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

TAKASHI OWADA AND ANDREW THOMAS

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 and Poisson regimes, as well when the connectivity radius is on the order of $n^{-1/d}$, the critical regime. We establish limit theorems in all of the aforementioned regimes, a central limit theorem for the sparse and critical regimes, and a Poisson limit theorem for the Poisson regime. When the connectivity radius of the Čech complex is $o(n^{-1/d})$, i.e., the sparse and Poisson regimes, we can decompose the limiting processes into a time-changed Brownian motion and 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.

1. Introduction

It’s easy enough to tell the difference between a donut and a mug on a sunny day—it’s much harder in a hurricane. In a state in which there are innumerable ways in which one may classify objects, it may be useful to see that compared to a baseball—the mug with its handle and the donut with its hole—both have something for you to put your hand through. The point being is that in a veritable storm surge of noise, the ability to categorize objects by their most essential structure is an important start in learning the sum of their properties. The problem of analyzing data in the presence of noise has always been a nuisance. With the advent of the application of algebraic topology to probabilistic structures, the ability to capture the most prominent of features of a space has never been closer at hand. These techniques and their corresponding theory typically fall under the umbrella of topological data analysis (TDA).

A brief introduction into concepts of algebraic topology is needed before moving onward. Though our introduction here will be theoretically impoverished, 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 [9, 24]. Treatments from a topological perspective for practitioners of all sorts can be seen in [14], and a rigorous

2010 Mathematics Subject Classification. Primary 60D05. Secondary 55U10, 60F05, 05E45.
Key words and phrases. Random topology, Betti number, Central limit theorem, Poisson limit theorem.
This research is partially supported by the NSF : Probability and Topology #1811428.
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 \( \beta_2(X) = 1 \).

In recent years, there have been growing interests in the theory of random topology \([2,6,16–18,26]\), 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. Additionally \([7]\) studied the maximal (persistent) \( k \)-cycles when an underlying distribution is a uniform Poisson process in the unit cube. Further, \([10]\) 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 \([25,26]\).

An elegant summary on recent progress in the field is provided by \([5]\). The topological objects in these studies are typically constructed from a geometric complex. Among many choices of geometric complexes (see, e.g., \([14]\)), 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 \( [x_{i_0}, \ldots, x_{i_k}] \) is in \( \check{C}(\mathcal{X}, t) \) if \( \bigcap_{j=0}^{k} B(x_i; t/2) \neq \emptyset \),

where \( B(x; r) = \{ y \in \mathbb{R}^d : |x - y| < r \} \) is an open ball of radius \( r \) around \( x \in \mathbb{R}^d \).
Figure 2. Čech complex $\hat{C}(\mathcal{X}, t)$ with $\mathcal{X} = \{x_1, \ldots, x_7\} \subset \mathbb{R}^2$. There are nine 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 $\hat{C}(\mathcal{X}, t)$, since the balls around these points have an non-empty intersection. The 3-simplex $[x_4, x_5, x_6, x_7]$ represents a tetrahedron.

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 $k \geq 0$

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

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 [6, 8, 17]. This setup necessarily dates back to the classical study on random geometric graphs as seen in the monograph [21]. 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 actually a 1-skeleton of a Čech complex. Let $\mathcal{X}_n$ be a set of random points on $\mathbb{R}^d$. Typically it 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, $r_n$ denotes a sequence of connectivity radii of a Čech complex (or a geometric graph). In this setting the behavior of $\hat{C}(\mathcal{X}_n, r_n)$ is classified into several different regimes, depending on how $nr_n^d$ varies as $n \to \infty$. 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 $nr_n^d \to 0$ as $n \to \infty$, the complex is so sparse that many 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., $nr_n^d \to \xi \in (0, \infty)$, then $\hat{C}(\mathcal{X}_n, r_n)$ 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 $nr_n^d \to \infty$ is the dense regime, for which the complex is highly connected with few topological holes. Detailed study of the Betti numbers has yielded a univariate central limit theorem for the sparse regime [17, 18] and for the critical regime [26]. In addition [17] has proven a Poisson convergence result of Betti numbers when $n^{k+2}r_n^{d(k+1)} \to \lambda \in (0, \infty)$ as $n \to \infty$, so that topological holes hardly ever occur.

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

\[
\beta_{k,n}(t) := \beta_k(\hat{C}(X_n, r_n(t))), \quad t > 0.
\]

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, such as the sparse and Poisson regimes, the limiting process of \( \beta_{k,n}(t) \) can be decomposed into the difference of well-known stochastic processes. Specifically, in the sparse regime we can decompose the limiting process into the difference of time-changed Brownian motions and in the Poisson regime we can decompose the limiting process as the difference of time-changed homogenous Poisson processes on the real half-line. In the critical regime however, the limiting process of \( \beta_{k,n}(t) \) 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.

The motivation of reformulating Betti numbers as a stochastic process 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 [9,13,24]. The first [13] is an essential and succinct overview. The second [9] gives a self-contained theoretical treatment of the topological and probabilistic aspects as well as detailed applications.The final one [24] gives an introduction to persistent homology and its applications from a statistical perspective. Theoretically rigorous treatment of persistence homology, especially the computational aspects, can be seen in [12,27]. Considering a family \( (\hat{C}(X_n, r_n(t)), t > 0) \) 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, i.e. the sum of life lengths of all \( k \)-cycles up to time (radius) \( t \), given by

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

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 include [19,20].

From the viewpoints of proof techniques we shall borrow ideas from [17,18,21] and apply sharper variance/covariance bounds than those given in [18] for the analysis of the critical regime. 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 [18]. The argument for the Poisson regime uses a completely different technique based on [11].

As a final remark, unlike [17,18,21] 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
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 \( ns_n^d \to 0 \) and \( n^{k+2}s_n^{d(k+1)} \to \infty \). The fourth section is about the critical regime, in which \( ns_n^d = 1 \), and Section 5 is dedicated to investigating the Poisson regime with \( n^{k+2}s_n^{d(k+1)} = 1 \). The major part of Section 6 is devoted to proving limit theorems for the critical and Poisson regimes. The proof for the sparse regime can be obtained immediately via simple modification of the critical regime case.

2. Setup

To begin, we start by defining some essential concepts towards proving the results in this paper. Due to the ease of proofs in the case of “Poissonization” we only 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 \( \lambda \) which we define as \( \lambda(A) = n \int_A f(x) \, dx \) for all measurable \( A \) in \( \mathbb{R}^d \). In the previous definition \( f \) is a probability density that is almost surely bounded and continuous with respect to Lebesgue measure. Throughout the paper, Lebesgue measure on \( \mathbb{R}^d \) is denoted by \( m_k \) and for convenience we let \( m := m_0 \).

As an aside, we have a few definitions to mention before commencing. First, let \( \|f\|_\infty \) be the essential supremum of the aforementioned \( f \), which is finite as \( f \) is almost surely 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 the quantity

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

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, with \( 1\{ \cdot \} \) denoting an indicator function.

It is useful to define the notion of a finite point cloud throughout the study. We let \( \mathcal{X}_m := \{X_1, X_2, \ldots, X_m\} \) where \( X_i \) are i.i.d with density \( f \) as mentioned before, though let it represent an arbitrary subset of \( \mathbb{R}^d \) of cardinality \( m \) as needed. Thus if \( N_n \) is a Poisson random variable with parameter \( n \), 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 \( \delta_x \) 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 \( h_t(x_1, \ldots, x_{k+2}), x_i \in \mathbb{R}^d \), to be the indicator that \( \tilde{C}(\{x_1, x_2, \ldots, x_{k+2}\}, t) \) contains an empty \((k+1)\)-simplex. This means that \( \tilde{C}(\{x_1, x_2, \ldots, x_{k+2}\}, t) \) 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

\[
h_t(x_1, \ldots, x_{k+2}) = h^+_t(x_1, \ldots, x_{k+2}) - h^-_t(x_1, \ldots, x_{k+2}),
\]
where we define

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

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

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

\[ h_s^\pm(x_1, \ldots, x_{k+2}) \leq h_t^\pm(x_1, \ldots, x_{k+2}) \]

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 \( n \to \infty \) 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

\[ \beta_{k,n}(t) = \sum_{i \geq k+2} \sum_{j > 0} jU_{i,j,n}(t), \quad t > 0. \]

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

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

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

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

In particular, denote

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

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

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

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 $S_{k,n}(t) = \sum_{\mathcal{Y} \in \mathcal{P}_n} g_{r_n(t)}(\mathcal{Y}, \mathcal{P}_n)$ and $U_{i,j,n}(t) = \sum_{\mathcal{Y} \in \mathcal{P}_n} g_{r_n(t)}^{(ij)}(\mathcal{Y}, \mathcal{P}_n)$. As a final bit of notation, let

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

where we require the left-most point of every subset $\mathcal{Y}$ to be an element of $A$ in the calculation of the $k$th Betti number. 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. **Sparse regime**

Throughout this section we assume that $ns_n^d \to 0$ and $\rho_n := n^{k+2}s_n^{d(k+1)} \to \infty$ as $n \to \infty$. The essence of the sparse regime is that Čech complexes are distributed sparsely with many separate connected components, because of a fast decay of $s_n$ as a result of $ns_n^d \to 0$. Consequently, all $k$-cycles in the limit are supported exactly on $k+2$ points ($k+2$ is a “minimum” number necessary to form a $k$-cycle). From a more analytic viewpoint, the behavior of the $k$th Betti number \([26]\) 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 \([18]\).

