LIMIT THEOREMS FOR BETTI NUMBERS OF EXTREME SAMPLE CLOUDS WITH APPLICATION TO PERSISTENCE BARCODES[1]

BY TAKASHI OWADA

Purdue University

We investigate the topological dynamics of extreme sample clouds generated by a heavy tail distribution on $\mathbb{R}^d$ by establishing various limit theorems for Betti numbers, a basic quantifier of algebraic topology. It then turns out that the growth rate of the Betti numbers and the properties of the limiting processes all depend on the distance of the region of interest from the weak core, that is, the area in which random points are placed sufficiently densely to connect with one another. If the region of interest becomes sufficiently close to the weak core, the limiting process involves a new class of Gaussian processes. We also derive the limit theorems for the sum of bar lengths in the persistence barcode plot, a graphical descriptor of persistent homology.

1. Introduction. The main focus of this paper lies in two areas, random topology and extreme value theory with the aim of revealing topological dynamics of extreme sample clouds far away from the origin, which are generated by heavy tailed distributions on $\mathbb{R}^d$.

For the construction of topological objects of our interest, we start with a point cloud $\mathcal{X} = \{x_1, \ldots, x_n\}$ of points in $\mathbb{R}^d$, from which more complex sets are constructed. Two such examples are the union of balls $\bigcup_{i=1}^n B(x_i; t)$, where $B(x; t)$ is a closed ball of radius t about the point x, and the *Čech complex*, $\check{C}(\mathcal{X}, t)$ (see Figure 1).

DEFINITION 1.1. Let $\mathcal{X}$ be a collection of points in $\mathbb{R}^d$ and t be a positive number. Then the Čech complex $\check{C}(\mathcal{X}, t)$ is defined as follows:

1. The 0-simplices are the points in $\mathcal{X}$.
2. A p-simplex $\sigma = [x_{i_0}, \ldots, x_{i_p}]$ belongs to $\check{C}(\mathcal{X}, t)$ whenever a family of closed balls $\{B(x_{i_j}; t/2), j = 0, \ldots, p\}$ has a nonempty intersection.

In addition to the Čech complex, there are many other *geometric simplicial complexes*, such as the Vietoris–Rips and alpha complexes (see, e.g., [28]). However,
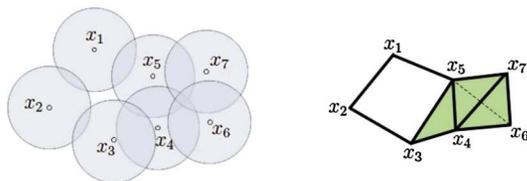
FIG. 1.   *Take* $\mathcal{X} = \{x_1, \ldots, x_7\} \subset \mathbb{R}^2$. *Since three balls with radius* $t/2$ *centered at* x_3, x_4, x_5 *have a common intersection, the 2-simplex* $[x_3, x_4, x_5]$ *belongs to* $\check{C}(\mathcal{X}; t)$. *There also exists a 3-simplex* $[x_4, x_5, x_6, x_7]$, *which adds a tetrahedron on the right figure.*

throughout the current paper, we concentrate on the Čech complex. One reason for doing so is its topological equivalence to the union of balls. An important result, known as the Nerve theorem (see, e.g., Theorem 10.7 of [9]), states that the Čech complex and the union of balls are homotopy equivalent. Furthermore, Čech complexes are regarded as higher-dimensional analogues of *geometric graphs* and, therefore, many of the techniques developed thus far in random geometric graph theory (see, e.g., [41]) are also applicable to random Čech complexes.

   Given a topological space X, the *0th homology group* $H_0(X)$ is an Abelian group and is a topological invariant generated by elements representing connected components in X. For $k \geq 1$, the *kth homology group* $H_k(X)$ is a topological invariant as well, which is generated by elements representing k-dimensional "holes" or "cycles" in X. For $k \geq 0$, the *kth Betti number* $\beta_k(X)$ is a basic quantifier of topology that is central to the study in this paper. Intuitively, $\beta_0(X)$ counts the number of connected components in X, while $\beta_k(X)$, $k \geq 1$, counts the number of k-dimensional holes or cycles in X. For example, as seen in Figure 2, a one-dimensional sphere, that is, a circle, has $\beta_0 = 1$, $\beta_1 = 1$, and $\beta_k = 0$ for all $k \geq 2$. A two-dimensional sphere has $\beta_0 = 1$, $\beta_1 = 0$ and $\beta_2 = 1$, and all others zero. In the case of a two-dimensional torus, the nonzero Betti numbers are $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$. At a more formal level, the kth Betti number $\beta_k(X)$ is defined as the dimension of the kth homology group $H_k(X)$. More rigorous coverage of homology theory can be found in, for example, [29, 46] and [37]. An excellent review [14] contains a gentle introduction of the topological concepts needed in the current paper.

