

LIMIT THEOREMS FOR PROCESS-LEVEL BETTI NUMBERS FOR SPARSE AND CRITICAL REGIMES

TAKASHI OWADA,* *Purdue University*

ANDREW M. THOMAS,** *Purdue University*

Abstract

The objective of this study is to examine the asymptotic behavior of Betti numbers of Čech complexes treated as stochastic processes and formed from random points in the d -dimensional Euclidean space \mathbb{R}^d . We consider the case where the points of the Čech complex are generated by a Poisson process with intensity nf for a probability density f . We look at the cases where the behavior of the connectivity radius of Čech complex causes simplices of dimension greater than $k+1$ to vanish in probability, the so-called sparse regime, as well when the connectivity radius is on the order of $n^{-1/d}$, the critical regime. We establish limit theorems in the aforementioned regimes: central limit theorems for the sparse and critical regimes, and a Poisson limit theorem for the sparse regime. When the connectivity radius of the Čech complex is $o(n^{-1/d})$, i.e., the sparse regime, we can decompose the limiting processes into a time-changed Brownian motion or a time-changed homogeneous Poisson process respectively. In the critical regime, the limiting process is a centered Gaussian process but has much more complicated representation, because the Čech complex becomes highly connected with many topological holes of any dimension.

Keywords: Random topology; Betti number; Central limit theorem; Poisson limit theorem.

2010 Mathematics Subject Classification: Primary 60D05.

Secondary 55U10; 60F05; 05E45.

* Postal address: Department of Statistics, Purdue University, IN, 47907, USA

** Postal address: Department of Statistics, Purdue University, IN, 47907, USA

1. Introduction

The problem of analyzing data in the presence of noise has always been a challenge. With the advent of the application of algebraic topology to probabilistic structures, the tools to capture the most prominent of features of a space have never been as numerous and versatile. These techniques and their corresponding theory typically fall under the umbrella of topological data analysis (TDA). This paper takes up the mantle of trying to capture the dynamic evolution of these features, by investigating stochastic processes of topological summaries, specifically Betti numbers.

A brief introduction to the concepts of algebraic topology is needed before moving onward. Though our introduction here will be somewhat informal, it will nonetheless provide an intuition for some of the concepts discussed in this study. Those wishing for an introduction to algebraic topology for statistical ends should see [10, 32]. Treatments from a topological perspective for practitioners of all sorts can be seen in [16], and a rigorous treatment can be seen in [18]. In many of the studies on TDA, especially those specific to random topology, the *Betti number* has been a main focus as a good quantifier of topological complexity beyond simple connectivity. Given a topological space X and an integer $k \geq 0$, the k th homology group $H_k(X)$ is the quotient group $\ker \partial_k / \text{im } \partial_{k+1}$, where $\partial_k, \partial_{k+1}$ are boundary maps for X . More intuitively, $H_k(X)$ represents a class of topological invariants representing k -dimensional “cycles” or “holes” as the boundary of a $(k+1)$ -dimensional body. The k th Betti number of X , denoted by $\beta_k(X)$, is defined as the rank of $H_k(X)$. Thus $\beta_k(X)$ captures, in essence, the number of k -dimensional cycles in X (in the following we write “ k -cycle” for short). Having dispatched with this formalism, it is useful to know that $\beta_0(X)$ represents the number of connected components of X , $\beta_1(X)$ the number of “closed loops” in X and $\beta_2(X)$ the number of “voids”. For a manifold embedded in \mathbb{R}^d these are features in one, two and three-dimensional subspaces respectively. Though it is the case that $\beta_k(X)$ is defined for all integers $k \geq 0$, in Figure 1 above $\beta_k(X) = 0$ for $k \geq 3$.

In recent years, there has been a growing interest in the theory of random topology [20, 5, 2, 21, 22, 34], exploring the probabilistic features of Betti numbers as well as related notions, for example, the number of critical points of a certain distance function with a fixed Morse index. In an additional study, [7] studied the maximal (persistent)

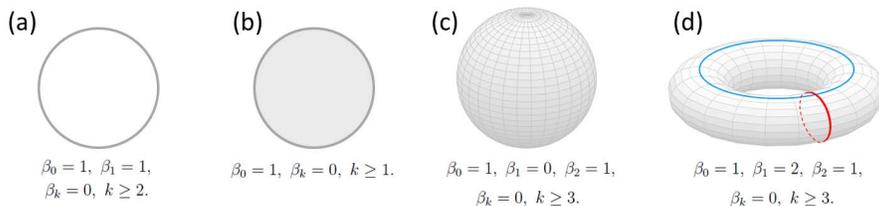


FIGURE 1: The object in (a) is a 1-sphere, or a circle, i.e., $S^1 = \{x \in \mathbb{R}^2 : |x| = 1\}$. The surface in (c) is a 2-sphere or S^2 . Finally, (d) is a two-dimensional torus. Denoting the space corresponding to the torus as X , the blue and red cycles represent the generators of $H_1(X)$, and are regarded as non-equivalent cycles. Note that the torus is hollow, thus $\beta_2(X) = 1$.

k -cycles when an underlying distribution is a uniform Poisson process in the unit cube. Further, [11, 9] investigated topology of a Poisson process on a d -dimensional torus. Those wishing to examine the properties of Betti numbers formed from points generated by a general stationary point process should consult [33, 34]. An elegant summary on recent progress in the field is provided by [6]. The topological objects in these studies are typically constructed from a *geometric complex*. Among many choices of geometric complexes (see, e.g., [16]), the present paper deals with one of the most studied ones, a *Čech complex*; see Figure 2.

Definition 1.1. If $t > 0$ and \mathcal{X} is a collection of points in \mathbb{R}^d , the Čech complex $\check{C}(\mathcal{X}, t)$ is defined as follows:

1. The 0-simplices are the points in \mathcal{X} .
2. A k -simplex $[x_{i_0}, \dots, x_{i_k}]$ is in $\check{C}(\mathcal{X}, t)$ if $\bigcap_{j=0}^k B(x_{i_j}; t/2) \neq \emptyset$,

where $B(x; r) = \{y \in \mathbb{R}^d : |x - y| < r\}$ is an open ball of radius r around $x \in \mathbb{R}^d$.

One good reason for concentrating on the Čech complex is its topological equivalence to the union of balls $\bigcup_{y \in \mathcal{X}} B(y; t/2)$. A fundamental result known as the Nerve lemma (see, e.g., Theorem 10.7 of [4]), asserts that the Čech complex and the union of balls are homotopy equivalent. In particular, they induce the same homology groups, that is for all $k \geq 0$,

$$H_k(\check{C}(\mathcal{X}, t)) \cong H_k\left(\bigcup_{y \in \mathcal{X}} B(y; t/2)\right).$$

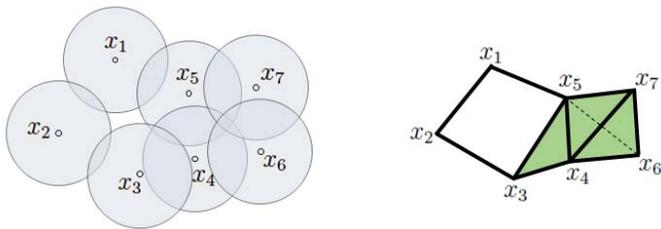


FIGURE 2: Čech complex $\check{C}(\mathcal{X}, t)$ with $\mathcal{X} = \{x_1, \dots, x_7\} \subset \mathbb{R}^2$. There are eleven 1-simplices with each adding a line segment joining a pair of the points. The 2-simplex $[x_3, x_4, x_5]$ belongs to $\check{C}(\mathcal{X}, t)$, since the balls around these points have a non-empty intersection. The 3-simplex $[x_4, x_5, x_6, x_7]$ represents a tetrahedron.

The objective of the current paper is to investigate how the k th Betti number fluctuates as the sample size increases under the setup of [21, 5, 8]. This setup dates back to the classical study on random geometric graphs as seen in the monograph [26]. This is due to the fact that a Čech complex can be seen as a higher-dimensional analogue of a geometric graph. In fact, a geometric graph is a 1-skeleton of a Čech complex. Let \mathcal{X}_n be a set of random points on \mathbb{R}^d . Typically \mathcal{X}_n represents n i.i.d random points sampled from a probability density f or a set of points taken from a Poisson process with intensity nf . Further, let r_n denote a sequence of connectivity radii of a Čech complex. In this setting the behavior of $\check{C}(\mathcal{X}_n, r_n)$ is classified into several different regimes, depending on how nr_n^d varies as $n \rightarrow \infty$. There is an intuitive meaning behind the quantity nr_n^d . It is actually the average number of points in a ball of radius r_n around a point $x \in \mathbb{R}^d$, up to a proportionality constant.

The first regime is that if $nr_n^d \rightarrow 0$ as $n \rightarrow \infty$, the complex is so sparse that separate connected components are scattered throughout the space. This is called the *sparse regime*. If the connectivity radii r_n decays to 0 more slowly, i.e., $nr_n^d \rightarrow \xi \in (0, \infty)$, then $\check{C}(\mathcal{X}_n, r_n)$ belongs to the *critical regime*, in which the complex begins to be connected, forming much larger components with topological holes of various dimensions. Finally the case when $nr_n^d \rightarrow \infty$ is the *dense regime*, for which the complex is highly connected with few topological holes. Detailed study of the Betti numbers has yielded univariate central limit theorems for the sparse regime [21, 22] and for the critical regime [34, 31]. The strong law of large numbers in the critical regime has

been established in [34, 17, 30]. In addition [21] has proven a Poisson convergence result of Betti numbers when $n^{k+2}r_n^{d(k+1)} \rightarrow \lambda \in (0, \infty)$ as $n \rightarrow \infty$, so that topological holes hardly ever occur, in the sense that a Poisson limit theorem ensures that the expectation of k th Betti number converges to a positive constant as $n \rightarrow \infty$, whereas the expectation diverges if $n^{k+2}r_n^{d(k+1)} \rightarrow \infty$ as $n \rightarrow \infty$. In other words, the occurrence of topological holes is rare.

The main objective of this study is to generalize Betti numbers as a stochastic process and provide comprehensive results on the central limit theorem and Poisson limit theorem. We shall consider the Betti number of a Čech complex with radius $r_n(t) := s_n t$, namely

$$\beta_{k,n}(t) := \beta_k(\check{C}(\mathcal{X}_n, r_n(t))), \quad t > 0. \quad (1.1)$$

Obviously (1.1) gives a stochastic process in parameter t with right continuous sample paths with left limits. With this functional setup, this paper reveals that when the Čech complex is relatively sparse, the limiting process of (properly normalized) $\beta_{k,n}(t)$ can be decomposed into the difference of well-known stochastic processes. Specifically, if $ns_n^d \rightarrow 0$ as in Section 3, we can decompose the limiting process into the difference of time-changed Brownian motions and if $n^{k+2}s_n^{d(k+1)} \rightarrow 1$ as in Section 5, we can decompose the limiting process as the difference of time-changed homogenous Poisson processes on the real half-line. In the critical regime of Section 4 however, the limiting process of the normalized $\beta_{k,n}(t)$ has much more complicated representation due to the emergence of connected components of larger size. In fact, the limiting process is denoted as the sum of infinitely many Gaussian processes with each representing connected components of size $i \geq k + 2$ with j topological holes. We would like to emphasize that various “non-functional” type limit theorems with t in (1.1) fixed, have been proven so far – as seen in the last paragraph. However, much less is known about the process-level Betti numbers in (1.1), and their corresponding “functional” limit theorems.

The motivation of reformulating Betti numbers as a stochastic process partially comes from an application to *persistent homology*. Persistent homology is perhaps the most prominent and ubiquitous tool in TDA. Those needing a quick introduction should consult [1]. For surveys of applications of persistent homology see [15, 10,

32]. The first [10] gives a self-contained theoretical treatment of the topological and probabilistic aspects as well as detailed applications. The second [15] is an essential and succinct overview. The final one [32] gives an introduction to persistent homology and its applications from a statistical perspective. Theoretically rigorous treatment of persistent homology, especially the computational aspects, can be seen in [14, 35]. In addition one can even prove vague and weak convergence for persistence diagrams based off of geometric complexes formed from stationary point processes [19].

Considering a family $(\check{C}(\mathcal{X}_n, r_n(t)), t > 0)$ of Čech complexes and increasing radii t , the k th persistent homology provides a list of pairs (birth, death), representing the birth time (radius) at which a k -cycle is born and the death time (radius) at which it gets filled in and disappears. One of the typical applications of our results is the analysis on the sum of persistence barcodes as seen in [24], that is, the sum of life lengths of all k -cycles up to time (radius) t , given by

$$L_{k,n}(t) = \int_0^t \beta_{k,n}(s) \, ds, \quad t > 0. \quad (1.2)$$

This is useful insofar as it allows us to capture the average length of the persistence barcode for the k th persistent homology of our random point cloud, which is potentially important in various statistical applications. Of course, the limiting process of (1.2) is impossible to obtain from non-functional Betti numbers that do not involve parameter t . According to our results, however, it can be obtained as an integral of the limiting process of $\beta_{k,n}(t)$. Similar treatments of the stochastic process approach to geometric/topological functionals include [24, 25].

Regarding proof techniques, we shall borrow ideas from [26, 21, 22] and apply sharper variance/covariance bounds than those given in [22]. Using these sharper bounds, the central limit theorem proven for the sparse regime no longer requires $s_n = o(n^{-1/d-\delta})$ for some $\delta > 0$ in the case that $n^{k+3}s_n^{d(k+2)}$ is bounded away from zero, as is assumed in [22]. Furthermore, we did not use techniques as seen in the [34, 17, 30, 31] for the critical regime because of the fact that this fundamentally alters the representation of the limiting process we obtain, and such a representation is integral to our contribution. We will highlight these differences later on in Section 4. The argument for the Poisson limit theorem uses a completely different technique based on [12].

As a final remark, unlike [26, 21, 22] we do not consider points generated by a binomial process. Further studies would have to perform “De-Poissonization” as seen in section 2.5 of [26]. We have skipped these results not only for brevity but because they are highly technical and add little to the intuition behind our results.

The structure of the paper is as follows. The second section details our setup and all the notation needed to appropriately and succinctly elucidate our results. The third section details the central limit theorem for the sparse regime, i.e. when we have $ns_n^d \rightarrow 0$ and $n^{k+2}s_n^{d(k+1)} \rightarrow \infty$. The fourth section is about the critical regime, in which $ns_n^d = 1$, and Section 5 is dedicated to investigating the Poisson limit theorem with $n^{k+2}s_n^{d(k+1)} = 1$. The major part of Section 6 is devoted to proving the central limit theorem for the critical regime and the Poisson limit theorem for the sparse regime. The proof for the central limit theorem in the sparse regime can be obtained immediately via simple modification of the critical regime case, but is nonetheless given a brief treatment in Section 6.2.