Before presenting the main result we define the limiting stochastic process

$$G_k(t) := \int_{\mathbb{R}^{d(k+1)}} h_t(0,y) G_k(dy),$$

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 $A_1, \ldots, A_m$ disjoint, $G_k(A_1), \ldots, G_k(A_m)$ are independent. As defined, $G_k(t)$ depends on the indicator $h_t$, meaning that due to sparsity of the Čech complex in this regime, the $k$-cycles affecting $G_k(t)$ 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

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

then $G_k^\pm(t)$ becomes a time-changed Brownian motion; see Proposition \([3.1]\) below. Hence $G_k(t) = G_k^+(t) - G_k^-(t)$ is a difference of two dependent time-changed Brownian motions, where dependence is due to the same Gaussian random measure $G_k$ shared by $G_k^+(t)$ and $G_k^-(t)$. Those wishing to examine this characterization in more detail should refer to \([20]\).
example, it is proven in [20] that the process $G_k(t)$ 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 $G_k^\pm(t)$ can be expressed as

$$(G_k^\pm(t), t \geq 0) \overset{d}{=} (B(C_{f,k}m_k(D_1^\pm)_{t,d(k+1)}), t \geq 0),$$

where $B$ is a standard Brownian motion and $D_1^\pm = \{y \in \mathbb{R}^{d(k+1)} : h_i^\pm(0, y) = 1\}$.

**Proof.** We prove only the result for $G_k^+$, as the proof for $G_k^-$ is the same. It is elementary to show that $G_k^+(t)$ has mean zero. Thus, it only remains to demonstrate the covariance result. Since $h_i^+$ is non-decreasing in $t$, we have $D^+_t \subset D^+_s$ for $0 \leq t_1 \leq t_2$; therefore,

$$
\mathbb{E}[G_k^+(t_1)G_k^+(t_2)] = \mathbb{E}[G_k(D^+_t)G_k(D^+_s)] = \mathbb{E}[G_k(D^+_t)^2] = C_{f,k}m_k(D^+_t)_{t,d(k+1)}.
$$

□

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 $\rho_n$.

**Theorem 3.2.** Suppose that $ns_n^d \to 0$ and $\rho_n = n^{k+2}s_n^{d(k+1)} \to \infty$. 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

$$
\rho_n^{-1/2}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]) \overset{f.d.}{\Rightarrow} G_k(t),
$$

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

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

weakly in $\mathbb{R}^m$.

4. CRITICAL REGIME

We now expand on the results of [26] 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 $i > k + 2$.

As a related work, [26] also established a central limit theorem for the critical regime (though [26] 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 stationary 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 [26] treats the static topology of random Čech complexes whereas we treat the dynamic topology. As a consequence, while the weak limit in [26] 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].

We now define the limiting Gaussian process $\mathcal{H}_k(t)$

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

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

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

Here $\delta_{i_1,i_2}$ 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 $G_k(t)$ 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 $b_{j,t}(x) = 0$ for all $j \geq j_0$, $t > 0$, and $x \in \mathbb{R}^{d_i}$. 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.

In the result below we let $n s_n^d = 1$, though we could easily suppose that $n s_n^d \to 1$ as $n \to \infty$. All proofs are collected in Section 6.1. Our proof is highly analytic in nature, borrowing techniques from [21] and [17], whereas the proof of [26] relies more on topological nature of the objects. 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 $n s_n^d = 1$ and $f$ is an almost everywhere bounded and continuous density function. If $0 < t_1 < t_2 < \cdots < t_m < (\varepsilon ||f||_{\infty} d)^{-1/2}$, 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
\[ n^{-1/2}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]) \xrightarrow{fidi} H_k(t). \]
This means that for every \(m \in \mathbb{N}\) we have
\[ n^{-1/2}(\beta_{k,n}(t_i) - \mathbb{E}[\beta_{k,n}(t_i)], i = 1, \ldots, m) \Rightarrow (H_k(t_i), i = 1, \ldots, m), \]
weakly in \(\mathbb{R}^m\).

**Remark 4.2.** Although Theorem 4.1 imposes a restriction on the range of \(t_i\)'s, we conjecture that the limit theorem holds without such restrictions. Indeed in the case of the “truncated” Betti number
\[ \beta^{(M)}_{k,n}(t) = \sum_{i=k+2}^{M} \sum_{j>0} jU_{i,j,n}(t), \quad M \in \mathbb{N}, \]
which itself is useful for the approximation arguments in our proof, the central limit theorem does hold for every \(t > 0\).

Before concluding this section we shall exploit Theorem 4.6 in [26] and present the strong law of large numbers of \(\beta_{k,n}(t)\). The proof is given at the end of Section 6.1.

**Corollary 4.3.** 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 \(n \to \infty\),
\[ \frac{\beta_{k,n}(t)}{n} \to \sum_{i=k+2}^{\infty} \sum_{j>0} \frac{j}{i!} \eta_{k,i,j,n}^{(i,j)}(t), \quad \text{a.s.} \]

5. **Poisson 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 the sparse regime.

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 \(A_1, \ldots, A_m\) are disjoint, \(M_k(A_1), \ldots, M_k(A_m)\) are independent. We are now ready to define the stochastic process
\[ \mathcal{V}_k(t) = \int_{\mathbb{R}^{d(k+1)}} h_t(0,y) M_k(dy), \]
which appears below as a weak limit in the main theorem. What is interesting about this is that if we define
\[ \mathcal{V}^\pm_k(t) := \int_{\mathbb{R}^{d(k+1)}} h^\pm_t(0,y) M_k(dy), \]
Proposition 5.1. The process $\mathcal{V}_k^\pm$ can be expressed as

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

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

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 $0 = t_0 < t_1 < \ldots < t_k < \infty$ and $\lambda_i > 0$, $i = 1, \ldots, k$, then by the non-decreasingness of $h_t^+$,

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

where $D_t^\pm \setminus D_{t_i-1}^\pm$ are disjoint and $M_k(D_t^\pm \setminus D_{t_i-1}^\pm)$, $i = 1, \ldots, k$, are independent. Moreover, $M_k(D_t^\pm \setminus D_{t_i-1}^\pm)$ is Poisson distributed with parameter $C_{f,k} m_k(D_t^\pm \setminus D_{t_i-1}^\pm) = C_{f,k} m_k(D_t^\pm)(t_i^{d(k+1)} - t_{i-1}^{d(k+1)})$ by a change of variable. Hence we have that

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

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

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

is Poisson with parameter $C_{f,k} m_k(D_t^\pm)t.$

In what follows we assume $\rho_n = 1$, though we could easily modify this to suppose that $\rho_n \to 1$ as $n \to \infty$. 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 11.

Theorem 5.2. 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

$$\beta_{k,n}(t) \overset{f.d.}{\to} \mathcal{V}_k(t),$$

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

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

weakly in $\mathbb{R}^m$. 

6. Proofs

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

In the sequel, we write \( x + y = (x_1 + y_1, \ldots, x_d + y_d) \) for \( x \in \mathbb{R}^d \) and \( y = (y_1, \ldots, y_d) \in \mathbb{R}^d \).