   The study of the geometric and topological properties of extreme sample clouds in a high-dimensional space belongs to Extreme Value Theory (EVT). EVT studies, as its name implies, the extremal behavior of stochastic processes. It is a highly active research area at the intersection of probability theory and statistics; an excellent treatment of the field is in [43] and a more recent exposition is in [18], with other key publications over the years including [15, 24, 26, 35] and [44]. Indeed, over the last decade or so, many studies have provided geometric descriptions of multivariate extremes in view of point process theory, among them [4, 5] and [6]. In particular, Poisson limits of point processes with a U-statistic structure were discussed in [17, 45] and [20], the last two also including a number of stochastic
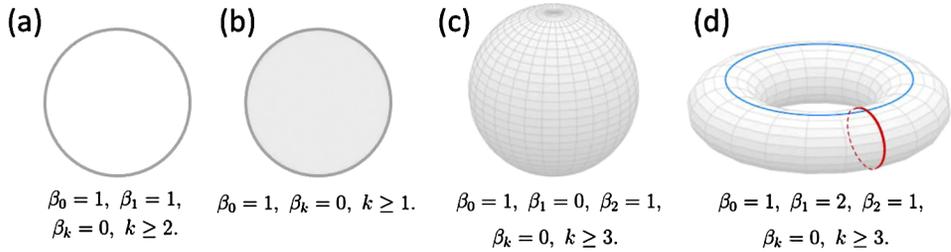
$\beta_0 = 1, \ \beta_1 = 1,$
$\beta_k = 0, \ k \geq 2.$

$\beta_0 = 1, \ \beta_k = 0, \ k \geq 1.$

$\beta_0 = 1, \ \beta_1 = 0, \ \beta_2 = 1,$
$\beta_k = 0, \ k \geq 3.$

$\beta_0 = 1, \ \beta_1 = 2, \ \beta_2 = 1,$
$\beta_k = 0, \ k \geq 3.$

FIG. 2. (a) *One-dimensional sphere*. (b) *One-dimensional disk*. (c) *Two-dimensional sphere*. (d) *Two-dimensional torus. The Betti number β_1 of a two-dimensional sphere is zero; even if one winds a closed loop around the sphere, the loop ultimately vanishes as it moves upward (or downward) along the sphere until the pole. The Betti number β_1 of a two-dimensional torus is 2 because of two independent closed loops (one is red and the other is blue).*

geometry examples. Furthermore, in [40] a recent extensive study of the general point process convergence of extreme sample clouds, leading to limit theorems for Betti numbers of extremes, is reported. The main contribution in [40] is a probabilistic investigation into a layered structure consisting of a collection of "rings" around the origin, with each ring containing extreme random points that exhibit different topological behaviors in terms of the Betti numbers. More formally, this ring-like structure is referred to as *topological crackle*, which was originally reported in [2].

We remark also that there has been increasing interest in the limiting behaviors of random geometric complexes, which are not necessarily related to extremes; see [31, 32, 47, 48] and [12]. In particular, [32] and [48] derived various central and Poisson limit theorems for the Betti numbers of the random Čech complexes $\check{C}(\mathcal{X}_n, r_n)$, with $\mathcal{X}_n$ a random point set in $\mathbb{R}^d$ and r_n a threshold radius. The resulting limit theorems depend heavily on the asymptotics of nr_n^d, as $n \to \infty$. For example, [32] investigated the *sparse regime* (i.e., $nr_n^d \to 0$) so that the spatial distribution of complexes is sparse, and they are observed mostly as isolated components. In contrast, the main focus of [48] was the *thermodynamic regime* [i.e., $nr_n^d \to \xi \in (0, \infty)$] in which complexes are large and highly connected. A nice survey on topology of random geometric complexes is provided by [10]. However, with a few exceptions of [2] and [40], already discussed above, none of these papers has results related to extreme sample clouds. The contribution of the present paper is to develop a fuller description of the ring-like structure and the crackle phenomena by establishing a variety of limit theorems, not only Poissonian-type but also central limit type, for Betti numbers of extremes.

One motivation for studying random geometric complexes comes from topological data analysis (TDA). TDA is a growing research area that broadly refers to the analysis of high-dimensional datasets, the main goal of which is to extract robust topological information from datasets. One of the most typical approaches to TDA is *persistent homology*, which originated in computational topology and appears in

a wide range of applications, including sensor networks [19], bioinformatics [16], computational chemistry [36], manifold learning [38] and linguistics [42].