2. Setup

To begin, we start by defining some essential concepts towards proving the results in this paper. We look at point clouds generated by \mathcal{P}_n , a Poisson process on \mathbb{R}^d , $d \geq 2$. We take \mathcal{P}_n to have the intensity measure $n \int_A f(x) dx$ for all measurable A in \mathbb{R}^d , where f is a probability density that is almost everywhere bounded and continuous with respect to Lebesgue measure. Throughout the paper, Lebesgue measure on $\mathbb{R}^{d(k+1)}$ is denoted by m_k and for convenience we let $m := m_0$.

As an aside, we have a few definitions to mention before we begin. First, let $\|f\|_\infty$ be the essential supremum of the aforementioned f , which is finite as f is almost everywhere bounded. Furthermore, define $\theta_d := m(B(0;1))$ to be the volume of the unit ball in \mathbb{R}^d . The constant $C_{f,k}$ is mentioned frequently in the study and is defined as

$$C_{f,k} := \frac{1}{(k+2)!} \int_{\mathbb{R}^d} f(x)^{k+2} dx.$$

Furthermore we let $\mathbb{R}_+ := [0, \infty)$ and \mathbb{N} be the positive integers and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ —the non-negative integers, and $\mathbf{1}\{\cdot\}$ denotes an indicator function.

It is useful to define the notion of a finite point cloud throughout the study. We

let $\mathcal{X}_m := \{X_1, X_2, \dots, X_m\}$ where X_i are i.i.d with density f as mentioned before, though it may represent an arbitrary subset of \mathbb{R}^d of cardinality m as needed. Thus if N_n is a Poisson random variable with parameter n , which is independent of X_i 's, then we can represent the Poisson process \mathcal{P}_n as

$$\mathcal{P}_n(A) = \sum_{i=1}^{N_n} \delta_{X_i}(A),$$

for all measurable $A \subset \mathbb{R}^d$, with δ_x a Dirac measure at $x \in \mathbb{R}^d$.

With this definition in tow, we turn towards the study of Betti numbers. Fixing $1 \leq k < d$, we define $h_t(x_1, \dots, x_{k+2})$, $x_i \in \mathbb{R}^d$, to be the indicator that $\check{C}(\{x_1, x_2, \dots, x_{k+2}\}, t)$ contains an *empty* $(k+1)$ -simplex. This means that $\check{C}(\{x_1, x_2, \dots, x_{k+2}\}, t)$ does not contain a $(k+1)$ -simplex but does contain all possible k -simplices.

With this in mind, we see that h_t can be represented as

$$h_t(x_1, \dots, x_{k+2}) = h_t^+(x_1, \dots, x_{k+2}) - h_t^-(x_1, \dots, x_{k+2}),$$

where we define

$$h_t^+(x_1, \dots, x_{k+2}) := \prod_{i=1}^{k+2} \mathbf{1} \left\{ \bigcap_{j=1, j \neq i}^{k+2} B(x_j; t/2) \neq \emptyset \right\},$$

$$h_t^-(x_1, \dots, x_{k+2}) := \mathbf{1} \left\{ \bigcap_{j=1}^{k+2} B(x_j; t/2) \neq \emptyset \right\}.$$

It is important to note that h_t^\pm is non-decreasing in t . That is,

$$h_s^\pm(x_1, \dots, x_{k+2}) \leq h_t^\pm(x_1, \dots, x_{k+2})$$

for all $0 \leq s < t$ and $x_i \in \mathbb{R}^d$.

Throughout the paper we interest ourselves in the k th Betti number for $\check{C}(\mathcal{P}_n, r_n(t))$ where $r_n(t) := s_n t$. Recall that the nature of how s_n decays to 0 as $n \rightarrow \infty$ is the object of our study. We denote by $S_{k,n}(t)$ the number of empty $(k+1)$ -simplex components of $\check{C}(\mathcal{P}_n, r_n(t))$. In other words, $S_{k,n}(t)$ represents the number of connected components C on $k+2$ points such that $\beta_k(C) = 1$. More generally, for integers $i \geq k+2$ and $j > 0$, we define $U_{i,j,n}(t)$ as the number of connected components C of $\check{C}(\mathcal{P}_n, r_n(t))$ such that $|C| = i$ and $\beta_k(C) = j$. Then the k th Betti number of $\check{C}(\mathcal{P}_n, r_n(t))$ can be represented as

$$\beta_{k,n}(t) = \sum_{i \geq k+2} \sum_{j > 0} j U_{i,j,n}(t), \quad t > 0. \quad (2.1)$$

Since $S_{k,n}(t) = U_{k+2,1,n}(t)$ and one cannot form multiple empty $(k+1)$ -simplices from $k+2$ points, (2.1) can also be represented as

$$\beta_{k,n}(t) = S_{k,n}(t) + \sum_{i>k+2} \sum_{j>0} jU_{i,j,n}(t), \quad t > 0. \quad (2.2)$$

In this setting it is instructive to introduce the following indicator functions to formalize these concepts for an arbitrary collection of points $\mathcal{Y} \subset \mathcal{X} \subset \mathbb{R}^d$:

- $J_{i,t}(\mathcal{Y}, \mathcal{X}) := \mathbf{1}\{\check{C}(\mathcal{Y}, t) \text{ is a connected component of } \check{C}(\mathcal{X}, t)\} \mathbf{1}\{|\mathcal{Y}| = i\}$.
- $b_{j,t}(\mathcal{Y}) := \mathbf{1}\{\beta_k(\check{C}(\mathcal{Y}, t)) = j\}$.
- $g_t^{(i,j)}(\mathcal{Y}, \mathcal{X}) := b_{j,t}(\mathcal{Y})J_{i,t}(\mathcal{Y}, \mathcal{X})$.

In particular, denote

$$g_t(\mathcal{Y}, \mathcal{X}) := g_t^{(k+2,1)}(\mathcal{Y}, \mathcal{X}) = b_{1,t}(\mathcal{Y})J_{k+2,t}(\mathcal{Y}, \mathcal{X}) = h_t(\mathcal{Y})J_{k+2,t}(\mathcal{Y}, \mathcal{X}).$$

Additionally, for $A \subset \mathbb{R}^d$, let

- $h_{t,A}(\mathcal{Y}) := h_t(\mathcal{Y})\mathbf{1}\{\text{LMP}(\mathcal{Y}) \in A\}$,
- $g_{t,A}^{(i,j)}(\mathcal{Y}, \mathcal{X}) := g_t^{(i,j)}(\mathcal{Y}, \mathcal{X})\mathbf{1}\{\text{LMP}(\mathcal{Y}) \in A\}$,

where $\text{LMP}(\mathcal{Y})$ is the left-most point, in dictionary order, of the set \mathcal{Y} .

With the above indicators now available, it is clear that $S_{k,n}(t) = \sum_{\mathcal{Y} \subset \mathcal{P}_n} g_{r_n(t)}(\mathcal{Y}, \mathcal{P}_n)$ and $U_{i,j,n}(t) = \sum_{\mathcal{Y} \subset \mathcal{P}_n} g_{r_n(t)}^{(i,j)}(\mathcal{Y}, \mathcal{P}_n)$. As a final bit of notation, let

$$\beta_{k,n,A}(t) = \sum_{i \geq k+2} \sum_{j > 0} jU_{i,j,n,A}(t) = S_{k,n,A}(t) + \sum_{i > k+2} \sum_{j > 0} jU_{i,j,n,A}(t),$$

where, in all functions above, we require the left-most point of every subset \mathcal{Y} to be an element of A . When brevity is paramount, we occasionally shorten $\sum_{i > k+2} \sum_{j > 0} jU_{i,j,n}(t)$ to $R_{k,n}(t)$ and $\sum_{i > k+2} \sum_{j > 0} jU_{i,j,n,A}(t)$ to $R_{k,n,A}(t)$ respectively.

3. Central limit theorem in sparse regime

Throughout this section we assume that $ns_n^d \rightarrow 0$ and $\rho_n := n^{k+2}s_n^{d(k+1)} \rightarrow \infty$ as $n \rightarrow \infty$. The essence is that Čech complexes are distributed sparsely with many separate connected components, because of a fast decay of s_n as a result of $ns_n^d \rightarrow 0$.

Consequently, all k -cycles are asymptotically supported on $k + 2$ points. From a more analytic viewpoint, the behavior of the k th Betti number (2.2) is completely determined by $S_{k,n}(t)$, whereas $R_{k,n}(t) = \beta_{k,n}(t) - S_{k,n}(t)$ is asymptotically negligible in the sense that $\mathbb{E}[S_{k,n}(t)]/\rho_n$ tends to a finite positive constant, but $\mathbb{E}[R_{k,n}(t)]/\rho_n \rightarrow 0$ as $n \rightarrow \infty$.

The most relevant study to this section is [21], in which the central limit theorem for the sparse regime is discussed. We have extended [21] (with the erratum paper [22]) in two directions. First, we develop the process-level central limit theorem. This highlights the chief contribution of this paper. Whereas [21, 22], as well as [34] in the ensuing section, treat the “static” topology of random Čech complexes (i.e., no time parameter t involved), the main focus of this paper is “dynamic” topology of the same complex, treating Betti numbers as a stochastic process. Second, our central limit theorem is for the entirety of the sparse regime, without requiring that $s_n = o(n^{-1/d-\delta})$ for some $\delta > 0$ as assumed in [22].

Before presenting the main result we define the limiting stochastic process

$$\mathcal{G}_k(t) := \int_{\mathbb{R}^{d(k+1)}} h_t(0, \mathbf{y}) G_k(d\mathbf{y}), \quad (3.1)$$

where G_k is a Gaussian random measure such that $G_k(A) \sim \mathcal{N}(0, C_{f,k} m_k(A))$ for all measurable A in $\mathbb{R}^{d(k+1)}$. Furthermore, for A_1, \dots, A_m disjoint, $G_k(A_1), \dots, G_k(A_m)$ are independent. As defined, $\mathcal{G}_k(t)$ depends on the indicator h_t , meaning that due to sparsity of the Čech complex in this regime, the k -cycles affecting $\mathcal{G}_k(t)$ must be always formed by connected components on $k + 2$ points (i.e., components of the smallest size).

The significance of the characterization of the process at (3.1) is that if we define

$$\mathcal{G}_k^\pm(t) := \int_{\mathbb{R}^{d(k+1)}} h_t^\pm(0, \mathbf{y}) G_k(d\mathbf{y}),$$

then $\mathcal{G}_k^\pm(t)$ becomes a time-changed Brownian motion; see Proposition 3.1 below. Hence $\mathcal{G}_k(t) = \mathcal{G}_k^+(t) - \mathcal{G}_k^-(t)$ is a difference of two dependent time-changed Brownian motions, where dependence is due to the same Gaussian random measure G_k shared by $\mathcal{G}_k^+(t)$ and $\mathcal{G}_k^-(t)$. Those wishing to examine this characterization in more detail should refer to [25]. For example, it is proven in [25] that the process $\mathcal{G}_k(t)$ is self-similar with exponent $H = d(k - 1)/2$ and is Hölder continuous of any order in $[0, 1/2)$.

Proposition 3.1. *The process $\mathcal{G}_k^\pm(t)$ can be expressed as*

$$(\mathcal{G}_k^\pm(t), t \geq 0) \stackrel{d}{=} \left(B(C_{f,k} m_k(D_1^\pm) t^{d(k+1)}), t \geq 0 \right),$$

where B is a standard Brownian motion and $D_t^\pm = \{\mathbf{y} \in \mathbb{R}^{d(k+1)} : h_t^\pm(0, \mathbf{y}) = 1\}$.

Proof. We prove only the result for \mathcal{G}_k^+ , as the proof for \mathcal{G}_k^- is the same. It is elementary to show that $\mathcal{G}_k^+(t)$ has mean zero. Thus, it only remains to demonstrate the covariance result. Since h_t^+ is non-decreasing in t , we have $D_{t_1}^+ \subset D_{t_2}^+$ for $0 \leq t_1 \leq t_2$; therefore,

$$\begin{aligned} \mathbb{E}[\mathcal{G}_k^+(t_1)\mathcal{G}_k^+(t_2)] &= \mathbb{E}[G_k(D_{t_1}^+)G_k(D_{t_2}^+)] = \mathbb{E}[G_k(D_{t_1}^+)^2] \\ &= C_{f,k}m_k(D_{t_1}^+) = C_{f,k}m_k(D_1^+)t_1^{d(k+1)}. \end{aligned}$$

□

Our main result can be seen below. The proof is briefly presented in Section 6.2 as a straightforward variant of the proof for the critical regime. For the proof we need to examine the asymptotic growth rate of expectations and covariances of $\beta_{k,n}(t)$. The detailed results are presented in Proposition 6.2, where it is seen that the expectation and covariance both grow at the rate ρ_n .

Theorem 3.1. *Suppose that $ns_n^d \rightarrow 0$ and $\rho_n = n^{k+2}s_n^{d(k+1)} \rightarrow \infty$. Assume that f is an almost everywhere bounded and continuous density function. Then, we have the following weak convergence in the finite dimensional sense, namely*

$$\rho_n^{-1/2}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]) \xrightarrow{fidi} \mathcal{G}_k(t),$$

meaning that for every $m \in \mathbb{N}$ and $0 < t_1 < t_2 < \dots < t_m < \infty$ we have

$$\rho_n^{-1/2}(\beta_{k,n}(t_i) - \mathbb{E}[\beta_{k,n}(t_i)], i = 1, \dots, m) \Rightarrow (\mathcal{G}_k(t_i), i = 1, \dots, m)$$

weakly in \mathbb{R}^m .

4. Central limit theorem in critical regime

We now expand on the results of [34] (see also [31]) by offering an explicit limit of appropriately scaled moments and a central limit theorem for $\beta_{k,n}(t)$. In the critical regime, the connectivity radius s_n is defined to be $s_n = n^{-1/d}$. This sequence decays more slowly than that in the previous section; hence, Čech complexes become highly connected with many topological holes of any dimension $k < d$. More analytically, all

terms in the sum (2.1) contribute to the k th Betti number, unlike in the sparse regime. This implies that the k -cycles in the limit could be supported not only on $k + 2$ points but also on i points for all possible $i > k + 2$.

In [34], the authors established the central limit theorem for the first time for the critical regime (though they referred to it as the “thermodynamic” regime). There are two key differences between that paper and ours. The first is that the Poisson process they consider is homogeneous with unit intensity, restricted to a set B_n such that $m(B_n) = n$. The second difference between the two, and equivalent to the contrast indicated in the sparse regime, is again that [34] treats the static topology of random Čech complexes whereas we treat the dynamic topology. As a consequence, while the weak limit in [34] is a simple Gaussian distribution with unknown variance, our limit is a Gaussian process having structure similar to that of the Betti number (2.1).

The other relevant article to our study is [31], which generalizes [34] to an inhomogeneous Poisson process case, but again only deals with static topology. We would like to emphasize that our proof techniques are significantly different from those in [34, 31]. In fact, our proof is highly analytic in nature, borrowing machinery from [26] and [21], whereas the proofs of [34, 31] rely more on the topological nature of the objects, including weakly/strongly stabilizing properties of Betti numbers, the notion of critical radius of percolation, and the theory of geometric functionals as in [27]; see also Remark 4.2. By virtue of our analytic approach, we can fully specify the structure of the limiting Gaussian process as in (4.1) below. This is actually the main objective of this study.

