}^d} \hat{\beta}_k(f(x), r) dx.$$

All the proofs will be given in Section 3 after discussing some basic properties of Betti numbers in the next section.

2 Simplicial complexes and Betti numbers

This section introduces some basic concepts in algebraic topology such as simplicial complexes and Betti numbers. It is mainly taken from the book [3].

An abstract simplicial complex \mathcal{K} on a finite set V is a collection of nonempty subsets of V which is closed under inclusion relation, that is, if $\sigma \in \mathcal{K}$, then $\tau \in \mathcal{K}$ for any nonempty subset $\tau \subset \sigma$. An element $\sigma \in \mathcal{K}$ with $|\sigma| = k + 1$ is called a k -simplex or a simplex of dimension k . A 0-simplex (resp. 1-simplex) is usually called a vertex (resp. edge). Čech complexes are examples of geometric complexes which are constructed over points in some metric space with respect to certain conditions.

We assign orientations on simplices in the following way. For a k -simplex $\sigma = \{v_0, \dots, v_k\}$ with $k > 0$, define two orderings of its vertex set to be equivalent if they differ from one other by an even permutation. The orderings of the vertices of σ then fall into two equivalent classes. Each of these classes is called an orientation of σ . We write $\langle v_0, \dots, v_k \rangle$ for an oriented simplex. Let us fix an ordering of the vertex set V . Then the notation $\langle \sigma \rangle$ means the oriented simplex which belongs to the equivalent class of a natural ordering. A 0-simplex has only one orientation.

Let \mathbf{F} be a field. For each k , let

$$C_k(\mathcal{K}) = \left\{ \sum \alpha_i \langle \sigma_i \rangle : \alpha_i \in \mathbf{F}, \sigma_i \in \mathcal{K}_k \right\}$$

be a vector space with the basis $\{\langle \sigma \rangle : \sigma \in \mathcal{K}_k\}$, where \mathcal{K}_k is the set of all k -simplices in \mathcal{K} . The space $C_k(\mathcal{K})$ is called a chain group.

The dimension of \mathcal{K} , denoted by $\dim(\mathcal{K})$, is defined to be the maximum dimension of simplices in \mathcal{K} . For $1 \leq k \leq \dim(\mathcal{K})$, a boundary operator $\partial_k: C_k(\mathcal{K}) \rightarrow C_{k-1}(\mathcal{K})$ is a linear map whose value on an oriented simplex $\langle v_0, \dots, v_k \rangle$ is given by

$$\partial_k(\langle v_0, \dots, v_k \rangle) = \sum_{i=0}^k (-1)^k \langle v_0, \dots, \hat{v}_i, \dots, v_k \rangle,$$

where the symbol $\hat{}$ over v_i indicates that the vertex is removed from the sequence. Then we get a chain

$$0 \rightarrow C_{\dim(\mathcal{K})} \xrightarrow{\partial_{\dim(\mathcal{K})}} \dots \rightarrow C_{k+1}(\mathcal{K}) \xrightarrow{\partial_{k+1}} C_k(\mathcal{K}) \xrightarrow{\partial_k} C_{k-1}(\mathcal{K}) \rightarrow \dots \xrightarrow{\partial_1} C_0(\mathcal{K}) \rightarrow 0.$$

The subspaces $B_k(\mathcal{K}) := \text{Im} \partial_{k+1}$ and $Z_k(\mathcal{K}) := \ker \partial_k$ are called the k th boundary group and the k th cycle group, respectively. By definition, it is straightforward to show that $\partial_k \circ \partial_{k+1} = 0$. Thus $B_k(\mathcal{K})$ becomes a subspace of Z_k . The quotient space

$$H_k(\mathcal{K}) = Z_k(\mathcal{K})/B_k(\mathcal{K})$$

is called the k th homology group of \mathcal{K} , and its rank is the k th Betti number,

$$\beta_k(\mathcal{K}) = \text{rank} H_k(\mathcal{K}).$$

In computational topology, it is convenient to consider the case where $\mathbf{F} = \mathbf{F}_2 = \{0, 1\}$ because in this case we do not need orientations of simplices.