Persistent homology keeps track of how topological features dynamically evolve in a filtered topological space. We do not give a formal description of persistent homology, but, alternatively, we present an illustrative example, which helps capture its essence. Readers interested in a more rigorous description of persistent homology may refer to [23, 49] and [22], while [1] and [27] provide an elegant review of the topics in an accessible way for nontopologists. Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a set of random points on $\mathbb{R}^d$, drawn from an unknown manifold $\mathcal{M} \subset \mathbb{R}^d$. First, we construct a union of balls

$$U(t) := \bigcup_{i=1}^n B(X_i; t), \qquad t \geq 0,$$

which defines a random filtration generated by balls with increasing radii $t \to \infty$, that is, $U(s) \subset U(t)$ holds for all $0 \leq s \leq t$. By virtue of the Nerve theorem, the filtration $\{U(t), t \geq 0\}$ conveys the same homological information as a collection of Čech complexes $\{\check{C}(\mathcal{X}_n; t), t \geq 0\}$. Utilizing $\{U(t), t \geq 0\}$ or $\{\check{C}(\mathcal{X}_n; t), t \geq 0\}$, we wish to recover the homology of $\mathcal{M}$. We expect that, provided that t is suitably chosen, the union of balls $U(t)$ is homotopy equivalent to $\mathcal{M}$, and hence its homology is the same as $\mathcal{M}$. In general, however, selecting such an appropriate t is not easy at all. To make this more transparent, we consider an example for which $\mathcal{M}$ represents an annulus (Figure 3). In this case, if t is chosen to be too small as in (a), $U(t)$ is homotopy equivalent to many distinct points, implying that we fail to recover the homology of an annulus. On the other hand, if t is extremely large as in (c), then $U(t)$ becomes contractible (i.e., can deform into a single point continuously) and, once again, $U(t)$ does not recover the homology of an annulus.

Persistent homology can extract more robust information of $\mathcal{M}$ by treating a possible range of t simultaneously. Typically, persistent homology can be visualized by two equivalent graphical descriptors known as the *persistence diagram* and
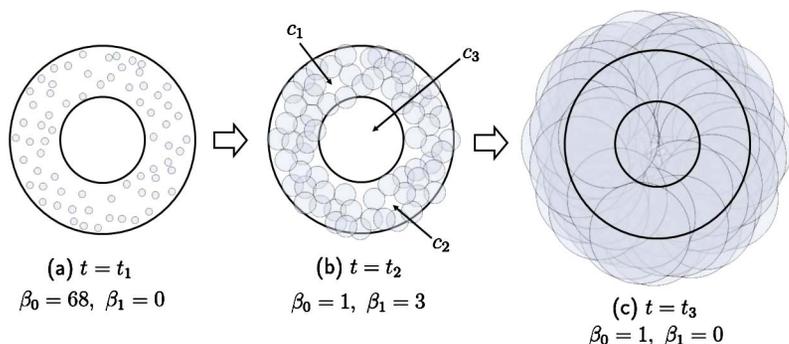

(a) $t = t_1$
$\beta_0 = 68, \ \beta_1 = 0$

(b) $t = t_2$
$\beta_0 = 1, \ \beta_1 = 3$

(c) $t = t_3$
$\beta_0 = 1, \ \beta_1 = 0$

FIG. 3. *Many random points are scattered over an annulus. We increase the radius $t/2$ of the balls about these random points.*
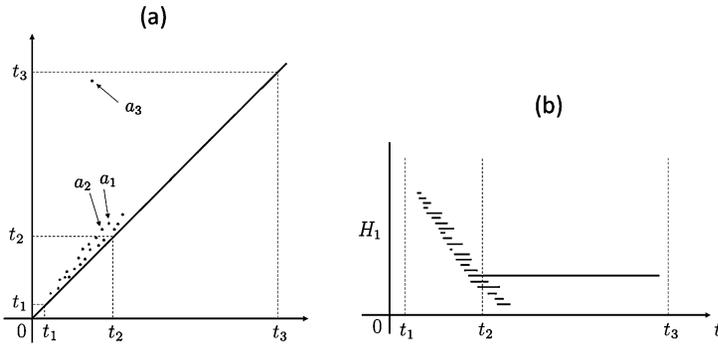
FIG. 4.   (a) *Persistence diagram for the first homology group H_1 represented by one-dimensional holes. In Figure* 3, *there exist two small holes c_1 and c_2 when $t = t_2$. The lifetimes of these holes are so short that they are represented by the points a_1 and a_2 near the diagonal line. On the other hand, c_3 is a robust hole, and thus, the corresponding point a_3 is placed far from the diagonal line.* (b) *Persistence barcode plot for H_1. The vertical line at level t_2 intersects horizontal bars three times, meaning that there are three holes when $t = t_2$. Although two of these quickly vanish, the remaining one has the largest persistence and generates the longest bar.*

persistence barcode plot. The persistence diagram consists of a multiset of points in the plane $\{(b_i, d_i) : i = 1, \ldots, m, 0 \leq b_i < d_i \leq \infty\}$, where each pair (b_i, d_i) describes the birth time and death time of each hole (or connected component). Alternatively, if we represent the pair (b_i, d_i) as an interval $[b_i, d_i]$, we obtain a set of horizontal bars, called the persistence barcode plot.