We now define the aforementioned limiting Gaussian process by

$$\mathcal{H}_k(t) = \sum_{i \geq k+2} \sum_{j > 0} j \mathcal{H}_k^{(i,j)}(t), \quad t > 0, \quad (4.1)$$

where $(\mathcal{H}_k^{(i,j)}, i \geq k + 2, j > 0)$ is a family of centered Gaussian processes with inter-process dependence between $\mathcal{H}_k^{(i_1, j_1)}$ and $\mathcal{H}_k^{(i_2, j_2)}$ determined by

$$\text{Cov}(\mathcal{H}_k^{(i_1, j_1)}(t_1), \mathcal{H}_k^{(i_2, j_2)}(t_2)) = \frac{1}{i_1!} \eta_{k, \mathbb{R}^d}^{(i_1, j_1, j_2)}(t_1, t_2) \delta_{i_1, i_2} + \frac{1}{i_1! i_2!} \nu_{k, \mathbb{R}^d}^{(i_1, i_2, j_1, j_2)}(t_1, t_2). \quad (4.2)$$

Here δ_{i_1, i_2} is the Kronecker delta, and the functions $\eta_{k, \mathbb{R}^d}^{(i_1, j_1, j_2)}$, $\nu_{k, \mathbb{R}^d}^{(i_1, i_2, j_1, j_2)}$ are explicitly defined during the proof of the main theorem (see (6.2) and (6.3)). From (4.2), the

covariance of $\mathcal{H}_k^{(i,j)}$ is given by

$$\text{Cov}(\mathcal{H}_k^{(i,j)}(t_1), \mathcal{H}_k^{(i,j)}(t_2)) = \frac{1}{i!} \eta_{k, \mathbb{R}^d}^{(i,j,j)}(t_1, t_2) + \frac{1}{(i!)^2} \nu_{k, \mathbb{R}^d}^{(i,i,j,j)}(t_1, t_2).$$

The main point here is that the Betti number (2.1) and the limit (4.1) are represented in a very similar fashion. In fact, the process $U_{i,j,n}(t)$ in (2.1) and $\mathcal{H}_k^{(i,j)}(t)$ in (4.1) both capture the spatial distribution of connected components C with $|C| = i$ and $\beta_k(C) = j$. In particular, $\mathcal{H}_k^{(k+2,1)}(t)$ represents the distribution of components C on $k+2$ points with $\beta_k(C) = 1$ (i.e., components of the smallest size) as does $\mathcal{G}_k(t)$ in the sparse regime. In the present regime however, many of the Gaussian processes in (4.1) beyond $\mathcal{H}_k^{(k+2,1)}(t)$, do contribute to the limit.

As a bit of a technical remark, note that for every $i \geq k+2$, there exists $j_0 > 0$ such that $b_{j,t}(\mathbf{x}) = 0$ for all $j \geq j_0$, $t > 0$, and $\mathbf{x} \in \mathbb{R}^{di}$. In this case,

$$\eta_{k, \mathbb{R}^d}^{(i,j,j)}(t, t) = \nu_{k, \mathbb{R}^d}^{(i,i,j,j)}(t, t) = 0,$$

and thus $\mathcal{H}_k^{(i,j)}$ becomes an identically zero process. For example, $\mathcal{H}_k^{(k+2,j)} \equiv 0$ for all $j \geq 2$, since one cannot create multiple k -cycles from $k+2$ points.

All proofs are collected in Section 6.1. In particular we will see that the growth rate of the expectation and variance of $\beta_{k,n}(t)$ is of order n —see Proposition 6.1. This indicates that the scaling constant for the central limit theorem must be of order $n^{1/2}$.

Theorem 4.1. *Suppose that $ns_n^d = 1$ and f is an almost everywhere bounded and continuous density function. If $0 < t_1 < t_2 < \dots < t_m < (e\|f\|_\infty \theta_d)^{-1/d}$, and $\mathcal{H}_k(t)$ is the centered Gaussian process defined above, then we have the following weak convergence in the finite dimensional sense, namely*

$$n^{-1/2}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]) \xrightarrow{fidi} \mathcal{H}_k(t).$$

This means that for every $m \in \mathbb{N}$ we have

$$n^{-1/2}(\beta_{k,n}(t_i) - \mathbb{E}[\beta_{k,n}(t_i)], i = 1, \dots, m) \Rightarrow (\mathcal{H}_k(t_i), i = 1, \dots, m),$$

weakly in \mathbb{R}^m .

Remark 4.1. Here we assume $ns_n^d = 1$, but we could generalize it to $ns_n^d \rightarrow 1$, $n \rightarrow \infty$. Indeed, throughout the proof of Proposition 6.1 we will frequently encounter

the integral expressions multiplied by $(ns_n^d)^{i-1}$ (e.g., (6.6)). If one assumes $ns_n^d \rightarrow 1$, the integral and $(ns_n^d)^{i-1}$ both converge. Thus, without loss of generality we may set $ns_n^d = 1$, so that we do not have to maintain $(ns_n^d)^{i-1}$ outside of the integral.

Remark 4.2. Although Theorem 4.1 imposes a restriction on the range of t_i 's, the central limit theorem does hold for every $t > 0$ in the case of the “truncated” Betti number

$$\beta_{k,n}^{(M)}(t) = \sum_{i=k+2}^M \sum_{j>0} j U_{i,j,n}(t), \quad M \in \mathbb{N},$$

which itself is useful for the approximation arguments in our proof.

The need for the restriction on t_i 's seems to be a delicate issue. First, according to [34], if the Poisson process is homogeneous, the non-functional central limit theorem holds for any fixed $t > 0$. However in the case of an inhomogeneous Poisson process, [31] has put a restriction on the value of t , despite significant difference in proof techniques with this paper. More specifically, in the notation of Theorem 4.1, the result of [31] indicates that t_m must be less than $r_c \|f\|_\infty^{-1/d}$, where r_c is the critical radius for percolation in a Boolean model in \mathbb{R}^d ([23]).

The restriction on t_i 's in Theorem 4.1 looks a little artificial; indeed it has only been used for demonstrating that

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var}(\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t)) = 0.$$

Nevertheless, removing this restriction is not easy, because of the “global” nature of Betti numbers. In fact, the Betti number consists of (infinitely many) component counts as in (2.1), and their moment calculations always involve an entire point set \mathcal{P}_n (see the definition of $J_{i,t}(\mathcal{Y}, \mathcal{X})$ after (2.2)). This property of Betti numbers seems to make it highly difficult to estimate the probability of large components, unless the radius (i.e., the value of t) is restricted in some way or other. For example, under our restriction $\mathbb{E}[\mathcal{P}_n(B(x; r_n(t)))]$ is bounded by e^{-1} for all $x \in \mathbb{R}^d$, while it can be as large as possible if t is unrestricted. Furthermore, as mentioned above, if the Betti number is truncated we no longer need to deal with large components, hence we do not need a restriction on the range of t .

Before concluding this section we shall exploit Theorem 4.6 in [34] and present the strong law of large numbers of $\beta_{k,n}(t)$. Though the strong law of large numbers has

already been proven in [17, 30] for any fixed $t > 0$, we shall state the result to highlight the novelty of our representation of the limit as the sum of contributions to the Betti number for variously sized components. The proof is short and given at the end of Section 6.1.

Corollary 4.1. *Under the condition of Theorem 4.1, we assume moreover that f has a compact, convex support such that $\inf_{x \in \text{supp}(f)} f(x) > 0$. Then we have, as $n \rightarrow \infty$,*

$$\frac{\beta_{k,n}(t)}{n} \rightarrow \sum_{i=k+2}^{\infty} \sum_{j>0} \frac{j}{i!} \eta_{k, \mathbb{R}^d}^{(i,j,j)}(t, t), \quad a.s.$$

5. Poisson limit theorem in sparse regime

Before concluding this paper we shall explore the random topology of Čech complexes when the complex is even more sparse than that in Section 3, so that k -cycles hardly ever occur. Then, the k th Betti number no longer follows a central limit theorem. Nevertheless, it does obey a Poisson limit theorem. In terms of the connectivity radii, we assume $\rho_n = n^{k+2} s_n^{d(k+1)} = 1$, equivalently, $s_n = n^{-(k+2)/d(k+1)}$, so that s_n converges to 0 more rapidly than in Section 3.

For the definition of a ‘‘Poissonian’’ type limiting process, we let M_k be a Poisson random measure with mean measure $C_{f,k} m_k$. Namely it is defined by

$$M_k(A) \sim \text{Poi}(C_{f,k} m_k(A))$$

for all measurable A in $\mathbb{R}^{d(k+1)}$. Further, if A_1, \dots, A_m are disjoint, $M_k(A_1), \dots, M_k(A_m)$ are independent. We are now ready to define the stochastic process

$$\mathcal{V}_k(t) = \int_{\mathbb{R}^{d(k+1)}} h_t(0, \mathbf{y}) M_k(d\mathbf{y}),$$

which appears below as a weak limit in the main theorem. What is interesting about this is that if we define

$$\mathcal{V}_k^\pm(t) := \int_{\mathbb{R}^{d(k+1)}} h_t^\pm(0, \mathbf{y}) M_k(d\mathbf{y}),$$

then $\mathcal{V}_k(t) = \mathcal{V}_k^+(t) - \mathcal{V}_k^-(t)$ is the difference of two dependent (time-changed) Poisson processes on \mathbb{R}_+ . Interestingly, this treatment is analogous to the statement of the Gaussian process limit in Section 3, and those wishing a deeper exploration of this in

a similar setting should refer to [24]. What is precisely meant by this can be seen in the following proposition.

Proposition 5.1. *The process \mathcal{V}_k^\pm can be expressed as*

$$(\mathcal{V}_k^\pm(t), t \geq 0) \stackrel{d}{=} \left(N_k^\pm(t^{d(k+1)}), t \geq 0 \right),$$

where N_k^\pm is a (homogeneous) Poisson process with intensity $C_{f,k}m_k(D_1^\pm)$ with $D_t^\pm = \{\mathbf{y} \in \mathbb{R}^{d(k+1)} : h_t^\pm(0, \mathbf{y}) = 1\}$.

Proof. As with Proposition 3.1, we prove only the result for \mathcal{V}_k^+ , as the proof for \mathcal{V}_k^- is the same. We can see that if $0 = t_0 < t_1 < \dots < t_k < \infty$ and $\lambda_i > 0$, $i = 1, \dots, k$, then by the non-decreasingness of h_t^+ ,

$$\mathbb{E} \left[\exp \left(- \sum_{i=1}^k \lambda_i (\mathcal{V}_k^+(t_i) - \mathcal{V}_k^+(t_{i-1})) \right) \right] = \mathbb{E} \left[\exp \left(- \sum_{i=1}^k \lambda_i M_k(D_{t_i}^+ \setminus D_{t_{i-1}}^+) \right) \right],$$

where $D_{t_i}^+ \setminus D_{t_{i-1}}^+$ are disjoint and $M_k(D_{t_i}^+ \setminus D_{t_{i-1}}^+)$, $i = 1, \dots, k$, are independent. Moreover, $M_k(D_{t_i}^+ \setminus D_{t_{i-1}}^+)$ is Poisson distributed with parameter

$$C_{f,k}m_k(D_{t_i}^+ \setminus D_{t_{i-1}}^+) = C_{f,k}m_k(D_1^+)(t_i^{d(k+1)} - t_{i-1}^{d(k+1)})$$

by a change of variable. Hence we have that

$$\mathbb{E} \left[\exp \left(- \sum_{i=1}^k \lambda_i M_k(D_{t_i}^+ \setminus D_{t_{i-1}}^+) \right) \right] = \prod_{i=1}^k \exp \left(C_{f,k}m_k(D_1^+)(t_i^{d(k+1)} - t_{i-1}^{d(k+1)})(e^{-\lambda_i} - 1) \right),$$

which implies that the process $\mathcal{V}_k^+(t^{1/d(k+1)})$ has independent increments and

$$\mathcal{V}_k^+((t+s)^{1/d(k+1)}) - \mathcal{V}_k^+(s^{1/d(k+1)})$$

is Poisson with parameter $C_{f,k}m_k(D_1^+)t$. \square

In what follows we assume $\rho_n = 1$, though we could easily modify this to suppose that $\rho_n \rightarrow 1$ as $n \rightarrow \infty$. For simplicity in our proofs we assert the former. The proof is again given in Section 6 and the main techniques there are those in [12].

Theorem 5.1. *Suppose that $\rho_n = 1$ and f is an almost everywhere bounded and continuous density function. Then, we have the following weak convergence in the finite dimensional sense, namely*

$$\beta_{k,n}(t) \xrightarrow{fidi} \mathcal{V}_k(t),$$

meaning that for every $m \in \mathbb{N}$ and $0 < t_1 < t_2 < \dots < t_m < \infty$ we have

$$(\beta_{k,n}(t_i), i = 1, \dots, m) \Rightarrow (\mathcal{V}_k(t_i), i = 1, \dots, m), \quad (5.1)$$

weakly in \mathbb{R}^m .

6. Proofs

In this section we prove the theorems seen in the sections above, with the exposition focused on Sections 4 and 5. We only briefly discuss the proof of Section 3, since the proof is considerably similar to (or even easier than) the critical regime case.

In the sequel, we write $x + \mathbf{y} = (x + y_1, \dots, x + y_m)$ for $x \in \mathbb{R}^d$ and $\mathbf{y} = (y_1, \dots, y_m) \in \mathbb{R}^{dm}$.

6.1. Central limit theorem in critical regime

The first step towards the required central limit theorem is to examine the asymptotic moments as follows. Before proceeding with the proof, let us define the “truncated” Betti numbers

$$\beta_{k,n,A}^{(M)}(t) := \sum_{i=k+2}^M \sum_{j>0} j U_{i,j,n,A}(t), \quad M \in \mathbb{N} \cup \{\infty\} \quad (6.1)$$

for any measurable $A \subset \mathbb{R}^d$. Clearly $\beta_{k,n,A}(t) = \beta_{k,n,A}^{(\infty)}(t)$.