Let A_k be the matrix representation of the boundary operator ∂_k with respect to the standard bases \mathcal{K}_k and \mathcal{K}_{k-1} . It is clear that A_k is a $f_{k-1} \times f_k$ matrix with $\{0, \pm 1\}$ -coefficients, where $f_k = \dim(C_k(\mathcal{K})) = |\mathcal{K}_k|$. Then the k th Betti number can be expressed as

$$\beta_k(\mathcal{K}) = \dim(\ker A_k) - \text{rank} A_{k+1} = f_k - \text{rank} A_k - \text{rank} A_{k+1}. \tag{3}$$

Let $\{\mathcal{K}^{(i)}\}_{i \in I}$ be a finite collection of disjoint simplicial complexes. Then $\cup_{i \in I} \mathcal{K}^{(i)}$ becomes a simplicial complex. We easily see that

$$\beta_k\left(\bigcup_{i \in I} \mathcal{K}^{(i)}\right) = \sum_{i \in I} \beta_k(\mathcal{K}^{(i)}). \tag{4}$$

Next assume that \mathcal{K} is a sub-complex of $\tilde{\mathcal{K}}$, that is, $\mathcal{K} \subset \tilde{\mathcal{K}}$. We use symbols like $\tilde{\beta}_k, \tilde{f}_k$ and \tilde{A}_k to denote corresponding quantities of $\tilde{\mathcal{K}}$. By arranging the basis in $C_k(\tilde{\mathcal{K}})$ such that first elements coincide with the elements in the basis of $C_k(\mathcal{K})$, the matrix \tilde{A}_k has the following form

$$\tilde{A}_k = \left(\begin{array}{c|c} A_k & \\ \hline \mathbf{0} & * \end{array} \right).$$

It then follows that $0 \leq \text{rank} \tilde{A}_k - \text{rank} A_k \leq \tilde{f}_k - f_k$. This inequality, together with the relation (3), implies the following result ([9, Lemma 2.2]).

Lemma 2.1 Let $\mathcal{K}, \tilde{\mathcal{K}}$ be two finite simplicial complexes such that $\mathcal{K} \subset \tilde{\mathcal{K}}$. Then for every $k \geq 1$,

$$\left| \beta_k(\mathcal{K}) - \beta_k(\tilde{\mathcal{K}}) \right| \leq \sum_{j=k}^{k+1} \# \left\{ j\text{-simplices in } \tilde{\mathcal{K}} \setminus \mathcal{K} \right\}. \tag{5}$$

We have mentioned that Betti numbers are nearly additive because of the two properties (4) and (5). Note that β_0 counts the number of connected components in the undirected graph $G = (V, E)$, where $E = \mathcal{K}_1$, which is independent of the underlying field \mathbf{F} .

3 Proofs of main theorems

We will use the following two important properties of Poisson point processes. Denote by $\mathcal{P}(f(x))$ the non-homogeneous Poisson point process with intensity function $f(x)$.

- (i) Scaling property. For any $\theta > 0$ and $t \in \mathbb{R}^d$,

$$\theta(\mathcal{P}(f(x)) - t) \stackrel{d}{=} \mathcal{P}(\theta^{-d}f(t + \theta^{-1}x)),$$

where ‘ $\stackrel{d}{=}$ ’ denotes the equality in distribution. In particular, $\theta(\mathcal{P}(\lambda) - t) \stackrel{d}{=} \mathcal{P}(\theta^{-d}\lambda)$.

- (ii) Coupling property. Let $\mathcal{P}(g(x))$ be a Poisson point process with intensity function $g(x)$ which is independent of $\mathcal{P}(f(x))$. Then

$$\mathcal{P}(f(x)) + \mathcal{P}(g(x)) \stackrel{d}{=} \mathcal{P}(f(x) + g(x)).$$

Here ‘+’ means the superposition of two point processes.

We begin with a result for the simplices counting function.

Lemma 3.1 (cf. [9, Lemma 3.2]) *Let $S_j(\lambda, r; L)$ be the number of j -simplices in $\mathcal{C}(\mathcal{P}_L(\lambda), r)$. Then for fixed $r > 0$,*

$$\frac{\mathbb{E}[S_j(\lambda, r; L)]}{L} \rightarrow \hat{S}_j(\lambda, r) \text{ as } L \rightarrow \infty, \text{ uniformly for } 0 \leq \lambda \leq \Lambda.$$

In addition, for fixed r , the limit $\hat{S}_j(\lambda, r)$ is a continuous function of λ on $[0, \infty)$.