For the annulus example in Figure 3, as we increase t, many small one-dimensional holes appear and quickly disappear (e.g., the holes c_1 and c_2). Since the birth time and death time of these nonrobust holes are close to each other, they are expressed in the persistence diagram as the points near the diagonal line [see the points a_1 and a_2 in Figure 4(a)]. The points near the diagonal line are usually viewed as "topological noise." In contrast, a robust hole for the annulus denoted by c_3 in Figure 3 has a much longer lifetime than any other small hole and, therefore, it can be represented by the point a_3 placed far above the diagonal line. From the viewpoint of the persistence barcode plot in Figure 4(b), the hole c_3 generates the longest bar, whereas other small holes generate only much shorter bars.

In the spirit of EVT, the main focus of the present paper is the topological features related to heavy-tailed distributions. We define Betti numbers as follows. Given a nonrandom sequence $(R_n, n \geq 1)$ growing to infinity and a nonnegative number t, we denote by $\check{C}(\mathcal{X}_n \cap B(0; R_n)^c, t)$ a Čech complex built over heavy-tailed random points $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ lying outside a growing ball $B(0; R_n)$. We then denote the corresponding kth Betti number by

$$\beta_{k,n}(t) := \beta_k\big(\check{C}(\mathcal{X}_n \cap B(0; R_n)^c, t)\big) = \beta_k\Bigg(\bigcup_{X \in \mathcal{X}_n \cap B(0; R_n)^c} B(X; t/2)\Bigg),$$

where the second equality holds by homotopy equivalence between the Čech complex and the union of balls. Clearly, $\beta_{k,n}(t)$ is viewed as a stochastic process (in the parameter t) possessing right-continuous sample paths with left limits. The main benefit of this scheme is functional level information obtainable about the birth and death of holes of various dimensions. This can be revealed, at least asymptotically, via the limiting process of $\beta_{k,n}(t)$, which will turn out to be certain functionals of Poisson or Gaussian processes.

Once we establish the limit theorems for $\beta_{k,n}(t)$, one fascinating direction would be connecting the asymptotics of Betti numbers to those of statistics concerned with persistent homology. For example, the current paper derives the limit theorems for

$$L_{k,n}(t) = \int_0^t \beta_{k,n}(s)\,ds, \qquad t \geq 0.$$

In the context of persistence barcode plot, $L_{k,n}(t)$ can be regarded as the *lifetime sum* up to parameter t, because, in the kth persistence barcode plot, the Betti number $\beta_{k,n}(s)$ equals the number of times the vertical line at level s intersects the horizontal bars (Figure 5).

The persistent homology originated in algebraic topology, and thus, there are only a limited number of probabilistic and statistical studies that have treated it. The present paper contains some of the earliest results from a pure probabilistic viewpoint, other papers being [11, 30] and [21]. In particular, [11] investigated probabilistic features of bars of the maximum size, from which they tried to capture "extremal" behavior of bars. On the contrary, the present paper tries to uncover the "average" behavior of bars by establishing the limit theorems for the sum of bars, while assuming inhomogeneous Poisson point process and paying more attention on distributional tails. Other references on the interdisciplinary studies between statistics and persistent homology include, for example, [13, 25] and [34].

Before commencing the main body of the paper, we remark that we consider only spherically symmetric distributions. Although the spherical symmetry assumption is far from being crucial, we adopt it to avoid unnecessary technicalities. Furthermore, this paper only treats the case in which the common density
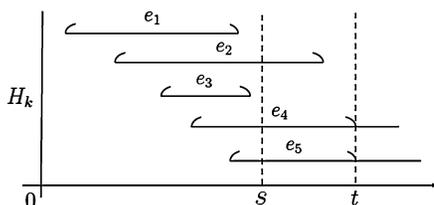

FIG. 5. *kth persistence barcode plot. The lifetime sum up to parameter t is $L_{k,n}(t) = \sum_{i=1}^5 e_i$. The vertical line at level s intersects the horizontal bars three times, implying that $\beta_{k,n}(s) = 3$. The integration of $\beta_{k,n}(s)$ from 0 to t coincides with $\sum_{i=1}^5 e_i$.*

of random points has a regularly varying tail. As is well known in EVT, in the one-dimensional case, regular variation of the tail in the density suffices for the distribution to be in the max-domain of attraction of the Fréchet law. However, results on [39] and [40] indicate that, under proper normalizing constants, all the limit theorems in the present paper can be carried over to the case in which the density has a subexponentially decaying tail. In particular, weak limits will coincide with each other up to constant multiplication.