6.1. 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

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

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

\[
\eta_{k,A}^{(i_1,j_1,j_2)}(t_1,t_2) := \int_{\mathbb{R}^{d(i_1-1)}} \int_{\mathbb{R}^d} \mathbf{1}\{C(\{0,y\}, t_1 \wedge t_2) \text{ is connected}\} \prod_{\ell=1}^{2} b_{j_\ell,t_\ell}(0,y) \times \exp \left( - (t_1 \vee t_2)^d f(x)m(B(\{0,y\};1)) \right) f(x)^i A(x) \, dx \, dy,
\]

and

\[
\nu_{k,A}^{(i_1,i_2,j_1,j_2)}(t_1,t_2) := \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^{d(i_1-1)}} dy_1 \int_{\mathbb{R}^{d(i_1-1)}} dy_2 \mathbf{1}\{C(\{0,y_1\}, t_1) \text{ is connected}\} \times \mathbf{1}\{C(\{0,y_2\}, t_2) \text{ is connected}\} b_{j_1,t_1}(0,y_1) b_{j_2,t_2}(y_2)
\times \left[ \left( \alpha_{t_1,t_2}(\{0,y_1\},y_2) - \alpha_{t_1,t_2}(\{0,y_1\},\{0,y_2\}) \right) e^{-f(x)m(B(\{0,y_1\};t_1) \cup B(\{0,y_2\};t_2))} \right.
\times \left. \left( \alpha_{t_1,t_2}(\{0,y_1\},\{0,y_2\}) e^{-f(x)m(B(\{0,y_1\};t_1))} e^{-f(x)m(B(\{0,y_2\};t_2))} \right) \right] f(x)^{i_1+i_2} A(x),
\]

where

\[
B(X; r) := \bigcup_{y \in \mathcal{X}} B(y; r)
\]

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

\[
\alpha_{r,s}(X_1, X_2) := \mathbf{1}\{B(X_1; r) \cap B(X_2; s) \neq \emptyset\},
\]

and \( \alpha_r(X_1, X_2) := \alpha_{r,r}(X_1, X_2) \). Finally we define for \( M \in \mathbb{N} \cup \{\infty\},
\[
\Phi_{k,A}^{(M)}(t_1,t_2) := \sum_{i_1=k+2}^{M} \sum_{i_2=k+2}^{M} \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{\eta_{k,A}^{(i_1,j_1,j_2)}(t_1,t_2)}{i_1!} + \frac{\nu_{k,A}^{(i_1,i_2,j_1,j_2)}(t_1,t_2)}{i_1! i_2!}
\]
where \( \delta_{i_1,i_2} \) 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 \( n s_n^d = 1 \) and \( A \subset \mathbb{R}^d \) is open with \( m(\partial A) = 0 \).

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

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

\[
n^{-1}Cov(\beta_{k,n,A}(t_1), \beta_{k,n,A}(t_2)) \to \Phi_{k,A}(t_1, t_2), \quad n \to \infty.
\]

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

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

\[
n^{-1}Cov(\beta_{k,n,A}(t_1), \beta_{k,n,A}(t_2)) \to \Phi_{k,A}(t_1, t_2), \quad n \to \infty
\]

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 [21], and the monotone convergence theorem supply that

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

where \( \mathcal{X}_i = (X_1, \ldots, X_i) \in \mathbb{R}^{di} \) 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

\[
n^{-1}E[g_{r_n(t)}^{(i,j)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)]
= n^{-1}E[b_{i,r_n(t)}(\mathcal{X}_i)E[J_{r_n(t)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) \mid \mathcal{X}_i]]
= n^{-1} \int_{\mathbb{R}^{di}} 1\{\tilde{C}(x, r_n(t)) \text{ is connected} \} b_{i,r_n(t)}(x) \exp \left(-nI_{r_n(t)}(x)\right) \prod_{j=1}^i f(x_j) \, dx,
\]

where

\[
I_{r_n(t)}(x) = I_{r_n(t)}(x_1, \ldots, x_i) = \int_{B(x;r_n(t))} f(z) \, dz.
\]

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

\[
(n s_n^d)^{i-1} \int_{\mathbb{R}^{d(i-1)}} \int_{\mathbb{R}^d} 1\{\tilde{C}(\{x, y, x + s_n y\}, r_n(t)) \text{ is connected} \} b_{i,r_n(t)}(x, x + s_n y)
\]

...
Indeed this can be derived from

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

Indeed this can be derived from

\[ P(\mathcal{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}) = \int_{\mathbb{R}^d} 1\{\mathcal{C}(x, r_n(t)) \text{ is connected}\} \prod_{j=1}^i f(x_j) \, dx \]
\[\begin{align*}
&= r_n(t)^{d(i-1)} \int_{\mathbb{R}^d} 1\{ \widetilde{C}\{0,y\}, 1 \} \text{ is connected} \} f(x) \prod_{j=1}^{i-1} f(x + r_n(t)y_j) \, dx \, dy \\
&\leq (r_n(t)^d \|f\|_\infty)^{i-1} \int_{\mathbb{R}^{d(i-1)}} 1\{ \widetilde{C}\{0,y\}, 1 \} \text{ is connected} \} \, dy \\
&\leq i^{i-2} (r_n(t)^d \|f\|_\infty \theta_d)^{i-1}.
\end{align*}\]

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 \(n s_n^d = 1\) we conclude that

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

It is easy to check that \(a_{i+1}/a_i \to et^d \|f\|_\infty \theta_d\) as \(i \to \infty\), 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

\[
E[\beta_{k,n}(t_1) \beta_{k,n}(t_2)]
= \sum_{i=k+2} \sum_{j_2=0}^{i+2} j_1 j_2 E \left[ \sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} g_{r_n(t_1)}^{(i_1,j_1)}(\mathcal{Y}_1, \mathcal{P}_n) \, g_{r_n(t_2)}^{(i_2,j_2)}(\mathcal{Y}_2, \mathcal{P}_n) \right]
\]

\[
= \sum_{i=k+2} \sum_{j_2=0}^{i+2} j_1 j_2 E \left[ \sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} g_{r_n(t_1)}^{(i_1,j_1)}(\mathcal{Y}_1, \mathcal{P}_n) \, g_{r_n(t_2)}^{(i_2,j_2)}(\mathcal{Y}_2, \mathcal{P}_n) \right]
\]

\[
+ \sum_{i=k+2} \sum_{j_2=0}^{i+2} j_1 j_2 E \left[ \sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} g_{r_n(t_1)}^{(i_1,j_1)}(\mathcal{Y}_1, \mathcal{P}_n) \, g_{r_n(t_2)}^{(i_2,j_2)}(\mathcal{Y}_2, \mathcal{P}_n) 1\{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = 0\} \right].
\]

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 \(\tilde{C}(\mathcal{Y}_2, r_n(t_2))\) cannot be an isolated component of \(\tilde{C}(\mathcal{P}_n, r_n(t_2))\)—so these terms are zero. Appealing to Palm theory again as seen in [17], we get that

\[
E[\beta_{k,n}(t_1) \beta_{k,n}(t_2)]
= \sum_{i=k+2} \sum_{j_2=0}^{i+2} j_1 j_2 \frac{n^i}{i!} E \left[ g_{r_n(t_1)}^{(i_1,j_1)}(\mathcal{X}_1, \mathcal{X}_1 \cup \mathcal{P}_n) \, g_{r_n(t_2)}^{(i_2,j_2)}(\mathcal{X}_2, \mathcal{X}_1 \cup \mathcal{P}_n) \right]
\]

\[
+ \sum_{i=k+2} \sum_{j_2=0}^{i+2} j_1 j_2 \frac{n^{i_1+i_2}}{i_1! i_2!}
\times E \left[ g_{r_n(t_1)}^{(i_1,j_1)}(\mathcal{X}_1, \mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{P}_n) \, g_{r_n(t_2)}^{(i_2,j_2)}(\mathcal{X}_2, \mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{P}_n) \right],
\]