Proof For convenience, let $A_l(\lambda) := S_j(\lambda, r; l^d) = S_j(\mathcal{C}(\mathcal{P}_{V_l}(\lambda), r))$, where $V_l = [-\frac{l}{2}, \frac{l}{2}]^d$. Our aim now is to show that

$$\frac{\mathbb{E}[A_l(\lambda)]}{l^d} \text{ uniformly converges for } \lambda \in [0, \Lambda] \text{ as } l \rightarrow \infty,$$

and that $\mathbb{E}[A_l(\lambda)]$ is continuous for $\lambda \in [0, \infty)$. Let us first show the continuity of $\mathbb{E}[A_l(\lambda)]$. For $0 \leq \lambda < \mu$, we use the coupling $\mathcal{P}(\mu) = \mathcal{P}(\lambda) + \mathcal{P}(\mu - \lambda)$. Here $\mathcal{P}(\lambda)$ and $\mathcal{P}(\mu - \lambda)$ are two independent Poisson point processes with densities λ and $(\mu - \lambda)$, respectively. Let N_λ (resp. $N_{\mu;\lambda}$) be the number of points of $\mathcal{P}(\lambda)$ (resp. $\mathcal{P}(\mu - \lambda)$) in V_l , which has Poisson distribution with parameter λl^d (resp. $(\mu - \lambda)l^d$). Then the continuity follows from a trivial estimate

$$0 \leq A_l(\mu) - A_l(\lambda) \leq N_{\mu;\lambda} (N_{\mu;\lambda} + N_\lambda)^j.$$

Next, we show the uniform convergence. The proof here is similar to that of the pointwise convergence ([9, Lemma 3.2]). Define the function

$$h(\mathcal{P}(\lambda)) := \frac{1}{j+1} \sum_{x \in \mathcal{P}_1(\lambda)} \# [j\text{-simplices in } \mathcal{C}(\mathcal{P}(\lambda), r) \text{ containing } x]. \tag{6}$$

Then for $l > 4r + 1$,

$$\sum_{z \in \mathbb{Z}^d \cap V_{l-4r-1}} h(\mathcal{P}(\lambda) - z) \leq A_l(\lambda) \leq \sum_{z \in \mathbb{Z}^d \cap V_{l+1}} h(\mathcal{P}(\lambda) - z).$$

Consequently, by the stationarity of the Poisson point process $\mathcal{P}(\lambda)$,

$$(l - 4r - 2)^d \mathbb{E}[h(\mathcal{P}(\lambda))] \leq \mathbb{E}[A_l(\lambda)] \leq (l + 2)^d \mathbb{E}[h(\mathcal{P}(\lambda))].$$

Note that $\mathbb{E}[h(\mathcal{P}(\lambda))]$ is non-decreasing in λ and for any $\lambda > 0$,

$$\mathbb{E}[h(\mathcal{P}(\lambda))] \leq \mathbb{E}[\mathcal{P}(\lambda; V_{1+4r})^{j+1}] < \infty.$$

Here $\mathcal{P}(\lambda; V_{1+4r})$ is the number of points of $\mathcal{P}(\lambda)$ in V_{1+4r} . Therefore uniformly for $0 \leq \lambda \leq \Lambda$,

$$\frac{\mathbb{E}[A_l(\lambda)]}{l^d} \rightarrow \mathbb{E}[h(\mathcal{P}(\lambda))] \text{ as } l \rightarrow \infty.$$

The proof is complete. □

For the sake of simplicity, we denote by $\beta_k(\lambda, r; L)$ the k th Betti number of the Čech complex $\mathcal{C}(\mathcal{P}_{W_L}(\lambda), r)$, where W_L is any rectangle of the form $x + [-\frac{L^{1/d}}{2}, \frac{L^{1/d}}{2}]^d$.