Let us introduce a few items useful for specifying the limiting covariances. In the following i, i_1, i_2, j_1 , and j_2 are positive integers, t_1, t_2 are non-negative reals, A is an open subset of \mathbb{R}^d with $m(\partial A) = 0$, and $a \wedge b := \min\{a, b\}$ with $a \vee b := \max\{a, b\}$. Additionally, we define the two functions

$$\begin{aligned} \eta_{k,A}^{(i,j_1,j_2)}(t_1, t_2) &:= \int_{\mathbb{R}^{d(i-1)}} \int_{\mathbb{R}^d} \mathbf{1}\{\check{C}(\{0, \mathbf{y}\}, t_1 \wedge t_2) \text{ is connected}\} \prod_{\ell=1}^2 b_{j_\ell, t_\ell}(0, \mathbf{y}) \\ &\quad \times \exp\left(- (t_1 \vee t_2)^d f(x) m(\mathcal{B}(\{0, \mathbf{y}\}; 1))\right) f(x)^i \mathbf{1}_A(x) \, dx \, d\mathbf{y}, \end{aligned} \quad (6.2)$$

and

$$\nu_{k,A}^{(i_1,i_2,j_1,j_2)}(t_1, t_2) := \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^{d(i_1-1)}} d\mathbf{y}_1 \int_{\mathbb{R}^{d i_2}} d\mathbf{y}_2 \mathbf{1}\{\check{C}(\{0, \mathbf{y}_1\}, t_1) \text{ is connected}\} \quad (6.3)$$

$$\begin{aligned}
& \times \mathbf{1}\{\check{C}(\mathbf{y}_2, t_2) \text{ is connected}\} b_{j_1, t_1}(0, \mathbf{y}_1) b_{j_2, t_2}(\mathbf{y}_2) \\
& \times \left[\left(\alpha_{t_1, t_2}(\{0, \mathbf{y}_1\}, \mathbf{y}_2) - \alpha_{(t_1 \vee t_2)/2}(\{0, \mathbf{y}_1\}, \mathbf{y}_2) \right) e^{-f(x)m(\mathcal{B}(\{0, \mathbf{y}_1\}; t_1) \cup \mathcal{B}(\mathbf{y}_2; t_2))} \right. \\
& \left. - \alpha_{t_1, t_2}(\{0, \mathbf{y}_1\}, \mathbf{y}_2) e^{-f(x)\{m(\mathcal{B}(\{0, \mathbf{y}_1\}; t_1)) + m(\mathcal{B}(\mathbf{y}_2; t_2))\}} \right] f(x)^{i_1 + i_2} \mathbf{1}_A(x),
\end{aligned}$$

where

$$\mathcal{B}(\mathcal{X}; r) := \bigcup_{y \in \mathcal{X}} B(y; r) \quad (6.4)$$

for a collection \mathcal{X} of \mathbb{R}^d -valued vectors and $r > 0$. Moreover,

$$\alpha_{r, s}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}) := \mathbf{1}\{\mathcal{B}(\mathcal{X}_{i_1}; r) \cap \mathcal{B}(\mathcal{X}_{i_2}; s) \neq \emptyset\},$$

and $\alpha_r(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}) := \alpha_{r, r}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})$. Finally we define for $M \in \mathbb{N} \cup \{\infty\}$,

$$\Phi_{k, A}^{(M)}(t_1, t_2) := \sum_{i_1=k+2}^M \sum_{i_2=k+2}^M \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \left(\frac{\eta_{k, \mathbb{R}^d}^{(i_1, j_1, j_2)}(t_1, t_2) \delta_{i_1, i_2}}{i_1!} + \frac{\nu_{k, \mathbb{R}^d}^{(i_1, i_2, j_1, j_2)}(t_1, t_2)}{i_1! i_2!} \right),$$

where δ_{i_1, i_2} is again the Kronecker delta and we define $\Phi_{k, A}(t_1, t_2) := \Phi_{k, A}^{(\infty)}(t_1, t_2)$.

Proposition 6.1. *Let f be an almost everywhere bounded and continuous density function. Let $ns_n^d = 1$ and $A \subset \mathbb{R}^d$ is open with $m(\partial A) = 0$.*

(i) *If $M < \infty$, then for $t, t_1, t_2 > 0$,*

$$n^{-1} \mathbb{E}[\beta_{k, n, A}^{(M)}(t)] \rightarrow \sum_{i=k+2}^M \sum_{j>0} \frac{j}{i!} \eta_{k, A}^{(i, j, j)}(t, t), \quad n \rightarrow \infty,$$

$$n^{-1} \text{Cov}(\beta_{k, n, A}^{(M)}(t_1), \beta_{k, n, A}^{(M)}(t_2)) \rightarrow \Phi_{k, A}^{(M)}(t_1, t_2), \quad n \rightarrow \infty.$$

(ii) *If $M = \infty$, then for $0 < t, t_1, t_2 < (e\|f\|_\infty \theta_d)^{-1/d}$,*

$$n^{-1} \mathbb{E}[\beta_{k, n, A}(t)] \rightarrow \sum_{i=k+2}^{\infty} \sum_{j>0} \frac{j}{i!} \eta_{k, A}^{(i, j, j)}(t, t), \quad n \rightarrow \infty,$$

$$n^{-1} \text{Cov}(\beta_{k, n, A}(t_1), \beta_{k, n, A}(t_2)) \rightarrow \Phi_{k, A}(t_1, t_2), \quad n \rightarrow \infty,$$

so that the limits above are finite non-zero constants.

Proof. We only establish the statements in (ii). We aim to demonstrate the convergence of the expectation in Part 1 and then in Part 2, the convergence of the covariance to $\Phi_{k, A}(t_1, t_2)$. For ease of description we treat only the case when $A = \mathbb{R}^d$. The argument for a general A will be the same except obvious minor changes.

Part 1: The definition in (2.1), the Palm theory for Poisson processes in [26], and the monotone convergence theorem supply that

$$n^{-1}\mathbb{E}[\beta_{k,n}(t)] = \sum_{i=k+2}^{\infty} \sum_{j>0} j \frac{n^{i-1}}{i!} \mathbb{E}[g_{r_n(t)}^{(i,j)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)], \quad (6.5)$$

where $\mathcal{X}_i = (X_1, \dots, X_i) \in \mathbb{R}^{di}$ is a collection of i.i.d random points in \mathbb{R}^d with common density f . By conditioning on \mathcal{X}_i we have that

$$\begin{aligned} & n^{i-1}\mathbb{E}[g_{r_n(t)}^{(i,j)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)] \quad (6.6) \\ &= n^{i-1}\mathbb{E}\left[b_{j,r_n(t)}(\mathcal{X}_i)\mathbb{E}[J_{i,r_n(t)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) \mid \mathcal{X}_i]\right] \\ &= n^{i-1} \int_{\mathbb{R}^{di}} \mathbf{1}\{\check{C}(\mathbf{x}, r_n(t)) \text{ is connected}\} b_{j,r_n(t)}(\mathbf{x}) \exp(-nI_{r_n(t)}(\mathbf{x})) \prod_{j=1}^i f(x_j) \, d\mathbf{x}, \end{aligned}$$

where

$$I_{r_n(t)}(\mathbf{x}) = I_{r_n(t)}(x_1, \dots, x_i) = \int_{\mathcal{B}(\mathbf{x}; r_n(t))} f(z) \, dz.$$

Subsequently we perform the change of variables $x_1 = x$ and $x_j = x + s_n y_{j-1}$ for $j = 2, \dots, i$, to get that (6.6) is equal to

$$\begin{aligned} & (ns_n^d)^{i-1} \int_{\mathbb{R}^{d(i-1)}} \int_{\mathbb{R}^d} \mathbf{1}\{\check{C}(\{x, x + s_n \mathbf{y}\}, r_n(t)) \text{ is connected}\} b_{j,r_n(t)}(x, x + s_n \mathbf{y}) \\ & \quad \times \exp(-nI_{r_n(t)}(x, x + s_n \mathbf{y})) f(x) \prod_{j=1}^{i-1} f(x + s_n y_j) \, dx \, d\mathbf{y} \\ &= \int_{\mathbb{R}^{d(i-1)}} \int_{\mathbb{R}^d} \mathbf{1}\{\check{C}(\{0, \mathbf{y}\}, t) \text{ is connected}\} b_{j,t}(0, \mathbf{y}) \\ & \quad \times \exp(-nI_{r_n(t)}(x, x + s_n \mathbf{y})) f(x) \prod_{j=1}^{i-1} f(x + s_n y_j) \, dx \, d\mathbf{y}, \end{aligned}$$

where the equality follows from the location and scale invariance of both of the indicator functions. By the continuity of f we have that $\prod_{j=1}^{i-1} f(x + s_n y_j) \rightarrow f(x)^{i-1}$ a.e. as $n \rightarrow \infty$. As for the convergence of the exponential term, we have

$$nI_{r_n(t)}(x, x + s_n \mathbf{y}) = n \int_{\mathcal{B}(\{x, x + s_n \mathbf{y}\}; r_n(t))} f(z) \, dz,$$

which after the change of variable $z = x + s_n v$, gives us

$$n \int_{\mathcal{B}(\{x, x + s_n \mathbf{y}\}; r_n(t))} f(z) \, dz \rightarrow t^d f(x) m(\mathcal{B}(\{0, \mathbf{y}\}; 1)).$$

It then follows from the dominated convergence theorem that

$$n^{i-1} \mathbb{E}[g_{r_n(t)}^{(i,j)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)] \rightarrow \eta_{k, \mathbb{R}^d}^{(i,j)}(t, t), \quad n \rightarrow \infty.$$

It remains to find a summable upper bound for (6.5) to apply the dominated convergence theorem for sums. To this end we use the inequality $j \leq \binom{i}{k+1}$ which is the result of the fact that there must be a k -simplex in $\check{C}(\mathcal{X}_i, r_n(t))$ whenever $\beta_k(\check{C}(\mathcal{X}_i, r_n(t))) > 0$. In addition, using an obvious inequality

$$J_{i, r_n(t)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) \leq \mathbf{1}\{\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}\}, \quad (6.7)$$

we get that

$$\begin{aligned} n^{-1} \mathbb{E}[\beta_{k,n}(t)] &\leq \sum_{i=k+2}^{\infty} \binom{i}{k+1} \frac{n^{i-1}}{i!} \sum_{j=1}^{\binom{i}{k+1}} \mathbb{E}[\mathbf{1}\{\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}\} b_{j, r_n(t)}(\mathcal{X}_i)] \\ &\leq \sum_{i=k+2}^{\infty} \binom{i}{k+1} \frac{n^{i-1}}{i!} \mathbb{P}(\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}). \end{aligned} \quad (6.8)$$

For further analysis we claim that

$$\mathbb{P}(\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}) \leq i^{i-2} (r_n(t)^d \|f\|_{\infty} \theta_d)^{i-1}. \quad (6.9)$$

Indeed this can be derived from

$$\begin{aligned} &\mathbb{P}(\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}) \\ &= \int_{\mathbb{R}^{di}} \mathbf{1}\{\check{C}(\mathbf{x}, r_n(t)) \text{ is connected}\} \prod_{j=1}^i f(x_j) \, d\mathbf{x} \\ &= r_n(t)^{d(i-1)} \int_{\mathbb{R}^{di}} \mathbf{1}\{\check{C}(\{0, \mathbf{y}\}, 1) \text{ is connected}\} f(x) \prod_{j=1}^{i-1} f(x + r_n(t)y_j) \, d\mathbf{x} \, d\mathbf{y} \\ &\leq (r_n(t)^d \|f\|_{\infty})^{i-1} \int_{\mathbb{R}^{d(i-1)}} \mathbf{1}\{\check{C}(\{0, \mathbf{y}\}, 1) \text{ is connected}\} \, d\mathbf{y} \\ &\leq i^{i-2} (r_n(t)^d \|f\|_{\infty} \theta_d)^{i-1}. \end{aligned} \quad (6.10)$$

The last inequality comes from the basic fact that there are i^{i-2} spanning trees on i vertices. Combining (6.8), (6.9), and $ns_n^d = 1$ we conclude that

$$n^{-1} \mathbb{E}[\beta_{k,n}(t)] \leq \frac{1}{(k+1)!} \sum_{i=k+2}^{\infty} \frac{i^{i-2}}{(i-k-1)!} (t^d \|f\|_{\infty} \theta_d)^{i-1} =: \frac{1}{(k+1)!} \sum_{i=k+2}^{\infty} a_i.$$

It is easy to check that $a_{i+1}/a_i \rightarrow et^d \|f\|_\infty \theta_d$ as $i \rightarrow \infty$, where the limit is less than 1 by our assumption. So the ratio test has shown that $\sum_{i=k+2}^\infty a_i$ converges as required.

Part 2: We assume $0 < t_1 \leq t_2 < (e\|f\|_\infty \theta_d)^{-1/d}$ and proceed with the fact that

$$\begin{aligned} & \mathbb{E}[\beta_{k,n}(t_1)\beta_{k,n}(t_2)] \\ &= \sum_{i_1=k+2}^\infty \sum_{i_2=k+2}^\infty \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \mathbb{E} \left[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} g_{r_n(t_1)}^{(i_1, j_1)}(\mathcal{Y}_1, \mathcal{P}_n) g_{r_n(t_2)}^{(i_2, j_2)}(\mathcal{Y}_2, \mathcal{P}_n) \right] \\ &= \sum_{i=k+2}^\infty \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \mathbb{E} \left[\sum_{\mathcal{Y} \subset \mathcal{P}_n} g_{r_n(t_1)}^{(i, j_1)}(\mathcal{Y}, \mathcal{P}_n) g_{r_n(t_2)}^{(i, j_2)}(\mathcal{Y}, \mathcal{P}_n) \right] \\ &+ \sum_{i_1=k+2}^\infty \sum_{i_2=k+2}^\infty \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \\ &\quad \times \mathbb{E} \left[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} g_{r_n(t_1)}^{(i_1, j_1)}(\mathcal{Y}_1, \mathcal{P}_n) g_{r_n(t_2)}^{(i_2, j_2)}(\mathcal{Y}_2, \mathcal{P}_n) \mathbf{1}\{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = 0\} \right]. \end{aligned}$$

The second equality comes from an observation that if $\mathcal{Y}_1 \neq \mathcal{Y}_2$ and the intersection of \mathcal{Y}_1 and \mathcal{Y}_2 is non-empty, then $\check{C}(\mathcal{Y}_2, r_n(t_2))$ cannot be an isolated component of $\check{C}(\mathcal{P}_n, r_n(t_2))$ —so these terms are zero. Appealing to Palm theory again as seen in [21], we get that

$$\begin{aligned} & \mathbb{E}[\beta_{k,n}(t_1)\beta_{k,n}(t_2)] \\ &= \sum_{i=k+2}^\infty \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^i}{i!} \mathbb{E} \left[g_{r_n(t_1)}^{(i, j_1)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i, j_2)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) \right] \\ &+ \sum_{i_1=k+2}^\infty \sum_{i_2=k+2}^\infty \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^{i_1+i_2}}{i_1! i_2!} \\ &\quad \times \mathbb{E} \left[g_{r_n(t_1)}^{(i_1, j_1)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i_2, j_2)}(\mathcal{X}_{i_2}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) \right], \end{aligned}$$

where \mathcal{X}_i and \mathcal{P}_n are independent, and \mathcal{X}_{i_1} , \mathcal{X}_{i_2} , and \mathcal{P}_n are also mutually independent such that \mathcal{X}_{i_1} and \mathcal{X}_{i_2} are disjoint.