where \(\mathcal{X}_1\) and \(\mathcal{P}_n\) are independent, and \(\mathcal{X}_1\), \(\mathcal{X}_2\), and \(\mathcal{P}_n\) are also mutually independent such that \(\mathcal{X}_1\) and \(\mathcal{X}_2\) are disjoint.
Applying (6.5) to each \(E[\beta_{k,n}(t_i)]\), \(i = 1, 2\), and utilizing the independence of \(\mathcal{X}_i\) and \(\mathcal{X}_j\),
we see that the covariance function can be written as
\[
\text{Cov}(\beta_{k,n}(t_1), \beta_{k,n}(t_2)) = A_{1,n} + A_{2,n},
\]
with
\[
A_{1,n} := \sum_{i=k+2}^{\infty} \sum_{j_1 > 0} \sum_{j_2 > 0} \frac{n^{i+1}}{i!} E[g_{r_n(t_1)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)g_{r_n(t_2)}^{(i,j_1,j_2)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)],
\]
and
\[
A_{2,n} := \sum_{j_1 = k+2}^{\infty} \sum_{j_2 = k+2}^{\infty} \sum_{j_3 > 0} \sum_{j_4 > 0} \frac{n^{i+1}}{i!} \frac{1}{j_1! j_2!}
\]
\[
\times E[g_{r_n(t_1)}^{(i,j_1)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{X}_2 \cup \mathcal{P}_n)g_{r_n(t_2)}^{(i,j_2)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{X}_2 \cup \mathcal{P}_n)
\]
\[
- g_{r_n(t_1)}^{(i,j_1)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)g_{r_n(t_2)}^{(i,j_2)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n')],
\]
where \(\mathcal{P}_n\) is an independent copy of \(\mathcal{P}_n\) and is also independent of \(\mathcal{X}_i\) and \(\mathcal{X}_j\).

Let us denote the expectation portions of \(A_{1,n}\) and \(A_{2,n}\) as \(E_{1,n}^{(i,j)}\) and \(E_{2,n}^{(i,j)}\), with \(i = (i_1, i_2)\), and \(j = (j_1, j_2)\) 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 \(n \to \infty\). For now we shall compute the limits of \(n^{-1}E_{1,n}^{(i,j)}\) and \(n^{-1+i_1}E_{2,n}^{(i,j)}\) for each \(i, i_1, i_2, j_1, j_2\), 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 \(t_1 \leq t_2\),
\[
n^{-1}E_{1,n}^{(i,j)} = n^{-1}E\left[\mathbf{1}\{\hat{C}(\mathcal{X}_i, r_n(t_1))\text{ is connected}\} \prod_{\ell=1}^{2} b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_i) \exp\left(-n I_{r_n(t_\ell)}(\mathcal{X}_i)\right)\right]
\]
\[
= \int_{\mathbb{R}^{d(i-1)}} \int_{\mathbb{R}^d} \mathbf{1}\{\hat{C}(\{0, y\}, t_1)\text{ is connected}\} \prod_{\ell=1}^{2} b_{j_\ell, t_\ell}(0, y)
\]
\[
\times \exp\left(-n I_{r_n(t_\ell)}(x, x + s_n y)\right) f(x) \prod_{j=1}^{i-1} f(x + s_n y_j) \, dx \, dy
\]
\[
\to \eta_{k,\mathbb{R}^d}^{(i,j_1,j_2)}(t_1, t_2) \quad \text{as} \quad n \to \infty.
\]
Hence the assumed dominated convergence theorem for sums concludes that
\[
n^{-1}A_{1,n} \to \sum_{i=k+2}^{\infty} \sum_{j_1 > 0} \sum_{j_2 > 0} \frac{j_1 j_2}{i!} \eta_{k,\mathbb{R}^d}^{(i,j_1,j_2)}(t_1, t_2) \quad n \to \infty.
\]

To demonstrate convergence for \(n^{1+i_1}E_{2,n}^{(i,j)}\), let us shorten \(g_{r_n(t_1)}\) to \(g_1\) and \(g_{r_n(t_2)}\) to \(g_2\) and decompose \(E_{2,n}^{(i,j)}\) into two terms:
\[
E_{2,n}^{(i,j)} = E\left[g_1(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{X}_2 \cup \mathcal{P}_n)g_2(\mathcal{X}_2, \mathcal{X}_i \cup \mathcal{X}_2 \cup \mathcal{P}_n) - g_1(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)g_2(\mathcal{X}_2, \mathcal{X}_2 \cup \mathcal{P}_n)\right].
\]
Note that for $\ell = 1, 2$,

$$g_{\ell}(X_{i_1}, X_{i_1} \cup X_{i_2} \cup P_n) = g_{\ell}(X_{i_1}, X_{i_2} \cup P_n) 1\{B(X_{i_1}, r_n(t_{\ell})/2) \cap B(X_{i_2}, r_n(t_{\ell})/2) = \emptyset\},$$

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

$$B_{1,n} = -\mathbb{E}[g_1(X_{i_1}, X_{i_1} \cup P_n) g_2(X_{i_2}, X_{i_2} \cup P_n) \alpha_{r_n(t_2)/2}(X_{i_1}, X_{i_2})].$$

At the same time, the spatial independence of $P_n$ justifies that

$$B_{2,n} = \mathbb{E}[g_1(X_{i_1}, X_{i_1} \cup P_n) (g_2(X_{i_2}, X_{i_2} \cup P_n) - g_2(X_{i_2}, X_{i_2} \cup P'_n)) \alpha_{r_n(t_1), r_n(t_2)}(X_{i_1}, X_{i_2})].$$

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

$$E_{2,n}^{(i,j)} = \mathbb{E}[g_1(X_{i_1}, X_{i_1} \cup P_n) (\alpha_{r_n(t_1), r_n(t_2)}(X_{i_1}, X_{i_2}) - \alpha_{r_n(t_2)/2}(X_{i_1}, X_{i_2}))]
\quad - \mathbb{E}[g_1(X_{i_1}, X_{i_1} \cup P_n) g_2(X_{i_2}, X_{i_2} \cup P'_n) \alpha_{r_n(t_1), r_n(t_2)}(X_{i_1}, X_{i_2})]
\quad := C_{1,n} - C_{2,n}.$$

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

$$n^{i_1 + i_2 - 1}C_{1,n} = n^{i_1 + i_2 - 1} \mathbb{E}\left[ \prod_{\ell=1}^{2} 1\{\hat{C}(X_{i_\ell}, r_n(t_{\ell})) \text{ is connected}\} b_{j_\ell, r_n(t_\ell)}(X_{i_\ell})
\quad \times (\alpha_{r_n(t_1), r_n(t_2)}(X_{i_1}, X_{i_2}) - \alpha_{r_n(t_2)/2}(X_{i_1}, X_{i_2}))
\quad \times \exp\left(-n \int_{B(X_{i_1}, r_n(t_1)) \cup B(X_{i_2}, r_n(t_2))} f(z) \, dz\right)\right]
\quad = \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^{d(i_1-1)}} dy_1 \int_{\mathbb{R}^{d(i_2)}} dy_2
\quad 1\{\hat{C}(0, y_1), t_1 \text{ is connected}\}
\quad \times 1\{\hat{C}(y_2, t_2) \text{ is connected}\} b_{j_1, t_1}(0, y_1) b_{j_2, t_2}(y_2)
\quad \times \left(\alpha_{t_1, t_2}(\{0, y_1\}, y_2) - \alpha_{t_2/2}(\{0, y_1\}, y_2)\right)
\quad \times \exp\left(-n \int_{B(\{x, x+s_n y_1\}; r_n(t_1)) \cup B(x+s_n 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})
\quad \rightarrow \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^{d(i_1-1)}} dy_1 \int_{\mathbb{R}^{d(i_2)}} dy_2
\quad 1\{\hat{C}(0, y_1), t_1 \text{ is connected}\}
\quad \times 1\{\hat{C}(y_2, t_2) \text{ is connected}\} b_{j_1, t_1}(0, y_1) b_{j_2, t_2}(y_2)
\quad \times \left(\alpha_{t_1, t_2}(\{0, y_1\}, y_2) - \alpha_{t_2/2}(\{0, y_1\}, y_2)\right)\]
we get that ratio test for sums. The last term is convergent by appealing to the assumption $y_1 = (y_{1,1}, \ldots, y_{1,i_1-1}) \in \mathbb{R}^{d(i_1-1)}$ and $y_2 = (y_{2,1}, \ldots, y_{2,i_2}) \in \mathbb{R}^{d i_2}$. Similarly one can see that

$$n^{i_1+i_2-1} C_{2,n} \to \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^{d(i_1-1)}} dy_1 \int_{\mathbb{R}^{d i_2}} dy_2 \mathbf{1}\{\tilde{C}(\{0, y_1\}, t_1) \text{ is connected}\}$$