2. Limit theorems for Betti numbers and the sum of bar lengths. Let $(X_i, i \geq 1)$ be an i.i.d. sequence of $\mathbb{R}^d$-valued random variables with common spherically symmetric density f of a regularly varying tail. Let S_{d-1} be the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$. Assume that for any $\theta \in S_{d-1}$ (equivalently for some $\theta \in S_{d-1}$) and for some $\alpha > d$,

$$(2.1) \qquad \lim_{r \to \infty} f(rt\theta)/f(r\theta) = t^{-\alpha} \qquad \text{for every } t > 0.$$

Denoting by RV_γ the family of regularly varying functions (at infinity) with exponent $\gamma \in \mathbb{R}$, this can be written as $f \in RV_{-\alpha}$. Let N_n be a Poisson random variable with mean n, independent of (X_i), and $\mathcal{P}_n = \{X_1, \ldots, X_{N_n}\}$ denote an inhomogeneous Poisson point process on $\mathbb{R}^d$ with intensity nf.

Given a sequence $(R_n, n \geq 1)$ growing to infinity and a nonnegative number $t \geq 0$, we denote by $\check{C}(\mathcal{P}_n \cap B(0; R_n)^c; t)$ a Čech complex built over random points in $\mathcal{P}_n$ lying outside a growing ball $B(0; R_n)$. Then a family of Čech complexes

$$\{\check{C}(\mathcal{P}_n \cap B(0; R_n)^c; t), t \geq 0\}$$

constitutes a "random filtration" parametrized by $t \geq 0$. That is, we have for all $0 \leq s \leq t$,

$$\check{C}(\mathcal{P}_n \cap B(0; R_n)^c; s) \subset \check{C}(\mathcal{P}_n \cap B(0; R_n)^c; t).$$

Choosing a positive integer $k \geq 1$, which remains fixed hereafter, we denote the kth Betti number of the Čech complex by

$$(2.2) \qquad \beta_{k,n}(t) := \beta_k\big(\check{C}(\mathcal{P}_n \cap B(0; R_n)^c; t)\big) = \beta_k\bigg(\bigcup_{X \in \mathcal{P}_n \cap B(0; R_n)^c} B(X; t/2)\bigg);$$

see Figure 6. We also denote the lifetime sum in the kth persistence barcode plot up to parameter t by

$$(2.3) \qquad L_{k,n}(t) = \int_0^t \beta_{k,n}(s)\, ds, \qquad t \geq 0.$$

The behavior of (2.2) and (2.3) splits into three different regimes, each of which is characterized by the growth rate of R_n:

(i) $n^{k+2} R_n^d f(R_n e_1)^{k+2} \to 1, \qquad n \to \infty,$

(ii) $n^{k+2} R_n^d f(R_n e_1)^{k+2} \to \infty, \qquad nf(R_n e_1) \to 0, n \to \infty,$

(iii) $nf(R_n e_1) \to \lambda, \qquad n \to \infty, \text{ for some } \lambda \in (0, \infty)$
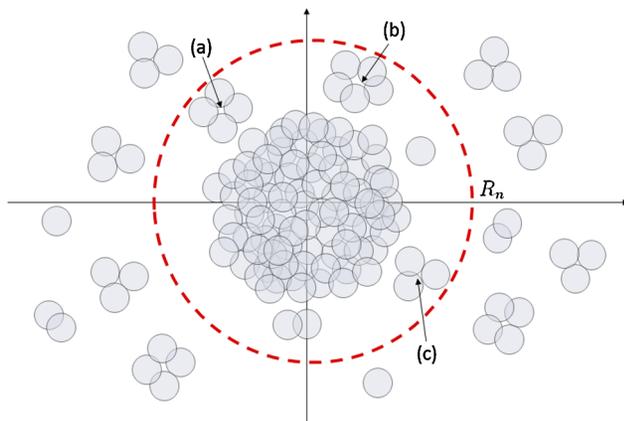
FIG. 6.   *For* $k = 1, d = 2$. *The Betti number* $\beta_{1,n}(t)$ *counts one-dimensional holes outside* $B(0; R_n)$, *while ignoring holes inside the ball* [*e.g.,* (a), (b) *and* (c)].

with $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^d$. Since (R_n) in case (i) grows fastest, the occurrence of k-dimensional holes outside $B(0; R_n)$ is the least likely of the three regimes. In other words, the Čech complex is so sparse that the appearance of holes is a rare event and we only observe "finitely" many holes. Hence, Betti numbers and related lifetime sums are both controlled by a Poisson limit theorem. In contrast, the R_n determined by (ii) grows more slowly than that in (i), so there appear "infinitely" many k-dimensional holes outside $B(0; R_n)$. In this case, the Betti number and its lifetime sum obey a central limit theorem. However, even in case (ii), all k-dimensional holes remaining in the limit will be the simplest one formed by $k + 2$ vertices. The R_n determined by (iii) grows most slowly, implying that the occurrence of k-dimensional holes outside $B(0; R_n)$ is the most likely of the three regimes. Unlike the previous regimes, the resulting Čech complex becomes highly connected and all of the connected components on i vertices for $i = k + 2, k + 3, \ldots$ remain in the limit.