Applying (6.5) to each $\mathbb{E}[\beta_{k,n}(t_i)]$, $i = 1, 2$, and utilizing the independence of \mathcal{X}_{i_1} and \mathcal{X}_{i_2} , we see that the covariance function can be written as

$$\text{Cov}(\beta_{k,n}(t_1), \beta_{k,n}(t_2)) = A_{1,n} + A_{2,n}, \quad (6.11)$$

with

$$A_{1,n} := \sum_{i=k+2}^\infty \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^i}{i!} \mathbb{E} \left[g_{r_n(t_1)}^{(i, j_1)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i, j_2)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) \right], \quad (6.12)$$

$$\begin{aligned}
A_{2,n} := & \sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^{i_1+i_2}}{i_1! i_2!} \\
& \times \mathbb{E} \left[g_{r_n(t_1)}^{(i_1, j_1)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i_2, j_2)}(\mathcal{X}_{i_2}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) \right. \\
& \left. - g_{r_n(t_1)}^{(i_1, j_1)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i_2, j_2)}(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}'_n) \right],
\end{aligned} \tag{6.13}$$

where \mathcal{P}'_n is an independent copy of \mathcal{P}_n and is also independent of \mathcal{X}_{i_1} and \mathcal{X}_{i_2} .

Let us denote the expectation portions of $A_{1,n}$ and $A_{2,n}$ as $E_{1,n}^{(\mathbf{i}, \mathbf{j})}$ and $E_{2,n}^{(\mathbf{i}, \mathbf{j})}$, with $\mathbf{i} = (\mathbf{i}_1, \mathbf{i}_2)$, and $\mathbf{j} = (\mathbf{j}_1, \mathbf{j}_2)$ respectively. Our goal is to show that $n^{-1}(A_{1,n} + A_{2,n})$ tends to $\Phi_{k, \mathbb{R}^d}(t_1, t_2)$ as $n \rightarrow \infty$. For now we shall compute the limits of $n^{i-1} E_{1,n}^{(\mathbf{i}, \mathbf{j})}$ and $n^{i_1+i_2-1} E_{2,n}^{(\mathbf{i}, \mathbf{j})}$ for each i, i_1, i_2, j_1 , and j_2 , while temporarily assuming that the dominated convergence theorem for sums is applicable for both $n^{-1} A_{1,n}$ and $n^{-1} A_{2,n}$. By mirroring the argument from Part 1 with the same change of variables and recalling $t_1 \leq t_2$,

$$\begin{aligned}
n^{i-1} E_{1,n}^{(\mathbf{i}, \mathbf{j})} &= n^{i-1} \mathbb{E} \left[\mathbf{1} \{ \check{C}(\mathcal{X}_i, r_n(t_1)) \text{ is connected} \} \right. \\
&\quad \left. \times \prod_{\ell=1}^2 b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_i) \exp(-n I_{r_n(t_2)}(\mathcal{X}_i)) \right] \\
&= \int_{\mathbb{R}^{d(i-1)}} \int_{\mathbb{R}^d} \mathbf{1} \{ \check{C}(\{0, \mathbf{y}\}, t_1) \text{ is connected} \} \prod_{\ell=1}^2 b_{j_\ell, t_\ell}(0, \mathbf{y}) \\
&\quad \times \exp(-n I_{r_n(t_2)}(x, x + s_n \mathbf{y})) f(x) \prod_{j=1}^{i-1} f(x + s_n y_j) dx d\mathbf{y} \\
&\rightarrow \eta_{k, \mathbb{R}^d}^{(i, j_1, j_2)}(t_1, t_2) \text{ as } n \rightarrow \infty.
\end{aligned}$$

Hence the assumed dominated convergence theorem for sums concludes that

$$n^{-1} A_{1,n} \rightarrow \sum_{i=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} \frac{j_1 j_2}{i!} \eta_{k, \mathbb{R}^d}^{(i, j_1, j_2)}(t_1, t_2) \quad n \rightarrow \infty. \tag{6.14}$$

To demonstrate convergence for $n^{i_1+i_2-1} E_{2,n}^{(\mathbf{i}, \mathbf{j})}$, let us shorten $g_{r_n(t_1)}^{(i_1, j_1)}$ to g_1 and $g_{r_n(t_2)}^{(i_2, j_2)}$ to g_2 and decompose $E_{2,n}^{(\mathbf{i}, \mathbf{j})}$ into two terms:

$$\begin{aligned}
E_{2,n}^{(\mathbf{i}, \mathbf{j})} &= \mathbb{E} \left[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) \right. \\
&\quad \left. - g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n) \right] \\
&\quad + \mathbb{E} \left[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) (g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n) - g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}'_n)) \right]
\end{aligned}$$

$$:= B_{1,n} + B_{2,n}.$$

Note that for $\ell = 1, 2$,

$$g_\ell(\mathcal{X}_{i_\ell}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) = g_\ell(\mathcal{X}_{i_\ell}, \mathcal{X}_{i_\ell} \cup \mathcal{P}_n) \mathbf{1}\{\mathcal{B}(\mathcal{X}_{i_1}; r_n(t_\ell)/2) \cap \mathcal{B}(\mathcal{X}_{i_2}; r_n(t_\ell)/2) = \emptyset\},$$

where $\mathcal{B}(\mathcal{X}; r)$ is defined in (6.4). Hence we have that

$$B_{1,n} = -\mathbb{E}\left[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n) \alpha_{r_n(t_2)/2}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})\right].$$

At the same time, the spatial independence of \mathcal{P}_n justifies that

$$\begin{aligned} B_{2,n} &= \mathbb{E}\left[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) (g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n) \right. \\ &\quad \left. - g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}'_n)) \alpha_{r_n(t_1), r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})\right]. \end{aligned}$$

Consequently we can rewrite $E_{2,n}^{(\mathbf{i}, \mathbf{j})}$ as

$$\begin{aligned} E_{2,n}^{(\mathbf{i}, \mathbf{j})} &= \mathbb{E}\left[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n) \right. \\ &\quad \left. \times (\alpha_{r_n(t_1), r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}) - \alpha_{r_n(t_2)/2}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}))\right] \\ &\quad - \mathbb{E}\left[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}'_n) \alpha_{r_n(t_1), r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})\right] \\ &:= C_{1,n} - C_{2,n}. \end{aligned} \tag{6.15}$$

After conditioning on $\mathcal{X}_{i_1} \cup \mathcal{X}_{i_2}$, the customary change of variable yields

$$\begin{aligned} n^{i_1+i_2-1} C_{1,n} &= n^{i_1+i_2-1} \mathbb{E}\left[\prod_{\ell=1}^2 \mathbf{1}\{\check{C}(\mathcal{X}_{i_\ell}, r_n(t_\ell)) \text{ is connected}\} b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_{i_\ell}) \right. \\ &\quad \left. \times (\alpha_{r_n(t_1), r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}) - \alpha_{r_n(t_2)/2}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})) \right. \\ &\quad \left. \times \exp\left(-n \int_{\mathcal{B}(\mathcal{X}_{i_1}; r_n(t_1)) \cup \mathcal{B}(\mathcal{X}_{i_2}; r_n(t_2))} f(z) dz\right)\right] \\ &= \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^{d(i_1-1)}} dy_1 \int_{\mathbb{R}^{d i_2}} dy_2 \mathbf{1}\{\check{C}(\{0, \mathbf{y}_1\}, t_1) \text{ is connected}\} \\ &\quad \times \mathbf{1}\{\check{C}(\mathbf{y}_2, t_2) \text{ is connected}\} b_{j_1, t_1}(0, \mathbf{y}_1) b_{j_2, t_2}(\mathbf{y}_2) \\ &\quad \times \left(\alpha_{t_1, t_2}(\{0, \mathbf{y}_1\}, \mathbf{y}_2) - \alpha_{t_2/2}(\{0, \mathbf{y}_1\}, \mathbf{y}_2)\right) \\ &\quad \times \exp\left(-n \int_{\mathcal{B}(\{x, x+s_n \mathbf{y}_1\}; r_n(t_1)) \cup \mathcal{B}(x+s_n \mathbf{y}_2; r_n(t_2))} f(z) dz\right) \\ &\quad \times f(x) \prod_{j=1}^{i_1-1} f(x + s_n y_{1,j}) \prod_{j=1}^{i_2} f(x + s_n y_{2,j}) \end{aligned}$$

$$\begin{aligned}
& \rightarrow \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^{d(i_1-1)}} dy_1 \int_{\mathbb{R}^{di_2}} dy_2 \mathbf{1}\{\check{C}(\{0, \mathbf{y}_1\}, t_1) \text{ is connected}\} \\
& \quad \times \mathbf{1}\{\check{C}(\mathbf{y}_2, t_2) \text{ is connected}\} b_{j_1, t_1}(0, \mathbf{y}_1) b_{j_2, t_2}(\mathbf{y}_2) \\
& \quad \times \left(\alpha_{t_1, t_2}(\{0, \mathbf{y}_1\}, \mathbf{y}_2) - \alpha_{t_2/2}(\{0, \mathbf{y}_1\}, \mathbf{y}_2) \right) \\
& \quad \times e^{-f(x)m(\mathcal{B}(\{0, \mathbf{y}_1\}; t_1) \cup \mathcal{B}(\mathbf{y}_2; t_2))} f(x)^{i_1+i_2},
\end{aligned}$$

where $\mathbf{y}_1 = (y_{1,1}, \dots, y_{1,i_1-1}) \in \mathbb{R}^{d(i_1-1)}$ and $\mathbf{y}_2 = (y_{2,1}, \dots, y_{2,i_2}) \in \mathbb{R}^{di_2}$.

Similarly one can see that

$$\begin{aligned}
n^{i_1+i_2-1} C_{2,n} & \rightarrow \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^{d(i_1-1)}} dy_1 \int_{\mathbb{R}^{di_2}} dy_2 \mathbf{1}\{\check{C}(\{0, \mathbf{y}_1\}, t_1) \text{ is connected}\} \\
& \quad \times \mathbf{1}\{\check{C}(\mathbf{y}_2, t_2) \text{ is connected}\} b_{j_1, t_1}(0, \mathbf{y}_1) b_{j_2, t_2}(\mathbf{y}_2) \alpha_{t_1, t_2}(\{0, \mathbf{y}_1\}, \mathbf{y}_2) \\
& \quad \times e^{-f(x)\{m(\mathcal{B}(\{0, \mathbf{y}_1\}; t_1)) + m(\mathcal{B}(\mathbf{y}_2; t_2))\}} f(x)^{i_1+i_2}.
\end{aligned}$$

Therefore,

$$n^{i_1+i_2-1} E_{2,n}^{(\mathbf{i}, \mathbf{j})} = n^{i_1+i_2-1} (C_{1,n} - C_{2,n}) \rightarrow \nu_{k, \mathbb{R}^d}^{(i_1, i_2, j_1, j_2)}(t_1, t_2), \quad n \rightarrow \infty.$$

Assuming convergence under summation, we have that

$$n^{-1} A_{2,n} \rightarrow \sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} \frac{j_1 j_2}{i_1! i_2!} \nu_{k, \mathbb{R}^d}^{(i_1, i_2, j_1, j_2)}(t_1, t_2), \quad n \rightarrow \infty. \quad (6.16)$$

From (6.14) and (6.16), it follows that $n^{-1}(A_{1,n} + A_{2,n}) \rightarrow \Phi_{k, \mathbb{R}^d}(t_1, t_2)$ as $n \rightarrow \infty$.

Now we would like to show that both $n^{i-1} E_{1,n}^{(\mathbf{i}, \mathbf{j})}$ and $n^{i_1+i_2-1} |E_{2,n}^{(\mathbf{i}, \mathbf{j})}|$ are bounded by a summable quantity, so that application of the dominated convergence theorem for sums is valid for both $n^{-1} A_{1,n}$ and $n^{-1} A_{2,n}$. Using the bounds (6.7), (6.9), together with $ns_n^d = 1$, we have

$$\begin{aligned}
n^{-1} A_{1,n} & \leq \sum_{i=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^{i-1}}{i!} \mathbb{E} \left[\mathbf{1}\{\check{C}(\mathcal{X}_i, r_n(t_1)) \text{ is connected}\} \prod_{\ell=1}^2 b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_i) \right] \\
& \leq \sum_{i=k+2}^{\infty} \binom{i}{k+1}^2 \frac{n^{i-1}}{i!} \mathbb{P}(\check{C}(\mathcal{X}_i, r_n(t_1)) \text{ is connected}) \\
& \leq \frac{1}{((k+1)!)^2} \sum_{i=k+2}^{\infty} \frac{i! i^{i-2}}{((i-k-1)!)^2} (t_1^d \|f\|_\infty \theta_d)^{i-1}.
\end{aligned} \quad (6.17)$$

The last term is convergent by appealing to the assumption $t_1 < (e\|f\|_\infty \theta_d)^{-1/d}$ and the ratio test for sums.

Subsequently we turn our attention to $n^{-1}A_{2,n}$. Returning to (6.15) and using obvious relations

$$\alpha_{r_n(t_1), r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}) \leq \alpha_{r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}), \quad \alpha_{r_n(t_2)/2}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}) \leq \alpha_{r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}),$$

we get that

$$|C_{1,n} - C_{2,n}| \leq 3\mathbb{E}\left[\prod_{\ell=1}^2 \mathbf{1}\{\check{C}(\mathcal{X}_{i_\ell}, r_n(t_2)) \text{ is connected}\} b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_{i_\ell}) \alpha_{r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})\right]$$

By virtue of this bound we have that

$$\begin{aligned} n^{-1}|A_{2,n}| &\leq 3 \sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^{i_1+i_2-1}}{i_1! i_2!} \\ &\times \mathbb{E}\left[\prod_{\ell=1}^2 \mathbf{1}\{\check{C}(\mathcal{X}_{i_\ell}, r_n(t_2)) \text{ is connected}\} b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_{i_\ell}) \alpha_{r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})\right] \\ &\leq 3 \sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \binom{i_1}{k+1} \binom{i_2}{k+1} \frac{n^{i_1+i_2-1}}{i_1! i_2!} \\ &\times \mathbb{P}\left(\check{C}(\mathcal{X}_{i_\ell}, r_n(t_2)) \text{ is connected for } \ell = 1, 2, \right. \\ &\quad \left. \mathcal{B}(\mathcal{X}_{i_1}; r_n(t_2)) \cap \mathcal{B}(\mathcal{X}_{i_2}; r_n(t_2)) \neq \emptyset\right). \end{aligned} \tag{6.18}$$

We claim here that

$$\begin{aligned} &\mathbb{P}\left(\check{C}(\mathcal{X}_{i_\ell}, r_n(t_2)) \text{ is connected for } \ell = 1, 2, \mathcal{B}(\mathcal{X}_{i_1}; r_n(t_2)) \cap \mathcal{B}(\mathcal{X}_{i_2}; r_n(t_2)) \neq \emptyset\right) \\ &\leq 2^d i_1^{i_1-1} i_2^{i_2-1} (r_n(t_2)^d \|f\|_\infty \theta_d)^{i_1+i_2-1}. \end{aligned} \tag{6.19}$$