$$\times \mathbf{1}\{\tilde{C}(y_2, t_2) \text{ is connected}\} b_{j_1 t_1}(0, y_1) b_{j_2 t_2}(y_2) \alpha_{t_1, t_2}(\{0, y_1\}, y_2)$$

$$\times e^{-f(x) m(B(\{0,y_1\};t_1)) m(B(y_2; t_2))} f(x)^{i_1+i_2}.$$

Therefore,

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

Assuming convergence under summation, we have that

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

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

Now we would like to show that both $n^{-1} E_{1,n}^{(i,j)}$ and $n^{i_1+i_2-1} |E_{2,n}^{(i,j)}|$ 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

$$n^{-1} A_{1,n} \leq \sum_{i=k+2}^{\infty} \sum_{j_1=0}^{i-1} \sum_{j_2=0}^{j_1} \frac{j_1 j_2}{i!} \mathbb{E}\left[\mathbf{1}\{\tilde{C}(X_i, r_n(t_1)) \text{ is connected}\}\prod_{\ell=1}^2 b_{j_\ell r_n(t_\ell)}(X_{i_\ell})\right]$$

$$\leq \sum_{i=k+2}^{\infty} \binom{i}{k+1} \frac{2n^{i-1}}{i!} \mathbb{P}(\tilde{C}(X_i, r_n(t_1)) \text{ is connected})$$

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

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

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

we get that

$$|C_{1,n} - C_{2,n}| \leq 3\mathbb{E}\left[\prod_{\ell=1}^2 \mathbf{1}\{\tilde{C}(X_{i_\ell}, r_n(t_2)) \text{ is connected}\}\right] b_{j_{\ell} r_n(t_\ell)}(X_{i_{\ell}}) \alpha_{r_n(t_2)}(X_{i_1}, X_{i_2}).$$
By virtue of this bound we have that

\[(6.18)\]

\[n^{-1}|A_{2,n}| \leq 3 \sum_{i_1+k+2}^{\infty} \sum_{i_2+k+2}^{\infty} \sum_{j_1>0}^{\infty} \sum_{j_2>0}^{\infty} \frac{n_{i_1+i_2-1}}{i_1!i_2!} j_1j_2 \]

\[\times E \left[ \prod_{\ell=1}^{2} 1 \{ \tilde{C}(\mathcal{X}_{i_{\ell}},r_n(t_{\ell})) \text{ is connected} \} b_{j_{\ell}r_n(t_{\ell})}(\mathcal{X}_{i_{\ell}},\mathcal{X}_{i_{\ell}}) \right] \]

\[\leq 3 \sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \left( \begin{array}{c} i_1 \\ k+1 \end{array} \right) \left( \begin{array}{c} i_2 \\ k+1 \end{array} \right) \frac{n_{i_1+i_2-1}}{i_1!i_2!} \]

\[\times P(\tilde{C}(\mathcal{X}_{i_{\ell}},r_n(t_{\ell})) \text{ is connected for } \ell = 1, 2, \mathcal{B}(\mathcal{X}_{i_1};r_n(t_{2})) \cap \mathcal{B}(\mathcal{X}_{i_2};r_n(t_{2})) \neq \emptyset) \]

We claim here that

\[(6.19)\]

\[P(\tilde{C}(\mathcal{X}_{i_{\ell}},r_n(t_{\ell})) \text{ is connected for } \ell = 1, 2, \mathcal{B}(\mathcal{X}_{i_1};r_n(t_{2})) \cap \mathcal{B}(\mathcal{X}_{i_2};r_n(t_{2})) \neq \emptyset) \]

\[\leq 2^{d}i_1^{-1}i_2^{-1}(r_n(t_{2})^d||f||_{\infty}^{\theta_d})^{i_1+i_2-1}.\]

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

\[P(\tilde{C}(\mathcal{X}_{i_{\ell}},r_n(t_{\ell})) \text{ is connected for } \ell = 1, 2, \mathcal{B}(\mathcal{X}_{i_1};r_n(t_{2})) \cap \mathcal{B}(\mathcal{X}_{i_2};r_n(t_{2})) \neq \emptyset) \]

\[\leq (r_n(t_{2})^d||f||_{\infty})^{i_1+i_2-1} \int_{\mathbb{R}^d(i_1+i_2-1)} \left[ \tilde{C}(\{0,y_1,\ldots,y_{i_1-1}\},1) \text{ is connected} \right] \]

\[\times 1 \{ \tilde{C}(\{y_{i_1},\ldots,y_{i_1+i_2-1}\},1) \text{ is connected} \} \]

\[\times 1 \{ \mathcal{B}(\{0,y_1,\ldots,y_{i_1-1}\};1) \cap \mathcal{B}(\{y_{i_1},\ldots,y_{i_1+i_2-1}\};1) \neq \emptyset \} \] dy.

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

\[(r_n(t_{2})^d||f||_{\infty})^{i_1+i_2-1} \frac{i_1^{i_1-2}i_2^{i_2-2}2^d\theta_d^{i_1+i_2-2}}{t_1t_2^d||f||_{\infty}^{\theta_d}} = 2^{d}i_1^{-1}i_2^{-1}(r_n(t_{2})^d||f||_{\infty}^{\theta_d})^{i_1+i_2-1}.\]

Now we have

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

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,

\[n^{-1/2}(\beta_{k,n}^{(M)}(t_i) - \mathbb{E}[\beta_{k,n}^{(M)}(t_i)], i = 1, \ldots, m) \Rightarrow (\mathcal{H}_{k}^{(M)}(t_i), i = 1, \ldots, m),\]
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 Cramèr-Wold device, 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 \((i)\):

\[
\text{Var}
\left(n^{-1/2} \sum_{i=1}^{m} a_i \beta_{k,n,A}^{(M)}(t_i)\right) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_ia_jn^{-1} \text{Cov}(\beta_{k,n,A}^{(M)}(t_i), \beta_{k,n,A}^{(M)}(t_j))
\]

\[
\rightarrow \sum_{i=1}^{m} \sum_{j=1}^{m} a_ia_j \Phi_{k,A}^{(M)}(t_i, t_j), \quad n \to \infty.
\]

(6.20)

Our proof exploits Stein’s normal approximation method for weakly dependent random variables, as in Theorem 2.4 in [21]. We assume the limit in (6.20) is positive as otherwise our proof is trivial. Define \( t := \max\{t_1, \ldots, t_m\} = t_m \) and let \( \{Q_{j,n}, j \in \mathbb{N}\} \) be an enumeration of almost disjoint closed cubes of side length \( r_n(t) \), such that \( \bigcup_{j \in \mathbb{N}} Q_{j,n} = \mathbb{R}^d \). Recalling \( A \) is bounded, we define

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

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 [21] for the definition) by declaring that for \( j, j' \in V_n \), \( j \sim j' \) if and only if \( d(Q_{j,n}, Q_{j',n}) \leq 2Mr_n(t) \). 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 \( |V_n| = O(s_n^{-d}) \). Let \( Y_{j,n} \) denote the number of points of \( P_n \) belonging to

\[
\text{Tube}(Q_{j,n}, Mr_n(t)) := \left\{ x \in \mathbb{R}^d : \inf_{y \in Q_{j,n}} |x - y| \leq Mr_n(t) \right\}.
\]

Then we have

\[
|\xi_{j,n}| \leq \sum_{i=1}^{m} |a_i| \beta_{k,n,A \cap Q_{j,n}}^{(M)}(t_i)
\]

\[
\leq \sum_{i=1}^{m} |a_i| \beta_k \left( \tilde{C}(P_n \cap \text{Tube}(Q_{j,n}, Mr_n(t)), r_n(t_i)) \right)
\]

\[
\leq \sum_{i=1}^{m} |a_i| \left( \frac{Y_{j,n}}{k + 1} \right),
\]
By definition, \( Y_{j,n} \) is Poisson distributed with parameter

\[
\lambda_{j,n} := n \int_{\text{Tube}(Q_{j,n}, Mr_n(t))} f(z) \, dz,
\]

which itself yields an upper bound of the form

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

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 \( \alpha \in \mathbb{N} \)

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

Letting

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

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 [21]) that for all \( x \in \mathbb{R} \),