Lemma 3.2 *For fixed $r > 0$, uniformly for $0 \leq \lambda \leq \Lambda$,*

$$\frac{\mathbb{E}[\beta_k(\lambda, r; L)]}{L} \rightarrow \hat{\beta}_k(\lambda, r) \text{ as } L \rightarrow \infty.$$

The limit $\hat{\beta}_k(\lambda, r)$ has the following scaling property,

$$\hat{\beta}_k(\lambda, r) = \frac{1}{\theta} \hat{\beta}_k\left(\lambda\theta, \frac{r}{\theta^{1/d}}\right), \text{ for any } \theta > 0.$$

In particular, $\hat{\beta}_k(\lambda, r) = \lambda \hat{\beta}_k(1, \lambda^{1/d}r)$ is a continuous function in both λ and r , and $\hat{\beta}_k(\lambda, r) > 0$, if $\lambda > 0$ and $r > 0$.

Proof For fixed $r > 0$ and fixed $\lambda > 0$, the convergence of the expectations of Betti numbers was shown in [9, Lemma 3.3]. The positivity is a consequence of [8, Theorem 4.2]. Here we will show the uniform convergence for $0 \leq \lambda \leq \Lambda$. We use the following criterion for the uniform convergence on a compact set, which is related to the Arzelà–Ascoli theorem. The sequence of continuous functions $\{a_l(\lambda)\}_{L>0}$ converges uniformly on $[0, \Lambda]$ if and only if it converges pointwise and is equicontinuous, that is, for any $\varepsilon > 0$, there are $\delta > 0$ and $L_0 > 0$ such that

$$|a_L(\lambda_1) - a_L(\lambda_2)| < \varepsilon \text{ for all } \lambda_1, \lambda_2 \in [0, \Lambda],$$

$$|\lambda_1 - \lambda_2| < \delta, \text{ and all } L > L_0. \tag{7}$$

Our task now is to show that the sequence $\{L^{-1}\mathbb{E}[\beta_k(\lambda, r; L)]\}$ is equicontinuous. Let $\lambda < \mu$. By using the coupling $\mathcal{P}(\mu) = \mathcal{P}(\lambda) + \mathcal{P}(\mu - \lambda)$, the Čech complex $\mathcal{C}(\mathcal{P}_L(\lambda), r)$ becomes a sub-complex of $\mathcal{C}(\mathcal{P}_L(\mu), r)$. Thus, by Lemma 2.1,

$$|\beta_k(\mu, r; L) - \beta_k(\lambda, r; L)|$$

$$\leq \sum_{j=k}^{k+1} \# \{j\text{-simplices in } \mathcal{C}(\mathcal{P}_L(\mu), r) \setminus \mathcal{C}(\mathcal{P}_L(\lambda), r)\}$$

$$= \sum_{j=k}^{k+1} (S_j(\mu, r; L) - S_j(\lambda, r; L)). \tag{8}$$

Therefore

$$\left| \frac{\mathbb{E}[\beta_k(\mu, r; L)]}{L} - \frac{\mathbb{E}[\beta_k(\lambda, r; L)]}{L} \right|$$

$$\leq \sum_{j=k}^{k+1} \left(\frac{\mathbb{E}[S_j(\mu, r; L)]}{L} - \frac{\mathbb{E}[S_j(\lambda, r; L)]}{L} \right). \tag{9}$$

The sequence $\{L^{-1}\mathbb{E}[S_j(\lambda, r; L)]\}$ converges uniformly on $[0, \Lambda]$ by Lemma 3.1, and hence, is equicontinuous, which then implies the equicontinuity of the sequence $\{L^{-1}\mathbb{E}[\beta_k(\lambda, r; L)]\}$.

By observing that $\theta^{-1/d}\mathcal{P}(\lambda)$ has the same distribution with $\mathcal{P}(\lambda\theta)$, we obtain the scaling property of $\hat{\beta}_k(\lambda, r)$. It then follows from the scaling property that $\hat{\beta}_k(\lambda, r)$ is continuous in both λ and r . The lemma is proved. \square

Let us now consider the scaled Poissonized version

$$\mathcal{P}_n = \left\{ n^{1/d}X_1, n^{1/d}X_2, \dots, n^{1/d}X_{N_n} \right\}.$$

Recall that N_n is independent of $\{X_n\}$ and has Poisson distribution with parameter n . Then $\mathcal{P}_n = n^{1/d}\bar{\mathcal{P}}_n$ is a non-homogeneous Poisson point process with the intensity function $f_n(x) := f(x/n^{1/d})$. It is clear that $\mathcal{C}(\mathcal{P}_n, n^{1/d}r_n) \cong \mathcal{C}(\bar{\mathcal{P}}_n, r_n)$. Thus Theorem 1.6 can be rewritten as follows.