To see this, by the change of variables as in (6.10), we have that

$$\begin{aligned} &\mathbb{P}\left(\check{C}(\mathcal{X}_{i_\ell}, r_n(t_2)) \text{ is connected for } \ell = 1, 2, \mathcal{B}(\mathcal{X}_{i_1}; r_n(t_2)) \cap \mathcal{B}(\mathcal{X}_{i_2}; r_n(t_2)) \neq \emptyset\right) \\ &\leq (r_n(t_2)^d \|f\|_\infty)^{i_1+i_2-1} \int_{\mathbb{R}^{d(i_1+i_2-1)}} \mathbf{1}\{\check{C}(\{0, y_1, \dots, y_{i_1-1}\}, 1) \text{ is connected}\} \\ &\quad \times \mathbf{1}\{\check{C}(\{y_{i_1}, \dots, y_{i_1+i_2-1}\}, 1) \text{ is connected}\} \\ &\quad \times \mathbf{1}\{\mathcal{B}(\{0, y_1, \dots, y_{i_1-1}\}; 1) \cap \mathcal{B}(\{y_{i_1}, \dots, y_{i_1+i_2-1}\}; 1) \neq \emptyset\} \, dy. \end{aligned}$$

Note that there are $i_1^{i_1-2}$ spanning trees on the set of points $\{0, y_1, \dots, y_{i_1-1}\}$ with unit connectivity radius, and there are $i_2^{i_2-2}$ spanning trees on $\{y_{i_1}, \dots, y_{i_1+i_2-1}\}$ with

unit connectivity radius as well. In addition there are $i_1 \times i_2$ possible ways of picking one vertex from $\{0, y_1, \dots, y_{i_1-1}\}$ and another from $\{y_{i_1}, \dots, y_{i_1+i_2-1}\}$, and connecting the two chosen vertices with connectivity radius 2. Therefore, the expression above is eventually bounded by

$$\begin{aligned} & (r_n(t_2)^d \|f\|_\infty)^{i_1+i_2-1} i_1^{i_1-2} i_2^{i_2-2} \theta_d^{i_1+i_2-2} (i_1 i_2 2^d \theta_d) \\ & = 2^d i_1^{i_1-1} i_2^{i_2-1} (r_n(t_2)^d \|f\|_\infty \theta_d)^{i_1+i_2-1}. \end{aligned}$$

Now we have

$$n^{-1} |A_{2,n}| \leq \frac{3 \cdot 2^d}{((k+1)!)^2 t_2^d \|f\|_\infty \theta_d} \left\{ \sum_{i=k+2}^{\infty} \frac{i^{i-1}}{(i-k-1)!} (t_2^d \|f\|_\infty \theta_d)^i \right\}^2.$$

The constraint $t_2 < (e \|f\|_\infty \theta_d)^{-1/d}$, together with the ratio test, guarantees that the last term converges. Hence the proof is completed. \square

Proof of Theorem 4.1. We begin by proving the corresponding result for the truncated Betti number in (6.1) for every $M \in \mathbb{N}$, that is,

$$n^{-1/2} (\beta_{k,n}^{(M)}(t_i) - \mathbb{E}[\beta_{k,n}^{(M)}(t_i)], i = 1, \dots, m) \Rightarrow (\mathcal{H}_k^{(M)}(t_i) i = 1, \dots, m),$$

where $\mathcal{H}_k^{(M)}$ is the ‘‘truncated’’ limiting centered Gaussian process given by

$$\mathcal{H}_k^{(M)}(t) = \sum_{i=k+2}^M \sum_{j>0} j \mathcal{H}_k^{(i,j)}(t).$$

We now restrict ourselves to the case in which the corresponding left most points belong to a fixed bounded set A . By the Cramér-Wold device [13, p. 176], we need to demonstrate a univariate central limit theorem for $\sum_{i=1}^m a_i \beta_{k,n,A}^{(M)}(t_i)$, where $a_i \in \mathbb{R}$, $m \geq 1$. The asymptotic variance of $\sum_{i=1}^m a_i \beta_{k,n,A}^{(M)}(t_i)$ scaled by $n^{-1/2}$ can be derived from Proposition 6.1 (i):

$$\begin{aligned} \text{Var} \left(n^{-1/2} \sum_{i=1}^m a_i \beta_{k,n,A}^{(M)}(t_i) \right) &= \sum_{i=1}^m \sum_{j=1}^m a_i a_j n^{-1} \text{Cov}(\beta_{k,n,A}^{(M)}(t_i), \beta_{k,n,A}^{(M)}(t_j)) \\ &\rightarrow \sum_{i=1}^m \sum_{j=1}^m a_i a_j \Phi_{k,A}^{(M)}(t_i, t_j), \quad n \rightarrow \infty. \end{aligned} \tag{6.20}$$

Our proof exploits Stein’s normal approximation method for weakly dependent random variables, as in Theorem 2.4 in [26]. We assume the limit in (6.20) is positive

as otherwise our proof is trivial. Recall that in the statement of Theorem 4.1, we have set $0 < t_1 < t_2 < \dots < t_m$, and let $(Q_{j,n}, j \in \mathbb{N})$ be an enumeration of almost disjoint closed cubes (i.e., their interiors are disjoint) of side length $r_n(t_m)$, such that $\cup_{j \in \mathbb{N}} Q_{j,n} = \mathbb{R}^d$.

$$V_n := \{j \in \mathbb{N} : Q_{j,n} \cap A \neq \emptyset\},$$

and

$$\xi_{j,n} := \sum_{i=1}^m a_i \beta_{k,n,A \cap Q_{j,n}}^{(M)}(t_i),$$

so that $\sum_{i=1}^m a_i \beta_{k,n,A}^{(M)}(t_i) = \sum_{j \in V_n} \xi_{j,n}$. We now turn V_n into the vertex set of a *dependency graph* (see Section 2.1 in [26] for the definition) by declaring that for $j, j' \in V_n$, $j \sim j'$ if and only if

$$\inf\{|x - y| : x \in Q_{j,n}, y \in Q_{j',n}\} \leq 2Mr_n(t_m).$$

It is easy to show that this provides us with the required independence properties, that is, for any vertex set $I_1, I_2 \subset V_n$ with no edges connecting vertices in I_1 and those in I_2 , we have that $(\xi_{j,n}, j \in I_1)$ and $(\xi_{j,n}, j \in I_2)$ are independent. Note moreover that the degree of (V_n, \sim) is uniformly bounded regardless of n . Since A is a bounded set, we have $|V_n| = \mathcal{O}(s_n^{-d})$. Let $Y_{j,n}$ denote the number of points of \mathcal{P}_n belonging to

$$\text{Tube}(Q_{j,n}, Mr_n(t_m)) := \{x \in \mathbb{R}^d : \inf_{y \in Q_{j,n}} |x - y| \leq Mr_n(t_m)\}.$$

Then we have

$$\begin{aligned} |\xi_{j,n}| &\leq \sum_{i=1}^m |a_i| \beta_{k,n,A \cap Q_{j,n}}^{(M)}(t_i) \\ &\leq \sum_{i=1}^m |a_i| \beta_k(\check{C}(\mathcal{P}_n \cap \text{Tube}(Q_{j,n}, Mr_n(t_m)), r_n(t_i))) \\ &\leq \sum_{i=1}^m |a_i| \binom{Y_{j,n}}{k+1}. \end{aligned}$$

By definition, $Y_{j,n}$ is Poisson distributed with parameter

$$\lambda_{j,n} := n \int_{\text{Tube}(Q_{j,n}, Mr_n(t_m))} f(z) \, dz,$$

which itself yields an upper bound of the form

$$\lambda_{j,n} \leq n \|f\|_\infty m \left(\text{Tube}(Q_{j,n}, Mr_n(t_m)) \right) := c. \quad (6.21)$$

This implies that $Y_{j,n}$ is stochastically dominated by a Poisson random variable, which we call Y , with parameter c . The assumption $ns_n^d = 1$ ensures that c does not depend on n , and for the rest of the proof, let C^* denote a generic positive constant which is independent of n but may vary between lines.

We get that for $\alpha \in \mathbb{N}$

$$\mathbb{E}[|\xi_{j,n}|^\alpha] \leq \left(\sum_{i=1}^m |a_i| \right)^\alpha \mathbb{E} \left[\left(\frac{Y_{j,n}}{k+1} \right)^\alpha \right] \leq \left(\sum_{i=1}^m |a_i| \right)^\alpha \mathbb{E} \left[\left(\frac{Y}{k+1} \right)^\alpha \right] = C^* \quad (6.22)$$

Letting

$$\xi'_{j,n} := \frac{\xi_{j,n} - \mathbb{E}[\xi_{j,n}]}{\sqrt{\text{Var}(\sum_{i=1}^m a_i \beta_{k,n,A}^{(M)}(t_i))}},$$

it is clear that (V_n, \sim) still constitutes a dependency graph for the $(\xi'_{j,n}, j \in \mathbb{N})$ because independence is not affected by affine transformations. Let Z be a standard normal random variable. It then follows from Stein's normal approximation method (i.e. Theorem 2.4 from [26]) that for all $x \in \mathbb{R}$,

$$\begin{aligned} \left| \mathbb{P} \left(\sum_{j \in V_n} \xi'_{j,n} \leq x \right) - \mathbb{P}(Z \leq x) \right| &\leq C^* \left(\sqrt{s_n^{-d} \mathbb{E}[|\xi'_{j,n}|^3]} + \sqrt{s_n^{-d} \mathbb{E}[|\xi'_{j,n}|^4]} \right) \\ &\leq C^* \left(\sqrt{s_n^{-d} n^{-3/2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^3]} + \sqrt{s_n^{-d} n^{-2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^4]} \right), \end{aligned}$$

where we have applied (6.20) for the second inequality.

Now we have by (6.22) that $\mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^p] \leq C^*$ for $p = 3, 4$, so that

$$s_n^{-d} n^{-p/2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^p] \leq C^* n^{1-p/2} \rightarrow 0, \quad n \rightarrow \infty.$$

From the argument thus far we conclude that

$$\sum_{j \in V_n} \xi'_{j,n} \Rightarrow Z,$$

which in turn implies

$$n^{-1/2} (\beta_{k,n,A}^{(M)}(t_i) - \mathbb{E}[\beta_{k,n,A}^{(M)}(t_i)], i = 1, \dots, m) \Rightarrow \mathcal{N}(0, (\Phi_{k,A}^{(M)}(t_i, t_j))_{i,j=1}^m)$$

for all bounded sets A . The case when A is unbounded can be established by standard approximation arguments nearly identical to those in [21] and [26], so we omit the details and conclude that as $n \rightarrow \infty$

$$n^{-1/2} (\beta_{k,n}^{(M)}(t_i) - \mathbb{E}[\beta_{k,n}^{(M)}(t_i)], i = 1, \dots, m) \Rightarrow \mathcal{N}(0, (\Phi_{k,\mathbb{R}^d}^{(M)}(t_i, t_j))_{i,j=1}^m).$$

This is equivalent to

$$n^{-1/2}(\beta_{k,n}^{(M)}(t_i) - \mathbb{E}[\beta_{k,n}^{(M)}(t_i)], i = 1, \dots, m) \Rightarrow (\mathcal{H}_k^{(M)}(t_i), i = 1, \dots, m),$$

as $n \rightarrow \infty$. Additionally, as $M \rightarrow \infty$

$$(\mathcal{H}_k^{(M)}(t_i), i = 1, \dots, m) \Rightarrow (\mathcal{H}_k(t_i), i = 1, \dots, m),$$

since $\Phi_{k,\mathbb{R}^d}^{(M)}(t_i, t_j) \rightarrow \Phi_{k,\mathbb{R}^d}(t_i, t_j)$ as $M \rightarrow \infty$. According to Theorem 3.2 in [3] it suffices to show that for every $t > 0$ and $\epsilon > 0$,

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left(|\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t) - \mathbb{E}[\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t)]| > \epsilon n^{1/2}\right) = 0. \quad (6.23)$$

By Chebyshev's inequality, the probability in (6.23) is bounded by

$$\frac{1}{\epsilon^2 n} \text{Var}(\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t)),$$

which itself converges to

$$\frac{1}{\epsilon^2} \sum_{i_1=M+1}^{\infty} \sum_{i_2=M+1}^{\infty} \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \left(\frac{\eta_{k,\mathbb{R}^d}^{(i_1, j_1, j_2)}(t_1, t_2) \delta_{i_1, i_2}}{i_1!} + \frac{\nu_{k,\mathbb{R}^d}^{(i_1, i_2, j_1, j_2)}(t_1, t_2)}{i_1! i_2!} \right), \quad n \rightarrow \infty. \quad (6.24)$$

Since $\Phi_{k,\mathbb{R}^d}(t, t)$ is a finite constant, (6.24) goes to 0 as $M \rightarrow \infty$. \square

Proof of Corollary 4.1. Theorem 4.6 in [34] verified that

$$\lim_{n \rightarrow \infty} n^{-1}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]) = 0$$

almost surely. Combining this with Proposition 6.1 (ii) proves the claim. \square

6.2. Central limit theorem in sparse regime

As with the critical regime case, the key results for proving a central limit theorem are those on asymptotic moments that can be seen in the proposition below. As discussed in Section 3, the probabilistic features of these moments are asymptotically determined by $S_{k,n}(t)$. Many functions and objects in Section 6.1 will be carried over for use in this section.

Proposition 6.2. *Let f be an almost everywhere bounded and continuous density function. If $ns_n^d \rightarrow 0$ and $A \subset \mathbb{R}^d$ is open with $m(\partial A) = 0$, then we have that for $t > 0$,*

$$\rho_n^{-1} \mathbb{E}[\beta_{k,n,A}(t)] \rightarrow \mu_{k,A}(t, t), \quad n \rightarrow \infty,$$

and for $t_1, t_2 > 0$,

$$\rho_n^{-1} \text{Cov}(\beta_{k,n,A}(t_1), \beta_{k,n,A}(t_2)) \rightarrow \mu_{k,A}(t_1, t_2), \quad n \rightarrow \infty,$$

where

$$\mu_{k,A}(t_1, t_2) := \frac{1}{(k+2)!} \int_A f(x)^{k+2} dx \int_{\mathbb{R}^{d(k+1)}} h_{t_1}(0, \mathbf{y}) h_{t_2}(0, \mathbf{y}) d\mathbf{y}.$$

Proof. We only discuss the covariance result in the case $A = \mathbb{R}^d$. Throughout the proof we assume $0 < t_1 \leq t_2$. We first derive the same expression as in (6.11) :

$$\text{Cov}(\beta_{k,n}(t_1), \beta_{k,n}(t_2)) = A_{1,n} + A_{2,n},$$

where $A_{1,n}$ and $A_{2,n}$ are given in (6.12), (6.13) respectively. Observing that $g_{r_n(t)}^{(k+2,j)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) = 0$ for all $j \geq 2$ and any $t > 0$, we can split $A_{1,n}$ into two parts, $A_{1,n} = D_{1,n} + D_{2,n}$, where

$$D_{1,n} := \frac{n^{k+2}}{(k+2)!} \mathbb{E}[g_{r_n(t_1)}(\mathcal{X}_{k+2}, \mathcal{X}_{k+2} \cup \mathcal{P}_n) g_{r_n(t_2)}(\mathcal{X}_{k+2}, \mathcal{X}_{k+2} \cup \mathcal{P}_n)],$$

$$D_{2,n} := A_{1,n} - D_{1,n} = \sum_{i=k+3}^{\infty} \sum_{j_1 > 0} \sum_{j_2 > 0} j_1 j_2 \frac{n^i}{i!} \mathbb{E}[g_{r_n(t_1)}^{(i,j_1)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i,j_2)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)],$$

Based on this decomposition, we claim that

$$\rho_n^{-1} D_{1,n} \rightarrow \mu_{k,\mathbb{R}^d}(t_1, t_2), \quad n \rightarrow \infty, \quad (6.25)$$

and $\rho_n^{-1} D_{2,n}$ and $\rho_n^{-1} A_{2,n}$ both converge to 0 as $n \rightarrow \infty$. An important implication of these convergence results is that

$$\rho_n^{-1} \text{Cov}(S_{k,n}(t_1), S_{k,n}(t_2)) \rightarrow \mu_{k,\mathbb{R}^d}(t_1, t_2), \quad n \rightarrow \infty;$$

namely, the covariance of $\beta_{k,n}(t)$ asymptotically coincides with that of $S_{k,n}(t)$.