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

\[
\leq C^* \left( \sqrt{s_n^{-d} n^{-3/2} \mathbb{E}[(\xi_{j,n} - \mathbb{E}[\xi_{j,n}])^3]} + \sqrt{s_n^{-d} n^{-2} \mathbb{E}[(\xi_{j,n} - \mathbb{E}[\xi_{j,n}])^4]} \right),
\]

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

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

From the argument thus far we conclude that

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

which in turn implies

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

for all bounded sets \( A \). The case when \( A \) is unbounded can be established by standard approximation arguments nearly identical to those in [17] and [21], so we omit the details and conclude that as \( n \to \infty \)

\[
n^{-1/2} \left( \beta_{k,n}(t_i) - \mathbb{E}[\beta_{k,n}(t_i)], i = 1, \ldots, m \right) = \mathcal{N}(0, (\Phi_{k,A}(t_i, t_j))_{i,j=1}^m).
\]
This is equivalent to
\[ n^{-1/2}(\beta_{k,n}^{(M)}(t_i) - \mathbb{E}[\beta_{k,n}^{(M)}(t_i)], i = 1, \ldots, m) \Rightarrow (H_k^{(M)}(t_i), i = 1, \ldots, m), \]
as \( n \to \infty \). Additionally, as \( M \to \infty \)
\[ (H_k^{(M)}(t_i), i = 1, \ldots, m) \Rightarrow (H_k(t_i), i = 1, \ldots, m), \]
since \( \Phi_{k,R,d}^{(M)}(t_i, t_j) \to \Phi_{k,R,d}(t_i, t_j) \) as \( M \to \infty \). According to Theorem 3.2 in [3] it suffices to show that for every \( t > 0 \) and \( \epsilon > 0 \),
\[ \lim_{M \to \infty} \lim_{n \to \infty} \mathbb{P}\left( |\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t) - \mathbb{E}[\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t)]| > \epsilon n^{1/2} \right) = 0. \]
By the Cauchy-Schwarz 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
\[ \frac{1}{\epsilon^2} \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} j_1 j_2 \left( \frac{\eta_{k,R,d}^{(i_1,j_1,j_2)}(t_1, t_2) \delta_{i_1,i_2}}{i_1!} + \frac{\nu_{k,R,d}^{(i_1,j_1,j_2)}(t_1, t_2)}{i_1! i_2!} \right) \text{ as } n \to \infty. \]
Since \( \Phi_{k,R,d}(t, t) \) is a finite constant, (6.24) goes to 0 as \( M \to \infty \). \( \square \)

Proof of Corollary 4.3. Theorem 4.6 in [26] verified that
\[ \lim_{n \to \infty} n^{-1}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]) = 0 \]
almost surely. Combining this with Proposition 6.1 (ii) proves the claim. \( \square \)

6.2. 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 \( n s_n^d \to 0 \) and \( A \subset \mathbb{R}^d \) is open with \( m(\partial A) = 0 \), then we have that for \( t > 0 \),
\[ \rho_n^{-1} \mathbb{E}[\beta_{k,n,A}(t)] \to \mu_{k,A}(t), \quad n \to \infty, \]
and for \( t_1, t_2 > 0 \),
\[ \rho_n^{-1} \text{Cov}(\beta_{k,n,A}(t_1), \beta_{k,n,A}(t_2)) \to \mu_{k,A}(t_1, t_2), \quad n \to \infty, \]
where
\[ \mu_{k,A}(t_1, t_2) := \frac{1}{(k + 2)!} \int_A f(x)^{k+2} \, dx \int_{\mathbb{R}^{d(k+1)}} h_{t_1}(0, y) h_{t_2}(0, y) \, dy. \]
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 $g_{r_n(t)}^{(k+2,j)}(\mathcal{X}_i, \mathcal{X}_i \cup P_n) = 0$ 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

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

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

Based on this decomposition, we claim that

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

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

$$\rho_n^{-1} D_{1,n} = \frac{\rho_n^{-1} n^{k+2}}{(k+2)!} \mathbb{E}[h_{r_n(t_1)}(\mathcal{X}_{k+2}) h_{r_n(t_2)}(\mathcal{X}_{k+2}) \mathbb{E}[J_{k+2,r_n(t_2)}(\mathcal{X}_{k+2}, \mathcal{X}_{k+2} \cup P_n)] | \mathcal{X}_{k+2}]$$

$$= \frac{\rho_n^{-1} n^{k+2}}{(k+2)!} \int_{\mathbb{R}^d} h_{r_n(t_1)}(x) h_{r_n(t_2)}(x) \exp(-nI_{r_n(t_2)}(x)) \prod_{j=1}^{k+2} f(x_j) \, dx$$

$$= \frac{1}{(k+2)!} \int_{\mathbb{R}^d(k+2)} \int_{\mathbb{R}^d} h_{t_1}(0,y) h_{t_2}(0,y) \exp(-nI_{r_n(t_2)}(x, x+s_n y))$$

$$\times f(x) \prod_{j=1}^{k+1} f(x+s_n y_j) \, dx \, dy.$$
Theorem 4.1, we get that for α ρ−1 ns \sum_{i=k+3}^{\infty} \binom{i}{k+1} \frac{\rho_{n}^{-1}}{(k+1)!} \mathbb{P}(\tilde{C}(\mathcal{X}_i, r_n(t_1))) is connected) \leq \sum_{i=k+3}^{\infty} b_{i,n},

where

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

Obviously \( b_{i,n} \to 0, n \to \infty \) for all \( i \geq k + 3 \). Since \( ns_n^d \to 0 \), 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 \( \rho_{n}^{-1} D_{2,n} \to 0 \) as \( n \to \infty \).

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

Since \( ns_n^d \to 0 \) as \( n \to \infty \), it follows from the dominated convergence theorem for sums that \( \rho_{n}^{-1} A_{2,n} \to 0, n \to \infty \), as desired. \hfill \Box

**Proof of Theorem 3.2.** 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 \( Y \subset \mathcal{P}_n \) to lie in an (open) bounded set \( A \subset \mathbb{R}^d \), with \( m(\partial A) = 0 \). Let \( V_n, Q_{j,n} \) and \( t \) be defined as in the proof of Theorem 4.1. 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 \( d(Q_{j,n}, Q_{j',n}) \leq 2(k+2)r_n(t) \).

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 Tube\((Q_{j,n}, (k+2)r_n(t))\). 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 \( cns_n^d \) 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 \( \alpha \in \mathbb{N} \),

\[ \mathbb{E}[|\xi_{j,n}|^\alpha] \leq C^*(ns_n^d)^{k+2}. \]
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
\[
\xi'_{j,n} := \frac{\xi_{j,n} - \mathbb{E}[\xi_{j,n}]}{\sqrt{\operatorname{Var}(\sum_{i=1}^m a_i S_{k,n,A}(t_i))}}
\]
and \( Z \sim \mathcal{N}(0, 1) \). As in the critical regime case, Stein’s normal approximation method gives
\[
\left| \mathbb{P}\left( \sum_{j \in V_n} \xi'_{j,n} \leq x \right) - \mathbb{P}(Z \leq x) \right| \leq C^* \left( \sqrt{s_n^d \rho_n^{-3/2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^3]} + \sqrt{s_n^d \rho_n^{-2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^4]} \right).
\]
The right-hand side vanishes as \( n \to \infty \), since for \( p = 3, 4 \),
\[
s_n^d \rho_n^{-p/2} \mathbb{E}[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^p] \leq C_* \rho_n^{1-p/2} \to 0, \quad n \to \infty.
\]
Thus we have obtained
\[
(6.27) \quad \rho_n^{-1/2} \left( S_{k,n}(t_i) - \mathbb{E}[S_{k,n}(t_i)], i = 1, \ldots, m \right) \Rightarrow \mathcal{N}(0, (\mu_{k, \mathbb{R}^d}(t_i, t_j))_{i,j=1}^m).
\]
The limiting covariance matrix above coincides with the covariance functions of the process \( \mathcal{G}_k \), i.e.,
\[
\mathbb{E}[\mathcal{G}_k(t_i) \mathcal{G}_k(t_j)] = C_{f,k} \int_{\mathbb{R}^{d(k+1)}} h_{t_i}(0, y) h_{t_j}(0, y) \, dy = \mu_{k, \mathbb{R}^d}(t_i, t_j), \quad i, j = 1, \ldots, m.
\]
Therefore (6.27) is equivalent to
\[
\rho_n^{-1/2} \left( S_{k,n}(t_i) - \mathbb{E}[S_{k,n}(t_i)], i = 1, \ldots, m \right) \Rightarrow (\mathcal{G}_k(t_i), i = 1, \ldots, m).
\]
Now we can finish the entire proof, provided that for every \( t > 0 \),
\[
\rho_n^{-1/2} \left( \beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)] \right) - \rho_n^{-1/2} \left( S_{k,n}(t) - \mathbb{E}[S_{k,n}(t)] \right) \overset{P}{\to} 0, \quad n \to \infty.
\]
This can be proved immediately by the Cauchy-Schwarz inequality. That is, for every \( \epsilon > 0 \),
\[
\mathbb{P}\left( \rho_n^{-1/2} |R_{k,n}(t) - \mathbb{E}[R_{k,n}(t)]| > \epsilon \right) \leq \frac{1}{\epsilon^2 \rho_n} \operatorname{Var}(R_{k,n}(t)) \to 0,
\]
where the convergence is a direct consequence of \( \rho_n^{-1} D_{2,n} \to 0 \) and \( \rho_n^{-1} A_{2,n} \to 0 \), which were verified in the proof of Proposition 6.2. \( \square \)