Before proceeding to specific subsections, we need to introduce one important notion.

DEFINITION 2.1.   Let f be a spherically symmetric density on $\mathbb{R}^d$. A *weak core* is a centered ball $B(0; R_n^{(w)})$ such that $nf(R_n^{(w)}e_1) \to 1$ as $n \to \infty$.

Weak cores are balls, centered at the origin with growing radii as n increases, in which random points are placed so densely that the balls with fixed (e.g., unit) radius about these random points become highly connected with one another and form a giant component of a geometric graph. For example, if f has a power-law tail

$$f(x) = C/\big(1 + \|x\|^\alpha\big), \qquad x \in \mathbb{R}^d$$

for some $\alpha > d$ and normalizing constant C ($\|\cdot\|$ denotes a Euclidean norm), then the radius of a weak core is given by $R_n^{(w)} = (Cn)^{1/\alpha}$. The properties of a weak core, together with those of the related notion of a *core*, were carefully explored in [39] for a wide class of distributions; see also [40] and [2]. Note that the R_n determined in (iii) coincides with the radius of a weak core up to multiplicative factors. Since there are essentially no holes inside the weak core, the case in which (R_n) satisfies $nf(R_n e_1) \to \infty$, $n \to \infty$ is expected to lead to the same asymptotic result as that in regime (iii). Moreover, as we can see below, regime (i) generates the same limit up to multiplicative factors, as long as the left-hand side in (i) converges to a positive and finite constant. Therefore, all nontrivial results regarding asymptotics of $\beta_{k,n}(t)$ and $L_{k,n}(t)$ can be completely covered by regimes (i)–(iii).

2.1. *Poissonian limit theorem in the first regime.* First, we assume that (R_n) satisfies condition (i), that is,

$$(2.4) \qquad n^{k+2} R_n^d f(R_n e_1)^{k+2} \to 1, \qquad n \to \infty.$$

It is then elementary to check that (R_n) is a regularly varying sequence (at infinity) with exponent

$$R_n \in RV_{1/(\alpha - d/(k+2))}.$$

Since this exponent depends on k, we write $R_n = R_{k,n}$ whenever it is an asymptotic solution to (2.4). Then the resulting Čech complex lying outside $B(0; R_n)$ is so sparse that there appear at most finitely many k-dimensional holes outside $B(0; R_n)$. Hence, the occurrence of k-dimensional holes outside $B(0; R_n)$ is seen to be "rare," and, consequently, the limiting processes for $\beta_{k,n}(t)$ and $L_{k,n}(t)$ are expressed as a natural functional of a certain Poisson random measure.

To define the limiting process more rigorously, we need some preparation. Let

$$(2.5) \qquad h_t(x_1, \ldots, x_{k+2}) := \mathbf{1}\{\beta_k(\check{C}(x_1, \ldots, x_{k+2}; t)) = 1\}, \qquad x_i \in \mathbb{R}^d.$$

This indicator function can be expressed as the difference between two other indicators:

$$h_t(x_1, \ldots, x_{k+2}) = \mathbf{1}\left\{ \bigcap_{j=1, j \neq j_0}^{k+2} B(x_j; t) \neq \varnothing \text{ for all } j_0 \in \{1, \ldots, k+2\} \right\}$$

$$(2.6) \qquad\qquad - \mathbf{1}\left\{ \bigcap_{j=1}^{k+2} B(x_j; t) \neq \varnothing \right\}$$

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

This decomposition comes from the fact that $h_t(x_1, \ldots, x_{k+2}) = 1$ if and only if $\{x_1, \ldots, x_{k+2}\}$ forms an *empty* $(k + 1)$-*simplex* with respect to t, that is, for

each $j_0 \in \{1, \ldots, k+2\}$, the intersection $\bigcap_{j=1, j \neq j_0}^{k+2} B(x_j; t)$ is nonempty, while $\bigcap_{j=1}^{k+2} B(x_j; t)$ is empty. Note that h_t^+ and h_t^- are nondecreasing functions in t:

$$(2.7) \qquad h_s^{\pm}(x_1, \ldots, x_{k+2}) \leq h_t^{\pm}(x_1, \ldots, x_{k+2})$$

for all $x_1, \ldots, x_{k+2} \in \mathbb{R}^d$ and $0 \leq s \leq t$. Hereafter, we denote $h(x_1, \ldots, x_{k+2}) := h_1(x_1, \ldots, x_{k+2})$ and $h^{\pm}(x_1, \ldots, x_{k+2}) := h_1^{\pm}(x_1, \ldots, x_{k+2})$.