Theorem 3.3 *Assume that the common probability density function $f(x)$ has compact support, is bounded and Riemann integrable. Then in the regime that $\tilde{r}_n \rightarrow r \in (0, \infty)$,*

$$\frac{\mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}_n, \tilde{r}_n))]}{n} \rightarrow \int_{\mathbb{R}^d} \hat{\beta}_k(f(x), r) dx$$

$$= \int_{\mathbb{R}^d} \hat{\beta}_k(1, f(x)^{1/d}r) f(x) dx \text{ as } n \rightarrow \infty. \tag{10}$$

Remark 3.4. Note that $\mathcal{P}'_n = (r/\tilde{r}_n)\mathcal{P}_n$ is also a non-homogeneous Poisson point process. Moreover, as a result of scaling, $\mathcal{C}(\mathcal{P}_n, \tilde{r}_n) \cong \mathcal{C}(\mathcal{P}'_n, r)$. Thus it is enough to prove Theorem 3.3 with $\tilde{r}_n = r$.

Lemma 3.5 *Assume that $f(x), g(x) \leq \Lambda$ in W_L , where $W_L \subset \mathbb{R}^d$ is any Borel set of volume L . Then there exists a constant $c = c(k, \Lambda L)$ such that*

$$|\mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}_{W_L}(f(x)), r))]| - \mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}_{W_L}(g(x)), r))]|$$

$$\leq c \int_{W_L} |f(x) - g(x)| dx. \tag{11}$$

Proof By considering $f(x) := f(x)|_{W_L}$ and $g(x) := g(x)|_{W_L}$, we omit the subscript W_L in formulae. Let $h(x) = \max\{f(x), g(x)\}$. A key idea here is the following coupling

$$\mathcal{P}(h(x)) = \mathcal{P}(f(x)) + \mathcal{P}(h(x) - f(x)).$$

Let $t = \int (h(x) - f(x)) dx = \int (g(x) - f(x))^+ dx$ and N_t be the number of points of $\mathcal{P}(h(x) - f(x))$ in W_L . Then N_t has Poisson distribution with parameter t . The total number of points of $\mathcal{P}(h(x))$ is bounded by $N_t + N_{\Lambda L - t}$, where $N_{\Lambda L - t}$ has Poisson distribution with parameter $(\Lambda L - t)$ which is independent of N_t . It now follows from Lemma 2.1 that

$$|\beta_k(\mathcal{C}(\mathcal{P}(f(x)), r)) - \beta_k(\mathcal{C}(\mathcal{P}(h(x)), r))|$$

$$\leq \sum_{j=k}^{k+1} S_j(\mathcal{C}(\mathcal{P}(h(x)), r) \setminus \mathcal{C}(\mathcal{P}(f(x)), r))$$

$$\leq 2N_t(N_t + N_{\Lambda L - t})^{k+1},$$

and hence,

$$|\mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}(f(x)), r))]| - \mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}(h(x)), r))]|$$

$$\leq 2\mathbb{E}[N_t(N_t + N_{\Lambda L - t})^{k+1}]. \tag{12}$$

The right hand side is a polynomial of t whose smallest order is 1 and note that $t \leq \Lambda L$, thus it is bounded by ct , where the constant $c = c(k, \Lambda L)$ depends only on k and ΛL , namely we have

$$|\mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}(f(x)), r))]| - \mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}(h(x)), r))]|$$

$$\leq c \int (g(x) - f(x))^+ dx. \tag{13}$$

An analogous estimate holds when we compare the k th Betti number of $\mathcal{C}(\mathcal{P}(g(x)), r)$ and that of $\mathcal{C}(\mathcal{P}(h(x)), r)$. The proof is complete. \square

Proof of Theorem 3.3 Let S be the support of f and $\Lambda := \sup f(x)$. Divide \mathbb{R}^d according to the lattice $(L/n)^{1/d}\mathbb{Z}^d$ and let $\{C_i\}$ be the cubes which intersect with S . Since we also consider the Poisson point process with density 0, we may assume that $S = \cup_i C_i$.