6.3. Poisson regime.

Proof of Theorem 5.2. We begin by defining
\[
H_{k,n}(t) := \sum_{Y \subseteq \mathcal{P}_n} h_{r_n(t)}(Y),
\]
and show that
\[
(6.28) \quad (H_{k,n}(t_i), i = 1, \ldots, m) \Rightarrow (V_k(t_i), i = 1, \ldots, m).
\]
Subsequently we shall verify that for every \( t > 0 \),
\[
(6.29) \quad H_{k,n}(t) - S_{k,n}(t) \overset{P}{\to} 0,
\]
\[
\beta_{k,n}(t) - S_{k,n}(t) \xrightarrow{p} 0.
\]

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 \(a_1, a_2, \ldots, a_m > 0, m \geq 1,\)
\[
\sum_{i=1}^{m} a_i H_{k,n}(t_i) \Rightarrow \sum_{i=1}^{m} a_i V_k(t_i).
\]

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 [11]. First let \((\Omega, \mathcal{F}, \mathbb{P})\) denote a generic probability space on which all objects are defined. Let \(\mathcal{N}(\mathbb{R}_+^+)\) be the set of finite counting measures on \(\mathbb{R}_+^+\). We equip \(\mathcal{N}(\mathbb{R}_+^+)\) with the vague topology; see, e.g., [22] for more information on the vague topology. Let us define a point process \(\xi_n : \Omega \rightarrow \mathcal{N}(\mathbb{R}_+^+)\) by
\[
\xi_n(\cdot) := \sum_{\mathcal{Y} \subset \mathcal{P}_n} 1\{\sum_{i=1}^{m} a_i h_{r_n(t_i)}(\mathcal{Y}) > 0\} \delta_{\sum_{i=1}^{m} a_i h_{r_n(t_i)}(\mathcal{Y})}(\cdot),
\]
where \(\delta\) is a Dirac measure.

Additionally let \(\zeta : \Omega \rightarrow \mathcal{N}(\mathbb{R}_+^+)\) denote a Poisson random measure with mean measure \(C_{f,k} \tau_k\) where
\[
\tau_k(A) := m_k \left\{ y \in \mathbb{R}^{d(k+1)} : \sum_{i=1}^{m} a_i h_{t_i}(0, y) \in A \setminus \{0\} \right\}, \quad A \subset \mathbb{R}_+.
\]

The rest of Part 1 is devoted to showing that (6.31)
\[
\xi_n \Rightarrow \zeta \text{ in } \mathcal{N}(\mathbb{R}_+^+)\]

According to Theorem 3.1 in [11], the following two conditions suffice for (6.31). Let \(L_n(\cdot) := \mathbb{E}[\xi_n(\cdot)]\) and \(M(\cdot) := \mathbb{E}[\zeta(\cdot)] = C_{f,k} \tau_k(\cdot)\). The first requirement for (6.31) is the convergence in terms of the total variation distance:
\[
(6.32) \quad d_{TV}(L_n, M) := \sup_{A \in \mathcal{B}(\mathbb{R}_+)} |L_n(A) - M(A)| \rightarrow 0, \quad n \rightarrow \infty,
\]
where \(\mathcal{B}(\mathbb{R}_+)\) is the Borel \(\sigma\)-field over \(\mathbb{R}_+\). In addition, the second requirement for (6.31) is
\[
(6.33) \quad v_n := \max_{1 \leq \ell \leq k+1} \int_{\mathbb{R}_+^d} \left( \int_{\mathbb{R}_+^{d(k+2)-\ell}}} 1\{\sum_{i=1}^{m} a_i h_{r_n(t_i)}(x_1, \ldots, x_{k+2}) > 0\} \right) \lambda^{k+2-\ell}(d(x_{\ell+1}, \ldots, x_{k+2})) \lambda^\ell(d(x_1, \ldots, x_\ell)) \rightarrow 0
\]
as \(n \rightarrow \infty\), where \(\lambda^m = \lambda \otimes \cdots \otimes \lambda\) 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. Let \(t := \max\{t_1, \ldots, t_m\} = t_m\). Then, for any \(A \in \mathcal{B}(\mathbb{R}_+)\) we have from Palm theory, the change of variables \(x_1 = x\),
\[ x_i = x + s_n y_{i-1} \text{ for } i = 2, \ldots, k + 2, \text{ and } \rho_n = 1 \] that
\[
L_n(A) = \frac{n^{k+2}}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} 1\left\{ \sum_{i=1}^{m} a_i h_{r_n(t_i)}(x) \in A \setminus \{0\} \right\} \prod_{j=1}^{k+2} f(x_j) \, dx
\]
\[
= \frac{1}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} 1\left\{ \sum_{i=1}^{m} a_i h_{t_i}(0, y) \in A \setminus \{0\} \right\} f(x) \prod_{j=1}^{k+1} f(x + s_n y_j) \, dx \, dy.
\]
Therefore,
\[
|L_n(A) - M(A)| \leq \frac{1}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} \prod_{i=1}^{k+2} 1\{ |y_i| \leq t \} f(x) \left| \prod_{j=1}^{k+1} f(x + s_n y_j) - f(x)^{k+1} \right| \, dx \, dy.
\]
If the indicator function above is equal to 1, then \( h_{t_i}(0, y) = 1 \) for at least one \( i \), which means that the distance of each component in \( y \) from the origin must be less than \( t \). Otherwise one cannot form a required empty \( (k+1) \)-simplex. Hence we have
\[
|L_n(A) - M(A)| \leq \frac{1}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} \prod_{i=1}^{k+2} 1\{ |y_i| \leq t \} f(x) \left| \prod_{j=1}^{k+1} f(x + s_n y_j) - f(x)^{k+1} \right| \, dx \, dy.
\]
We have by continuity of \( f \) that \( \prod_{j=1}^{k+1} f(x + s_n y_j) - f(x)^{k+1} \) converges to 0 a.e. as \( n \to \infty \) and is bounded by \( 2\|f\|_{\infty}^{k+1} < \infty \). So the dominated convergence theorem applies to get \( |L_n(A) - M(A)| \to 0 \) as \( n \to \infty \). 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
\[
v_n = \max_{1 \leq \ell \leq k+1} n^{2k+4-\ell} \int_{\mathbb{R}^{d(2k+4-\ell)}} 1\left\{ \sum_{i=1}^{m} a_i h_{r_n(t_i)}(x_1, \ldots, x_{k+2}) > 0 \right\}
\]
\[
\times 1\left\{ \sum_{i=1}^{m} a_i h_{r_n(t_i)}(x_1, \ldots, x_\ell, x_{k+3}, \ldots, x_{2k+4-\ell}) > 0 \right\} \prod_{j=1}^{2k+4-\ell} f(x_j) \, dx.
\]
Making a change of variables with \( x_1 = x \) and \( x_i = x + s_n y_{i-1} \) for \( i = 2, \ldots, 2k+4 - \ell \), while using \( f(x + s_n y_{i-1}) \leq \|f\|_\infty \), we get that
\[
v_n \leq \|f\|_\infty^{2k+3-\ell} \max_{1 \leq \ell \leq k+1} n^{2k+4-\ell} s_n^{d(2k+3-\ell)} \int_{\mathbb{R}^{d(2k+3-\ell)}} 1\left\{ \sum_{i=1}^{m} a_i h_{t_i}(0, y_1, \ldots, y_{k+1}) > 0 \right\}
\]
\[
\times 1\left\{ \sum_{i=1}^{m} a_i h_{t_i}(0, y_1, \ldots, y_{\ell-1}, y_{k+2}, \ldots, y_{2k+3-\ell}) > 0 \right\} \, dy.
\]
Obviously the above integral is finite, and
\[
\max_{1 \leq \ell \leq k+1} n^{2k+4-\ell} s_n^{d(2k+3-\ell)} = \max_{1 \leq \ell \leq k+1} (n s_n^d)^{k+2-\ell} \to 0, \quad n \to \infty,
\]
by the assumption \( \rho_n = 1 \). So \( v_n \to 0 \) follows and (6.33) is obtained.
Part 2: Define the map $\hat{T} : N(\mathbb{R}_+) \to \mathbb{R}_+$ by $\hat{T}(\sum_n \delta_{x_n}) = \sum_n x_n$. This map is continuous because it is defined on the space of finite counting measures. Applying the continuous mapping theorem to (6.31) gives $\hat{T}(\xi_n) \Rightarrow \hat{T}(\zeta)$. Equivalently, we have