Next, we give a Poissonian structure to the limiting process. Let

$$(2.8) \qquad C_k = \frac{s_{d-1}}{(k+2)!(\alpha(k+2) - d)},$$

where s_{d-1} is a surface area of the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$. Writing Leb for the Lebesgue measure on $(\mathbb{R}^d)^{k+1}$, the *Poisson random measure M_k* with intensity measure C_k Leb is defined by the finite-dimensional distributions

$$\mathbb{P}\{M_k(A) = m\} = e^{-C_k \operatorname{Leb}(A)}(C_k \operatorname{Leb}(A))^m / m!, \qquad m = 0, 1, 2, \ldots$$

for all measurable $A \subset (\mathbb{R}^d)^{k+1}$ with $\operatorname{Leb}(A) < \infty$. Furthermore, if $A_1, \ldots, A_m$ are disjoint subsets in $(\mathbb{R}^d)^{k+1}$, then $M_k(A_1), \ldots, M_k(A_m)$ are independent.

We now state the main result of this subsection, the proof of which is, however, deferred to the Appendix. In the following, $\Rightarrow$ denotes weak convergence in the space considered, for example, $D[0, \infty)$ of right-continuous functions with left limits or $C[0, \infty)$ of continuous functions. In particular, $\overset{\text{fidi}}{\Rightarrow}$ means weak convergence in a finite-dimensional sense.

THEOREM 2.2. *Suppose that $R_n = R_{k,n}$ satisfies* (2.4). *Then*

$$(2.9) \qquad \beta_{k,n}(t) \Rightarrow V_k(t) := \int_{(\mathbb{R}^d)^{k+1}} h_t(0, \mathbf{y}) M_k(d\mathbf{y}) \qquad \text{in } D[0, \infty).$$

Furthermore,

$$(2.10) \qquad L_{k,n}(t) \Rightarrow \int_0^t V_k(s)\, ds \qquad \text{in } C[0, \infty).$$

Recalling the definition of h_t, one may state that the k-dimensional holes contributing to the limit are always formed by connected components on $k+2$ vertices, while other components on more than $k+2$ vertices never appear in the limit. Since there need to be at least $k+2$ vertices to form a single k-dimensional hole, all the k-dimensional holes remaining in the limit are necessarily formed by components of the smallest size.

Because of the decomposition (2.6), we can denote $\mathbf{V}_k = (V_k(t), t \geq 0)$ as

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

$$:= V_k^+(t) - V_k^-(t).$$

The following proposition shows that $\mathbf{V}_k^+$ and $\mathbf{V}_k^-$ can be represented as a time-changed Poisson process.

PROPOSITION 2.3. *The process $\mathbf{V}_k^{\pm}$ is represented in law as*

$$\left(V_k^{\pm}(t), t \ge 0\right) \stackrel{d}{=} \left(N_k^{\pm}\left(t^{d(k+1)}\right), t \ge 0\right),$$

where $N_k^{\pm}$ is a Poisson process with intensity $C_k \int_{(\mathbb{R}^d)^{k+1}} h^{\pm}(0, \mathbf{y}) \, d\mathbf{y}$.

PROOF. It is straightforward to calculate the moment generating function of $(V_k^{\pm}(t_1), \ldots, V_k^{\pm}(t_m))$ for $0 \le t_1 < \cdots < t_m < \infty$. For $\lambda_1, \ldots, \lambda_m \ge 0$, we have
(2.11)

$$\mathbb{E}\left\{\exp\left\{-\sum_{j=1}^{m} \lambda_j V_k^{\pm}(t_j)\right\}\right\} = \exp\left\{-C_k \int_{(\mathbb{R}^d)^{k+1}} \left(1 - e^{-\sum_{j=1}^{m} \lambda_j h_{t_j}^{\pm}(0, \mathbf{y})}\right) d\mathbf{y}\right\}.$$

Exploiting this result, one can easily see that $\mathbf{V}_k^{\pm}$ has independent increments, while for $0 \le s \le t$, $V_k^{\pm}(t) - V_k^{\pm}(s)$ has a Poisson law with mean $C_k \int_{(\mathbb{R}^d)^{k+1}} h^{\pm}(0, \mathbf{y}) \, d\mathbf{y}(t^{d(k+1)} - s^{d(k+1)})$. Now, the claim follows. $\square$

REMARK 2.4. By the moment generating function (2.11), it is easy to see that for each $t \ge 0$, $V_k(t)$ has a Poisson distribution with mean $C_k \int_{(\mathbb{R}^d)^{k+1}} h(0, \mathbf{y}) \, d\mathbf{y} t^{d(k+1)}$. Nevertheless, the process $\mathbf{V}_k$ cannot be represented as a (time-changed) Poisson process, since the sample paths of $\mathbf{V}_k$ allow for both upward and downward jumps.