Let W_i be the image of C_i under the map $x \mapsto n^{1/d}x$. Then W_i is a cube of length $L^{1/d}$. Let $\beta_k(W_i, r)$ be the k th Betti number of $\mathcal{C}(\mathcal{P}_n|_{W_i}, r)$. We now compare the k th Betti number of $\mathcal{C}(\mathcal{P}_n, r)$ and that of $\cup_i \mathcal{C}(\mathcal{P}_n|_{W_i}, r)$ by taking into account Lemma 2.1,

$$\begin{aligned} \left| \beta_k(\mathcal{C}(\mathcal{P}_n, r)) - \beta_k\left(\bigcup_i \mathcal{C}(\mathcal{P}_n|_{W_i}, r)\right) \right| &\leq \sum_{j=k}^{k+1} S_j(\mathcal{C}(\mathcal{P}_n, r) \setminus \bigcup_i \mathcal{C}(\mathcal{P}_n|_{W_i}, r)) \\ &\leq \sum_{j=k}^{k+1} S_j(\mathcal{P}_n, r; \cup_i (\partial W_i)^{(2r)}). \end{aligned}$$

Here $S_j(\mathcal{P}_n, r; A)$ is the number of j -simplices in $\mathcal{C}(\mathcal{P}_n, r)$ which has a vertex in A , ∂A denotes the boundary of the set A and $A^{(2r)}$ is the set of points with distance at most $2r$ from A . The second inequality holds because any simplex in $\mathcal{C}(\mathcal{P}_n, r) \setminus \cup_i \mathcal{C}(\mathcal{P}_n|_{W_i}, r)$ must have a vertex in $\cup_i (\partial W_i)^{(2r)}$.

Next, by the coupling $\mathcal{P}(\Lambda) = \mathcal{P}_n + \mathcal{P}(\Lambda - f(x/n^{1/d}))$, it follows that for any bounded Borel set A ,

$$\begin{aligned} \mathbb{E}[S_j(\mathcal{P}_n, r; A)] &\leq \mathbb{E}[S_j(\mathcal{P}(\Lambda), r; A)] \\ &\leq \mathbb{E}\left[\sum_{x \in \mathcal{P}(\Lambda) \cap A} \mathcal{P}(\Lambda; B_{2r}(x))^j\right] =: \mu_{\Lambda, r, j}(A) < \infty. \end{aligned}$$

It turns out that $\mu_{\Lambda, r, j}$ becomes a translation invariant measure on \mathbb{R}^d which is finite on bounded Borel sets. Thus $\mu_{\Lambda, r, j}(A) = c(\Lambda, r, j)|A|$ for some constant $c(\Lambda, r, j)$ depending only on Λ, r and j . Now by taking the expectation in (5), we get

$$\begin{aligned} \left| \mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}_n, r))] - \sum_i \mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}_n|_{W_i}, r))] \right| \\ \leq c \sum_i |(\partial W_i)^{(2r)}| \leq c' \frac{n|S|}{L} L^{(d-1)/d} = c' \frac{n|S|}{L^{1/d}}, \end{aligned}$$

where c and c' are constants which do not depend on n and L . Therefore,

$$\limsup_{n \rightarrow \infty} \left| \frac{\mathbb{E}[\beta_k(\mathcal{C}(\mathcal{P}_n, r))]}{n} - \frac{1}{n} \sum_i \mathbb{E}[\beta_k(W_i, r)] \right| \leq c' \frac{|S|}{L^{1/d}}. \tag{14}$$

Let $f_i^* := \sup_{x \in C_i} f(x)$ and $\beta_k(f_i^*, r; L)$ be the k th Betti number of the Čech complex built on a homogeneous Poisson point process $\mathcal{P}_{W_i}(f_i^*)$ with density f_i^* restricted on W_i . Then by Lemma 3.5,

$$\begin{aligned} |\mathbb{E}[\beta_k(W_i, r)] - \mathbb{E}[\beta_k(f_i^*, r; L)]| &\leq c(k, \Lambda L) \int_{W_i} (f_i^* - f(x/n^{1/d})) dx \\ &= c(k, \Lambda L)n \int_{C_i} (f_i^* - f(x)) dx. \end{aligned}$$

Here $c(k, \Lambda L)$ is the constant in Lemma 3.5. Consequently,

$$\begin{aligned} \left| \frac{1}{n} \sum_i \mathbb{E}[\beta_k(W_i, r)] - \frac{1}{n} \sum_i \mathbb{E}[\beta_k(f_i^*, r; L)] \right| \\ \leq c(k, \Lambda L) \sum_i \int_{C_i} (f_i^* - f(x)) dx \rightarrow 0 \text{ as } n \rightarrow \infty, \end{aligned}$$

because the function $f(x)$ is assumed to be Riemann integrable.