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

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

$$
\zeta \overset{d}{=} \sum_{i=1}^M \delta_{Y_i},
$$

where $Y_1, Y_2, \ldots$ 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 [23]) that for every $\lambda > 0$,

$$
\mathbb{E}\left[\exp\left(-\lambda \sum_{i=1}^m a_i V_k(t_i)\right)\right] = \mathbb{E}\left[\exp\left(-\int_{\mathbb{R}^{d(k+1)}} \lambda \sum_{i=1}^m a_i h_{ti}(0,y) M_k(dy)\right)\right]
$$

$$
= \exp\left(-C_{f,k} \int_{\mathbb{R}^{d(k+1)}} (1 - e^{-\lambda \sum_{i=1}^m a_i h_{ti}(0,y)}) dy\right)
$$

On the other hand it is straightforward to compute that

$$
\mathbb{E}\left[\exp\left(-\lambda \hat{T}(\zeta)\right)\right] = \mathbb{E}\left[\exp\left(-\lambda \sum_{i=1}^M Y_i\right)\right] = \exp\left(-C_{f,k} \tau_k(\mathbb{R}_+)(1 - \mathbb{E}[e^{-\lambda Y_1}])\right)
$$

$$
= \exp\left(-C_{f,k} \int_{\mathbb{R}^{d(k+1)}} (1 - e^{-\lambda \sum_{i=1}^m a_i h_{ti}(0,y)}) dy\right),
$$

implying $\hat{T}(\zeta) \overset{d}{=} \sum_{i=1}^m a_i V_k(t_i)$ 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 $t_1 = t_2$, that

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

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,

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

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 $\mathbb{E}[R_{k,n}(t)] \to 0$ as $n \to \infty$. Mimicking the derivation of (6.8) with $\rho_n = 1$, we get that

$$
\mathbb{E}[R_{k,n}(t)] \leq \sum_{i=k+3}^{\infty} \left( \frac{n}{k+1} \right) \frac{n^i}{i!} \mathbb{P}(\bar{C}(X_i, r_n(t)) \text{ is connected}).
$$
Recalling the bound in (6.9), we have
\[
\mathbb{E}[R_{k,n}(t)] \leq \frac{(t^d \|f\|_{\infty} \theta_d)^{k+1} \sum_{i=k+3}^{\infty} \frac{i^{-2}}{(i-k-1)!} (nr_n(t)^d \|f\|_{\infty} \theta_d)^{i-(k+2)}}{(k+1)!} \to 0
\]
as \(n \to \infty\). □

**REFERENCES**

[1] R. J. Adler, O. Bobrowski, M. S. Borman, E. Subag, S. Weinberger, et al., *Persistent homology for random fields and complexes*, Borrowing strength: theory powering applications—a festschrift for lawrence d. brown, 2010, pp. 124–143.

[2] R. J. Adler, O. Bobrowski, and S. Weinberger, *Crackle: The homology of noise*, Discrete & Computational Geometry 52 (2014), no. 4, 680–704.

[3] P. Billingsley, *Convergence of probability measures*, 2nd edition, Wiley, New York, 1999.

[4] A. Björner, *Topological methods. in handbook of combinatorics*, Elsevier, Amsterdam, 1995.

[5] O. Bobrowski and M. Kahle, *Topology of random geometric complexes: a survey*, Journal of Applied and Computational Topology 1 (2018), 331–364.

[6] O. Bobrowski and R. J. Adler, *Distance functions, critical points, and the topology of random Čech complexes*, Homology, Homotopy and Applications 16 (2014), no. 2, 311–344, available at [arXiv:math/1107.4775](http://arxiv.org/abs/1107.4775).

[7] O. Bobrowski, M. Kahle, and P. Skraba, *Maximally persistent cycles in random geometric complexes*, The Annals of Applied Probability 27 (2017), no. 4, 2032–2060.

[8] O. Bobrowski and S. Mukherjee, *The topology of probability distributions on manifolds*, Probability Theory and Related Fields 161 (2015), no. 3, 651–686.

[9] G. Carlsson, *Topological pattern recognition for point cloud data*, Acta Numerica 23 (2014), 289–368.

[10] L. Decreusefond, E. Ferraz, H. Randriambololona, and A. Vergne, *Simplicial homology of random configurations*, Advances in Applied Probability 46 (2014), no. 2, 325–347.

[11] L. Decreusefond, M. Schulte, C. Thäle, et al., *Functional Poisson approximation in Kantorovich–Rubinstein distance with applications to u-statistics and stochastic geometry*, The Annals of Probability 44 (2016), no. 3, 2147–2197.

[12] H. Edelsbrunner and J. Harer, *Computational topology: an introduction*, American Mathematical Soc., 2010.

[13] R. Ghrist, *Barcodes: The persistent topology of data*, Bulletin of the American Mathematical Society 45 (Oct. 2007), no. 01, 61–76.

[14] R. W. Ghrist, *Elementary applied topology*, Createspace Seattle, 2014.

[15] A. Hatcher, *Algebraic topology*, Cambridge University Press, 2001.

[16] M. Kahle, *Random geometric complexes*, Discrete & Computational Geometry 45 (2011), no. 3, 553–573.

[17] M. Kahle and E. Meckes, *Limit theorems for Betti numbers of random simplicial complexes*, Homology, Homotopy and Applications 15 (2013), no. 1, 343–374.

[18] ______, *Erratum to Limit theorems for Betti numbers of random simplicial complexes*, Homology, Homotopy and Applications 18 (2016), no. 1, 129–142.

[19] T. Owada, *Limit theorems for betti numbers of extreme sample clouds with application to persistence barcodes*, Ann. Appl. Probab. 28 (Oct. 2018), no. 5, 2814–2854.

[20] ______, *Functional central limit theorem for subgraph counting processes*, Electronic Journal of Probability 22 (2017).

[21] M. Penrose, *Random geometric graphs*, Oxford University Press, 2003.

[22] S. I. Resnick, *Extreme values, regular variation and point processes*, Springer-Verlag, New York, 1987.

[23] ______, *Heavy-tail phenomena: Probabilistic and statistical modeling*, Springer, New York, 2007.
[24] L. Wasserman, *Topological Data Analysis*, Annual Review of Statistics and Its Application 5 (2016), no. 1, 501–532, available at arXiv:stat/1609.08227.

[25] D. Yogeshwaran and R. J. Adler, *On the topology of random complexes built over stationary point processes*, The Annals of Applied Probability 25 (2015), no. 6, 3338–3380.

[26] D. Yogeshwaran, E. Subag, and R. J. Adler, *Random geometric complexes in the thermodynamic regime*, Probability Theory and Related Fields 167 (2017), no. 1-2, 107–142.

[27] A. Zomorodian and G. Carlsson, *Computing persistent homology*, Discrete & Computational Geometry 33 (2005), no. 2, 249–274.

**Department of Statistics, Purdue University, IN, 47907, USA**

E-mail address: owada@purdue.edu, thoma186@purdue.edu