By what should now be a familiar argument and the customary change of variable, we see that

$$\begin{aligned} \rho_n^{-1} D_{1,n} &= \frac{\rho_n^{-1} n^{k+2}}{(k+2)!} \mathbb{E}[h_{r_n(t_1)}(\mathcal{X}_{k+2}) h_{r_n(t_2)}(\mathcal{X}_{k+2})] \\ &\quad \times \mathbb{E}[J_{k+2, r_n(t_2)}(\mathcal{X}_{k+2}, \mathcal{X}_{k+2} \cup \mathcal{P}_n) | \mathcal{X}_{k+2}] \\ &= \frac{\rho_n^{-1} n^{k+2}}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} h_{r_n(t_1)}(\mathbf{x}) h_{r_n(t_2)}(\mathbf{x}) \exp(-n I_{r_n(t_2)}(\mathbf{x})) \prod_{j=1}^{k+2} f(x_j) d\mathbf{x} \end{aligned} \quad (6.26)$$

$$\begin{aligned}
&= \frac{1}{(k+2)!} \int_{\mathbb{R}^{d(k+1)}} \int_{\mathbb{R}^d} h_{t_1}(0, \mathbf{y}) h_{t_2}(0, \mathbf{y}) \exp(-nI_{r_n(t_2)}(x, x + s_n \mathbf{y})) \\
&\quad \times f(x) \prod_{j=1}^{k+1} f(x + s_n y_j) \, dx \, d\mathbf{y}.
\end{aligned}$$

By the continuity of f it holds that $\prod_{j=1}^{k+1} f(x + s_n y_j) \rightarrow f(x)^{k+1}$ a.e. as $n \rightarrow \infty$.

Moreover, the exponential term converges to 1 because we see that

$$nI_{r_n(t_2)}(x, x + s_n \mathbf{y}) \leq ns_n^d \|f\|_\infty m(\mathcal{B}(\{0, \mathbf{y}\}; t_2)) \rightarrow 0, \quad n \rightarrow \infty.$$

Thus (6.25) follows from the dominated convergence theorem.

Next let us turn to the asymptotics of $\rho_n^{-1} D_{2,n}$. Proceeding as in (6.17), while applying (6.7) and (6.9), we have that

$$\begin{aligned}
\rho_n^{-1} D_{2,n} &\leq \sum_{i=k+3}^{\infty} \sum_{j_1 > 0} \sum_{j_2 > 0} j_1 j_2 \frac{\rho_n^{-1} n^i}{i!} \\
&\quad \times \mathbb{E} \left[\mathbf{1} \{ \check{C}(\mathcal{X}_i, r_n(t_1)) \text{ is connected} \} \prod_{\ell=1}^2 b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_i) \right] \\
&\leq \sum_{i=k+3}^{\infty} \binom{i}{k+1}^2 \frac{\rho_n^{-1} n^i}{i!} \mathbb{P}(\check{C}(\mathcal{X}_i, r_n(t_1)) \text{ is connected}) \\
&\leq \frac{(t_1^d \|f\|_\infty \theta_d)^{k+1}}{((k+1)!)^2} \sum_{i=k+3}^{\infty} b_{i,n},
\end{aligned}$$

where

$$b_{i,n} := \frac{i! i^{i-2}}{((i-k-1)!)^2} (nr_n(t_1)^d \|f\|_\infty \theta_d)^{i-(k+2)}.$$

Obviously $b_{i,n} \rightarrow 0$, $n \rightarrow \infty$ for all $i \geq k+3$. Since $ns_n^d \rightarrow 0$, it is easy to find a summable upper bound $c_i \geq b_{i,n}$ for sufficiently large n . Now the dominated convergence theorem for sums concludes $\rho_n^{-1} D_{2,n} \rightarrow 0$ as $n \rightarrow \infty$.

For the evaluation of $n^{-1} |A_{2,n}|$, we apply (6.19) to the right hand side at (6.18). Slightly changing the description of the resulting bound, we obtain

$$\begin{aligned}
\rho_n^{-1} |A_{2,n}| &\leq 3 \cdot 2^d \frac{(t_2^d \|f\|_\infty \theta_d)^{k+1}}{((k+1)!)^2} \\
&\quad \times \sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \frac{i_1^{i_1-1} i_2^{i_2-1}}{(i_1-k-1)!(i_2-k-1)!} (nr_n(t_2)^d \|f\|_\infty \theta_d)^{i_1+i_2-(k+2)}.
\end{aligned}$$

Since $ns_n^d \rightarrow 0$ as $n \rightarrow \infty$, it follows from the dominated convergence theorem for sums that $\rho_n^{-1} A_{2,n} \rightarrow 0$, $n \rightarrow \infty$, as desired. \square

Proof of Theorem 3.1. We first establish the central limit theorem for $S_{k,n}(t)$ by proceeding in an almost identical fashion to Theorem 4.1. As in that proof, we require that the left-most point of each subset $\mathcal{Y} \subset \mathcal{P}_n$ to lie in an (open) bounded set $A \subset \mathbb{R}^d$, with $m(\partial A) = 0$. Let $V_n, Q_{j,n}$ be defined as in the proof of Theorem 4.1 and recall one more that we assume $0 < t_1 < t_2 < \dots < t_m$. In this case however, we let V_n be the vertex set of a dependency graph by letting $j \sim j'$ if and only if

$$\inf\{|x - y| : x \in Q_{j,n}, y \in Q_{j',n}\} \leq 2(k+2)r_n(t_m).$$

We modify $\xi_{j,n}$ to be defined as

$$\xi_{j,n} := \sum_{i=1}^m a_i \sum_{\mathcal{Y} \subset \mathcal{P}_n} g_{r_n(t_i), A \cap Q_{j,n}}(\mathcal{Y}, \mathcal{P}_n)$$

so that $\sum_{i=1}^m a_i S_{k,n,A}(t_i) = \sum_{j \in V_n} \xi_{j,n}$. Furthermore, $Y_{j,n}$ denotes the number of points of \mathcal{P}_n in $\text{Tube}(Q_{j,n}, (k+2)r_n(t_m))$. Then,

$$|\xi_{j,n}| \leq \sum_{i=1}^m |a_i| \binom{Y_{j,n}}{k+2}.$$

It is easy to demonstrate that the Poisson parameter of $Y_{j,n}$ is bounded by $cn s_n^d$ for some constant $c > 0$ —see (6.21). Letting C^* be a general positive constant as in the proof of Theorem 4.1, we get that for $\alpha \in \mathbb{N}$,

$$\mathbb{E}[|\xi_{j,n}|^\alpha] \leq C^* (n s_n^d)^{k+2}.$$

This in turn implies $\mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^p] \leq C^* (n s_n^d)^{k+2}$ for $p = 3, 4$. Let

$$\xi'_{j,n} := \frac{\xi_{j,n} - \mathbb{E}[\xi_{j,n}]}{\sqrt{\text{Var}(\sum_{i=1}^m a_i S_{k,n,A}(t_i))}}$$

and $Z \sim \mathcal{N}(0, 1)$. As in the critical regime case, Stein's normal approximation method gives

$$\begin{aligned} & \left| \mathbb{P}\left(\sum_{j \in V_n} \xi'_{j,n} \leq x\right) - \mathbb{P}(Z \leq x) \right| \\ & \leq C^* \left(\sqrt{s_n^{-d} \rho_n^{-3/2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^3]} + \sqrt{s_n^{-d} \rho_n^{-2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^4]} \right), \end{aligned}$$

The right-hand side vanishes as $n \rightarrow \infty$, since for $p = 3, 4$,

$$s_n^{-d} \rho_n^{-p/2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^p] \leq C^* \rho_n^{1-p/2} \rightarrow 0, \quad n \rightarrow \infty.$$

Thus we have obtained

$$\rho_n^{-1/2}(S_{k,n}(t_i) - \mathbb{E}[S_{k,n}(t_i)], i = 1, \dots, m) \Rightarrow \mathcal{N}(0, (\mu_{k, \mathbb{R}^d}(t_i, t_j))_{i,j=1}^m). \quad (6.27)$$

The limiting covariance matrix above coincides with the covariance functions of the process \mathcal{G}_k , i.e.,

$$\mathbb{E}[\mathcal{G}_k(t_i)\mathcal{G}_k(t_j)] = C_{f,k} \int_{\mathbb{R}^{d(k+1)}} h_{t_i}(0, \mathbf{y})h_{t_j}(0, \mathbf{y}) d\mathbf{y} = \mu_{k, \mathbb{R}^d}(t_i, t_j), \quad i, j = 1, \dots, m.$$

Therefore (6.27) is equivalent to

$$\rho_n^{-1/2}(S_{k,n}(t_i) - \mathbb{E}[S_{k,n}(t_i)], i = 1, \dots, m) \Rightarrow (\mathcal{G}_k(t_i), i = 1, \dots, m).$$

Now we can finish the entire proof, provided that for every $t > 0$,

$$\rho_n^{-1/2}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]) - \rho_n^{-1/2}(S_{k,n}(t) - \mathbb{E}[S_{k,n}(t)]) \xrightarrow{P} 0, \quad n \rightarrow \infty.$$

This can be proved immediately by Chebyshev's inequality. That is, for every $\epsilon > 0$,

$$\mathbb{P}\left(\rho_n^{-1/2}|\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]| > \epsilon\right) \leq \frac{1}{\epsilon^2 \rho_n} \text{Var}(\beta_{k,n}(t)) \rightarrow 0,$$

where the convergence is a direct consequence of $\rho_n^{-1}D_{2,n} \rightarrow 0$ and $\rho_n^{-1}A_{2,n} \rightarrow 0$, which were verified in the proof of Proposition 6.2. \square

6.3. Poisson limit theorem in sparse regime

Proof of Theorem 5.1. We begin by defining

$$H_{k,n}(t) := \sum_{\mathcal{Y} \subset \mathcal{P}_n} h_{r_n(t)}(\mathcal{Y}),$$

and show that

$$(H_{k,n}(t_i), i = 1, \dots, m) \Rightarrow (\mathcal{V}_k(t_i), i = 1, \dots, m). \quad (6.28)$$

Subsequently we shall verify that for every $t > 0$,

$$H_{k,n}(t) - S_{k,n}(t) \xrightarrow{P} 0, \quad (6.29)$$

$$\beta_{k,n}(t) - S_{k,n}(t) \xrightarrow{P} 0. \quad (6.30)$$

Then the proof of (5.1) will be complete.

Part 1: For the proof of (6.28), it is sufficient to show that for any $a_1, a_2, \dots, a_m > 0$, $m \geq 1$,

$$\sum_{i=1}^m a_i H_{k,n}(t_i) \Rightarrow \sum_{i=1}^m a_i \mathcal{V}_k(t_i).$$

We may use positive constants because of the fact that the Laplace transform characterizes a random vector with values in \mathbb{R}_+^m . We proceed by using Theorem 3.1 from [12]. First let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a generic probability space on which all objects are defined. Let $\mathbf{N}(\mathbb{R}_+)$ be the set of finite counting measures on \mathbb{R}_+ . We equip $\mathbf{N}(\mathbb{R}_+)$ with the vague topology; see, e.g., [28] for more information on the vague topology. Let us define a point process $\xi_n : \Omega \rightarrow \mathbf{N}(\mathbb{R}_+)$ by

$$\xi_n(\cdot) := \sum_{\mathcal{Y} \subset \mathcal{P}_n} \mathbf{1}_{\left\{ \sum_{i=1}^m a_i h_{r_n(t_i)}(\mathcal{Y}) > 0 \right\}} \delta_{\sum_{i=1}^m a_i h_{r_n(t_i)}(\mathcal{Y})}(\cdot),$$

where δ is a Dirac measure.

Additionally let $\zeta : \Omega \rightarrow \mathbf{N}(\mathbb{R}_+)$ denote a Poisson random measure with mean measure $C_{f,k} \tau_k$ where

$$\tau_k(A) := m_k \left\{ \mathbf{y} \in \mathbb{R}^{d(k+1)} : \sum_{i=1}^m a_i h_{t_i}(0, \mathbf{y}) \in A \setminus \{0\} \right\}, \quad A \subset \mathbb{R}_+.$$

The rest of Part 1 is devoted to showing that

$$\xi_n \Rightarrow \zeta \text{ in } \mathbf{N}(\mathbb{R}_+). \quad (6.31)$$

According to Theorem 3.1 in [12], the following two conditions suffice for (6.31). Let $\mathbf{L}_n(\cdot) := \mathbb{E}[\xi_n(\cdot)]$ and $\mathbf{M}(\cdot) := \mathbb{E}[\zeta(\cdot)] = C_{f,k} \tau_k(\cdot)$. The first requirement for (6.31) is the convergence in terms of the *total variation distance*:

$$d_{\text{TV}}(\mathbf{L}_n, \mathbf{M}) := \sup_{A \in \mathcal{B}(\mathbb{R}_+)} |\mathbf{L}_n(A) - \mathbf{M}(A)| \rightarrow 0, \quad n \rightarrow \infty, \quad (6.32)$$

where $\mathcal{B}(\mathbb{R}_+)$ is the Borel σ -field over \mathbb{R}_+ . In addition, the second requirement for (6.31) is

$$v_n := \max_{1 \leq \ell \leq k+1} \int_{\mathbb{R}^{d\ell}} \left(\int_{\mathbb{R}^{d(k+2-\ell)}} \mathbf{1}_{\left\{ \sum_{i=1}^m a_i h_{r_n(t_i)}(x_1, \dots, x_{k+2}) > 0 \right\}} \lambda^{k+2-\ell}(d(x_{\ell+1}, \dots, x_{k+2})) \right)^2 \lambda^\ell(d(x_1, \dots, x_\ell)) \rightarrow 0 \quad (6.33)$$

as $n \rightarrow \infty$, where $\lambda^m = \lambda \otimes \dots \otimes \lambda$ is a product measure on \mathbb{R}^m with $\lambda(\cdot) = n \int f(z) dz$.