By comparing $\mathbb{E}[\beta_k(f_i^*, r; L)]$ with the limit $\hat{\beta}_k(\lambda, r)$, we get

$$\begin{aligned} \left| \frac{1}{n} \sum_i \mathbb{E}[\beta_k(f_i^*, r; L)] - \frac{L}{n} \sum_i \hat{\beta}_k(f_i^*, r) \right| \\ \leq \frac{L}{n} \#\{C_i\} \sup_{0 \leq \lambda \leq \Lambda} \left| \frac{\mathbb{E}[\beta_k(\lambda, r; L)]}{L} - \hat{\beta}_k(\lambda, r) \right| \\ = |S| \sup_{0 \leq \lambda \leq \Lambda} \left| \frac{\mathbb{E}[\beta_k(\lambda, r; L)]}{L} - \hat{\beta}_k(\lambda, r) \right|. \end{aligned}$$

Note that for fixed L , as $n \rightarrow \infty$,

$$\sum_i \hat{\beta}_k(f_i^*, r) \frac{L}{n} \rightarrow \int_S \hat{\beta}_k(f(x), r) dx.$$

Therefore

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_i \mathbb{E}[\beta_k(W_i, r)] - \int_S \hat{\beta}_k(f(x), r) dx \right| \\ \leq |S| \sup_{0 \leq \lambda \leq \Lambda} \left| \frac{\mathbb{E}[\beta_k(\lambda, r; L)]}{L} - \hat{\beta}_k(\lambda, r) \right|. \tag{15} \end{aligned}$$

Combining the two estimates (14) and (15) and then let $L \rightarrow \infty$, we get the desired result. The proof is complete. \square

The result for binomial point processes will follow from Theorem 1.6 and the following result.

Lemma 3.6 As $n \rightarrow \infty$ with $n^{1/d}r_n \rightarrow r$,

$$\left| \frac{\mathbb{E}[\beta_k(\mathcal{C}(\bar{\mathcal{P}}_n, r_n))]}{n} - \frac{\mathbb{E}[\beta_k(\mathcal{C}(\mathfrak{X}_n, r_n))]}{n} \right| \rightarrow 0.$$

Proof By Lemma 2.1 again, we have,

$$\begin{aligned} |\beta_k(\mathcal{C}(\bar{\mathcal{P}}_n, r_n)) - \beta_k(\mathcal{C}(\mathfrak{X}_n, r_n))| \\ \leq \sum_{j=k}^{k+1} |S_j(\mathcal{C}(\bar{\mathcal{P}}_n, r_n)) - S_j(\mathcal{C}(\mathfrak{X}_n, r_n))|. \end{aligned}$$

The right hand side, divided by n , converges to 0 as a consequence of general results in [6, 7] applied to S_j . Here we will give an easy proof.

For any m , let

$$S_j(m, n) = |S_j(\mathcal{C}(\mathfrak{X}_m, r_n)) - S_j(\mathcal{C}(\mathfrak{X}_n, r_n))|.$$

Since the probability density function $f(x)$ is bounded, in the regime that $n^{1/d}r_n \rightarrow r$, the probability that $\{X_1 \in B_x(r_n)\}$ is bounded by

$$\mathbb{P}(X_1 \in B_x(r_n)) \leq \frac{c}{n},$$

for some constant c which does not depend on n .

For $m > n \geq j$, since each j -simplex in $\mathcal{C}(\mathcal{X}_m, r_n) \setminus \mathcal{C}(\mathcal{X}_n, r_n)$ must contain at least one vertex in $\{X_{n+1}, \dots, X_m\}$, we have

$$\begin{aligned} \mathbb{E}[S_j(m, n)] &\leq (m - n)\mathbb{E}[\#\{j\text{-simplices in } \mathcal{C}(\mathcal{X}_m, r_n) \text{ containing } X_m\}] \\ &\leq (m - n) \binom{m}{j} \mathbb{P}(X_1 \in B_{X_m}(r_n), \dots, X_j \in B_{X_m}(r_n)) \\ &\leq (m - n) \frac{m!}{j!(m-j)!} \left(\frac{c}{n}\right)^j \\ &\leq c_1(m - n) \left(\frac{m}{n}\right)^j. \end{aligned}$$

When $j \leq m < n$, we change the role of m and n to get

$$\mathbb{E}[S_j(m, n)] \leq (n - m) \binom{n}{j} \left(\frac{c}{n}\right)^j \leq c_2(n - m).$$

Combining two estimates, we have

$$\mathbb{E}[S_j(m, n)] \leq c_3|m - n| \left[1 + \left(\frac{m}{n}\right)^j\right].$$

Therefore,

$$\begin{aligned} &\mathbb{E}[|S_j(\mathcal{C}(\bar{\mathcal{P}}_n, r_n)) - S_j(\mathcal{C}(\mathcal{X}_n, r_n))|] \\ &\leq c_3\mathbb{E}\left[|N_n - n| \left(1 + \frac{(N_n)^j}{n^j}\right)\right] \\ &\leq c_3\mathbb{E}[(N_n - n)^2]^{1/2} \mathbb{E}\left[\left(1 + \frac{(N_n)^j}{n^j}\right)^2\right]^{1/2}. \end{aligned}$$

Here in the last inequality, we have used the Cauchy–Schwarz inequality. Note that $\mathbb{E}[(N_n)^j]$ is a polynomial in n of degree j . Thus the second factor in the above estimate remains bounded as $n \rightarrow \infty$. Note also that

$$\mathbb{E}[(N_n - n)^2] = \text{Var}[N_n] = n.$$

Therefore,

$$\frac{\mathbb{E}[|S_j(\mathcal{C}(\bar{\mathcal{P}}_n, r_n)) - S_j(\mathcal{C}(\mathcal{X}_n, r_n))|]}{n} \leq \frac{c_4}{n^{1/2}} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

which completes the proof of Lemma 3.6. □

Acknowledgements

The author would like to thank Professor Tomoyuki Shirai and Dr. Kenkichi Tsunoda for useful discussions. The author would also like to thank the referees for their valuable comments. This work is partially supported by JST CREST Mathematics (15656429). The author is partially supported by JSPS KAKENHI Grant No. 16K17616.

Competing interests

The author declares that he has no competing interests.

Received: 14 November 2016 Revised: 12 January 2017 Accepted: 27 February 2017

Published online: 06 March 2017

References

1. Bobrowski, O, Kahle, M: Topology of random geometric complexes: a survey. *Topology in Statistical Inference, the Proceedings of Symposia in Applied Mathematics*. to appear
2. Meester, R, Roy, R: *Continuum Percolation*. In: *Cambridge Tracts in Mathematics*, p. 238. Cambridge University Press, Cambridge, (1996)
3. Munkres, JR: *Elements of Algebraic Topology*. Addison-Wesley Publishing Company, Menlo Park (1984)
4. Nakamura, T, Hiraoka, Y, Hirata, A, Escobar, EG, Nishiura, Y: Persistent homology and many-body atomic structure for medium-range order in the glass. *Nanotechnology*. **26**(30), 304001 (2015)
5. Penrose, M: *Random Geometric Graphs*. In: *Oxford Studies in Probability*, p. 330. Oxford University Press, Oxford, (2003)
6. Penrose, MD: Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli*. **13**(4), 1124–1150 (2007)
7. Penrose, MD, Yukich, JE: Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.* **13**(1), 277–303 (2003)
8. Yogeshwaran, D, Adler, RJ: On the topology of random complexes built over stationary point processes. *Ann. Appl. Probab.* **25**(6), 3338–3380 (2015)
9. Yogeshwaran, D, Subag, E, Adler, RJ: Random geometric complexes in the thermodynamic regime. *Probab. Theory Relat. Fields*. **167**(1–2), 107–142 (2017)

Submit your manuscript to a SpringerOpen journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com