Let us now return to (6.32) and present its proof here. As usual, we have assumed $0 < t_1 < t_2 < \dots < t_m$. Then, for any $A \in \mathcal{B}(\mathbb{R}_+)$ we have from Palm theory, the change of variables $x_1 = x$, $x_i = x + s_n y_{i-1}$ for $i = 2, \dots, k+2$, and $\rho_n = 1$ that

$$\begin{aligned} \mathbf{L}_n(A) &= \frac{n^{k+2}}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} \mathbf{1}\left\{\sum_{i=1}^m a_i h_{r_n(t_i)}(\mathbf{x}) \in A \setminus \{0\}\right\} \prod_{j=1}^{k+2} f(x_j) \, d\mathbf{x} \\ &= \frac{1}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} \mathbf{1}\left\{\sum_{i=1}^m a_i h_{t_i}(0, \mathbf{y}) \in A \setminus \{0\}\right\} f(x) \prod_{j=1}^{k+1} f(x + s_n y_j) \, dx \, d\mathbf{y}. \end{aligned}$$

Therefore,

$$\begin{aligned} |\mathbf{L}_n(A) - \mathbf{M}(A)| &\leq \frac{1}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} \mathbf{1}\left\{\sum_{i=1}^m a_i h_{t_i}(0, \mathbf{y}) \in A \setminus \{0\}\right\} \\ &\quad \times f(x) \left| \prod_{j=1}^{k+1} f(x + s_n y_j) - f(x)^{k+1} \right| \, dx \, d\mathbf{y}. \end{aligned}$$

If the indicator function above is equal to 1, then $h_{t_i}(0, \mathbf{y}) = 1$ for at least one i , which means that the distance of each component in \mathbf{y} from the origin must be less than t_m . Otherwise one cannot form a required empty $(k+1)$ -simplex. Hence we have

$$\begin{aligned} |\mathbf{L}_n(A) - \mathbf{M}(A)| &\leq \frac{1}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} \prod_{i=1}^{k+2} \mathbf{1}\{|y_i| \leq t_m\} \\ &\quad \times f(x) \left| \prod_{j=1}^{k+1} f(x + s_n y_j) - f(x)^{k+1} \right| \, dx \, d\mathbf{y}. \end{aligned}$$

We have by continuity of f that $\left| \prod_{j=1}^{k+1} f(x + s_n y_j) - f(x)^{k+1} \right|$ converges to 0 a.e. as $n \rightarrow \infty$ and is bounded by $2\|f\|_\infty^{k+1} < \infty$. So the dominated convergence theorem applies to get $|\mathbf{L}_n(A) - \mathbf{M}(A)| \rightarrow 0$ as $n \rightarrow \infty$. Since this convergence holds uniformly for all $A \in \mathcal{B}(\mathbb{R}_+)$, we have now established (6.32).

Next we turn to proving (6.33). First we can immediately see that

$$\begin{aligned} v_n &= \max_{1 \leq \ell \leq k+1} n^{2k+4-\ell} \int_{\mathbb{R}^{d(2k+4-\ell)}} \mathbf{1}\left\{\sum_{i=1}^m a_i h_{r_n(t_i)}(x_1, \dots, x_{k+2}) > 0\right\} \\ &\quad \times \mathbf{1}\left\{\sum_{i=1}^m a_i h_{r_n(t_i)}(x_1, \dots, x_\ell, x_{k+3}, \dots, x_{2k+4-\ell}) > 0\right\} \prod_{j=1}^{2k+4-\ell} f(x_j) \, d\mathbf{x}. \end{aligned}$$

Making a change of variables with $x_1 = x$ and $x_i = x + s_n y_{i-1}$ for $i = 2, \dots, 2k+4-\ell$, while using $f(x + s_n y_{i-1}) \leq \|f\|_\infty$, we get that

$$v_n \leq \|f\|_\infty^{2k+3-\ell} \max_{1 \leq \ell \leq k+1} n^{2k+4-\ell} s_n^{d(2k+3-\ell)} \int_{\mathbb{R}^{d(2k+3-\ell)}} \mathbf{1}\left\{\sum_{i=1}^m a_i h_{t_i}(0, y_1, \dots, y_{k+1}) > 0\right\}$$

$$\times \mathbf{1} \left\{ \sum_{i=1}^m a_i h_{t_i}(0, y_1, \dots, y_{\ell-1}, y_{k+2}, \dots, y_{2k+3-\ell}) > 0 \right\} dy.$$

Obviously the above integral is finite, and

$$\max_{1 \leq \ell \leq k+1} n^{2k+4-\ell} s_n^{d(2k+3-\ell)} = \max_{1 \leq \ell \leq k+1} (ns_n^d)^{k+2-\ell} \rightarrow 0, \quad n \rightarrow \infty,$$

by the assumption $\rho_n = 1$. So $v_n \rightarrow 0$ follows and (6.33) is obtained.

Part 2: Define the map $\widehat{T} : \mathbf{N}(\mathbb{R}_+) \rightarrow \mathbb{R}_+$ by $\widehat{T}(\sum_n \delta_{x_n}) = \sum_n x_n$. This map is continuous because it is defined on the space of *finite* counting measures. Applying the continuous mapping theorem to (6.31) gives $\widehat{T}(\xi_n) \Rightarrow \widehat{T}(\zeta)$. Equivalently, we have

$$\sum_{i=1}^m a_i H_{k,n}(t_i) \Rightarrow \sum_{i=1}^m a_i \mathcal{V}_k(t_i).$$

To see such equivalence, note that $\widehat{T}(\xi_n) = \sum_{i=1}^m a_i H_{k,n}(t_i)$, so it now suffices to show that $\widehat{T}(\zeta)$ is equal in distribution to $\sum_{i=1}^m a_i \mathcal{V}_k(t_i)$. To this aim let us represent ζ as

$$\zeta \stackrel{d}{=} \sum_{i=1}^{M_n} \delta_{Y_i},$$

where Y_1, Y_2, \dots are i.i.d with common distribution $\tau_k(\cdot)/\tau_k(\mathbb{R}_+)$ and M_n is Poisson distributed with parameter $C_{f,k}\tau_k(\mathbb{R}_+)$. Further, (Y_i) and M_n are independent. On one hand, it follows from the Laplace functional of a Poisson random measure (see Theorem 5.1 in [29]) that for every $\lambda > 0$,

$$\begin{aligned} \mathbb{E} \left[\exp \left(-\lambda \sum_{i=1}^m a_i \mathcal{V}_k(t_i) \right) \right] &= \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}^{d(k+1)}} \lambda \sum_{i=1}^m a_i h_{t_i}(0, \mathbf{y}) M_k(d\mathbf{y}) \right) \right] \\ &= \exp \left(-C_{f,k} \int_{\mathbb{R}^{d(k+1)}} (1 - e^{-\lambda \sum_{i=1}^m a_i h_{t_i}(0, \mathbf{y})}) d\mathbf{y} \right). \end{aligned}$$

On the other hand it is straightforward to compute that

$$\begin{aligned} \mathbb{E} \left[\exp \left(-\lambda \widehat{T}(\zeta) \right) \right] &= \mathbb{E} \left[\exp \left(-\lambda \sum_{i=1}^{M_n} Y_i \right) \right] = \exp \left(-C_{f,k}\tau_k(\mathbb{R}_+) (1 - \mathbb{E}[e^{-\lambda Y_1}]) \right) \\ &= \exp \left(-C_{f,k} \int_{\mathbb{R}^{d(k+1)}} (1 - e^{-\lambda \sum_{i=1}^m a_i h_{t_i}(0, \mathbf{y})}) d\mathbf{y} \right), \end{aligned}$$

implying $\widehat{T}(\zeta) \stackrel{d}{=} \sum_{i=1}^m a_i \mathcal{V}_k(t_i)$ as required.

Part 3: It remains to show (6.29) and (6.30). As for (6.29), we know from (6.25) with $\rho_n = 1$ and $t_1 = t_2 = t$, that

$$\mathbb{E}[S_{k,n}(t)] \rightarrow \mu_{k, \mathbb{R}^d}(t, t), \quad n \rightarrow \infty.$$

Since the exponential term in (6.26) converges to 1 without affecting the value of the limit, it must be that the $\mathbb{E}[H_{k,n}(t)]$ and $\mathbb{E}[S_{k,n}(t)]$ have the same limit. That is,

$$\mathbb{E}[H_{k,n}(t)] \rightarrow \mu_{k,\mathbb{R}^d}(t, t), \quad n \rightarrow \infty,$$

and thus, the Markov inequality gives (6.29).

Finally we turn our attention to (6.30). By Markov's inequality, it suffices to show that $\mathbb{E}[R_{k,n}(t)] \rightarrow 0$ as $n \rightarrow \infty$. Mimicking the derivation of (6.8) with $\rho_n = 1$, we get that

$$\mathbb{E}[R_{k,n}(t)] \leq \sum_{i=k+3}^{\infty} \binom{i}{k+1} \frac{n^i}{i!} \mathbb{P}(\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}).$$

Recalling the bound in (6.9), we have

$$\mathbb{E}[R_{k,n}(t)] \leq \frac{(t^d \|f\|_{\infty} \theta_d)^{k+1}}{(k+1)!} \sum_{i=k+3}^{\infty} \frac{i^{i-2}}{(i-k-1)!} (nr_n(t)^d \|f\|_{\infty} \theta_d)^{i-(k+2)} \rightarrow 0$$

as $n \rightarrow \infty$. □

Acknowledgements: TO's research is partially supported by the National Science Foundation (NSF) grant, Division of Mathematical Science (DMS), #1811428. The authors are very grateful for the detailed and useful comments received from two anonymous referees and an anonymous Associate Editor. These comments helped the authors to introduce a number of improvements to the paper.

References

- [1] ADLER, R. J., BOBROWSKI, O., BORMAN, M. S., SUBAG, E., WEINBERGER, S. ET AL. (2010). Persistent homology for random fields and complexes. In *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*. Institute of Mathematical Statistics pp. 124–143.
- [2] ADLER, R. J., BOBROWSKI, O. AND WEINBERGER, S. (2014). Crackle: The homology of noise. *Discrete & Computational Geometry* **52**, 680–704.
- [3] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edition. Wiley, New York.

- [4] BJÖRNER, A. (1995). *Topological Methods. In Handbook of Combinatorics.* Elsevier, Amsterdam.
- [5] BOBROWSKI, O. AND ADLER, R. J. (2014). Distance functions, critical points, and the topology of random Čech complexes. *Homology, Homotopy and Applications* **16**, 311–344.
- [6] BOBROWSKI, O. AND KAHLE, M. (2018). Topology of random geometric complexes: a survey. *Journal of Applied and Computational Topology* **1**, 331–364.
- [7] BOBROWSKI, O., KAHLE, M. AND SKRABA, P. (2017). Maximally persistent cycles in random geometric complexes. *Ann. Appl. Prob.* **27**, 2032–2060.
- [8] BOBROWSKI, O. AND MUKHERJEE, S. (2015). The topology of probability distributions on manifolds. *Prob. Theory Rel. Fields* **161**, 651–686.
- [9] BOBROWSKI, O. AND WEINBERGER, S. (2017). On the vanishing of homology in random čech complexes. *Random Structures & Algorithms* **51**, 14–51.
- [10] CARLSSON, G. (2014). Topological pattern recognition for point cloud data. *Acta Numerica* **23**, 289–368.
- [11] DECREUSEFOND, L., FERRAZ, E., RANDRIAMBOLOLONA, H. AND VERGNE, A. (2014). Simplicial homology of random configurations. *Adv. Appl. Prob.* **46**, 325–347.
- [12] DECREUSEFOND, L., SCHULTE, M., THÄLE, C. ET AL. (2016). Functional Poisson approximation in Kantorovich–Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Prob.* **44**, 2147–2197.
- [13] DURRETT, R. (2010). *Probability: Theory and Examples* 4th ed. Cambridge university press.
- [14] EDELSBRUNNER, H. AND HARER, J. (2010). *Computational Topology: an Introduction.* American Mathematical Soc.
- [15] GHRIST, R. Barcodes: The persistent topology of data. *Bull. Amer. Math. Soc.* **45**, 61–76.

- [16] GHRIST, R. W. (2014). *Elementary Applied Topology*. Createspace Seattle.
- [17] GOEL, A., TRINH, K. D. AND TSUNODA, K. Strong law of large numbers for Betti numbers in the thermodynamic regime. *Journal of Statistical Physics* **174**, 865–892.
- [18] HATCHER, A. (2001). *Algebraic Topology*. Cambridge University Press.
- [19] HIRAOKA, Y., SHIRAI, T. AND TRINH, K. D. (2018). Limit theorems for persistence diagrams. *Ann. Appl. Prob.* **28**, 2740–2780.
- [20] KAHLE, M. (2011). Random geometric complexes. *Discrete & Computational Geometry* **45**, 553–573.
- [21] KAHLE, M. AND MECKES, E. (2013). Limit theorems for Betti numbers of random simplicial complexes. *Homology, Homotopy and Applications* **15**, 343–374.
- [22] KAHLE, M. AND MECKES, E. (2016). Erratum to Limit theorems for Betti numbers of random simplicial complexes. *Homology, Homotopy and Applications* **18**, 129–142.
- [23] MEESTER, R. AND ROY, R. (1996). *Continuum percolation* vol. 119. Cambridge University Press.
- [24] OWADA, T. Limit theorems for Betti numbers of extreme sample clouds with application to persistence barcodes. *Ann. Appl. Prob.* **28**, 2814–2854.
- [25] OWADA, T. (2017). Functional central limit theorem for subgraph counting processes. *Electronic Journal of Probability* **22**, 38.
- [26] PENROSE, M. (2003). *Random Geometric Graphs*. Oxford University Press.
- [27] PENROSE, M. D. AND YUKICH, J. E. (2003). Weak laws of large numbers in geometric probability. *Ann. Appl. Prob.* **13**, 277–303.
- [28] RESNICK, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.
- [29] RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.

- [30] TRINH, K. D. A remark on the convergence of Betti numbers in the thermodynamic regime. *Pacific Journal of Mathematics for Industry* **9**, 7.
- [31] TRINH, K. D. (2018). On central limit theorems in stochastic geometry. arXiv:1804.02823.
- [32] WASSERMAN, L. (2016). Topological Data Analysis. *Annual Review of Statistics and Its Application* **5**, 501–532.
- [33] YOGESHWARAN, D. AND ADLER, R. J. (2015). On the topology of random complexes built over stationary point processes. *Ann. Appl. Prob.* **25**, 3338–3380.
- [34] YOGESHWARAN, D., SUBAG, E. AND ADLER, R. J. (2017). Random geometric complexes in the thermodynamic regime. *Prob. Theory Rel. Fields* **167**, 107–142.
- [35] ZOMORODIAN, A. AND CARLSSON, G. (2005). Computing persistent homology. *Discrete & Computational Geometry* **33**, 249–274.