2.2. *Central limit theorem in the second regime.* In this subsection, we turn to the second regime, which is characterized by

$$(2.12) \qquad n^{k+2} R_n^d f(R_n e_1)^{k+2} \to \infty, \qquad n f(R_n e_1) \to 0, \qquad n \to \infty,$$

for which (R_n) exhibits a slower divergence rate than that in the previous regime. Thus, we expect that, in an asymptotic sense, there appear infinitely many k-dimensional holes outside $B(0; R_n)$, and accordingly, instead of a Poissonian limit theorem, some sort of central limit theorem (CLT) governs the behavior of $\beta_{k,n}(t)$ and $L_{k,n}(t)$.

To formulate the limiting processes, we need some preliminary work. As before, let Leb denote the Lebesgue measure on $(\mathbb{R}^d)^{k+1}$ and C_k a positive constant given in (2.8). Denote by G_k a *Gaussian C_k Leb-noise*, such that

$$G_k(A) \sim N\left(0, C_k \operatorname{Leb}(A)\right)$$

for measurable sets $A \subset (\mathbb{R}^d)^{k+1}$ with $\operatorname{Leb}(A) < \infty$; if $A \cap B = \varnothing$, then $G_k(A)$ and $G_k(B)$ are independent.

We define a Gaussian process $\mathbf{Y}_k = (Y_k(t), t \ge 0)$ by

$$Y_k(t) = \int_{(\mathbb{R}^d)^{k+1}} h_t(0, \mathbf{y}) G_k(d\mathbf{y}), \qquad t \ge 0,$$

where h_t is given in (2.5). This process involves the same indicator function as $\mathbf{V}_k$, which implies that, similar to the last regime, the k-dimensional holes affecting $\mathbf{Y}_k$ must be always formed by connected components on $k + 2$ vertices (i.e., components of the smallest size).

We now state the main limit theorem. The proof is presented in the Appendix.

THEOREM 2.5. *Suppose that (R_n) satisfies (2.12). Then*

$$\left(n^{k+2} R_n^d f(R_n e_1)^{k+2}\right)^{-1/2} \left(\beta_{k,n}(t) - \mathbb{E}\{\beta_{k,n}(t)\}\right) \overset{\text{fidi}}{\Rightarrow} Y_k(t),$$

and

$$\left(n^{k+2} R_n^d f(R_n e_1)^{k+2}\right)^{-1/2} \left(L_{k,n}(t) - \mathbb{E}\{L_{k,n}(t)\}\right) \Rightarrow \int_0^t Y_k(s)\,ds \qquad in\ C[0, \infty).$$

REMARK 2.6. Unlike Theorem 2.2, the weak convergence for $\beta_{k,n}(t)$ is described only in a finite-dimensional sense. We believe that this holds even in the space $D[0, \infty)$; however, since the required tightness will need huge amount of calculation, we have decided not to pursuit that direction.

For further clarification of the structure of $\mathbf{Y}_k$, we express the process as

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

$$:= Y_k^+(t) - Y_k^-(t).$$

We claim that $\mathbf{Y}_k^+$ and $\mathbf{Y}_k^-$ are represented as a time-changed Brownian motion. Note, however, that, although $\mathbf{Y}_k$ is a Gaussian process, it cannot be denoted as a (time-changed) Brownian motion.

PROPOSITION 2.7. *The process $\mathbf{Y}_k^\pm$ can be represented in law as*

$$\left(Y_k^\pm(t), t \geq 0\right) \overset{d}{=} \left(B^\pm(D_k^\pm t^{d(k+1)}), t \geq 0\right),$$

where $B^\pm$ denotes the standard Brownian motion, and

$$D_k^\pm := C_k \int_{(\mathbb{R}^d)^{k+1}} h^\pm(0, \mathbf{y})\,d\mathbf{y}.$$

PROOF. It suffices to prove that the covariance functions on both sides coincide. It follows from (2.7) that for $0 \leq s \leq t$,

$$\mathbb{E}\{Y_k^\pm(t) Y_k^\pm(s)\} = C_k \int_{(\mathbb{R}^d)^{k+1}} h_t^\pm(0, \mathbf{y}) h_s^\pm(0, \mathbf{y})\,d\mathbf{y}$$

$$= s^{d(k+1)} D_k^\pm$$

$$= \mathbb{E}\{B^\pm(D_k^\pm t^{d(k+1)}) B^\pm(D_k^\pm s^{d(k+1)})\}. \qquad \square$$

2.3. *Central limit theorem in the third regime.* Finally, we turn to the third regime in which (R_n) is determined by

(2.13) $nf(R_n e_1) \to \lambda$ as $n \to \infty$

for some $\lambda > 0$. In this case, the formation of k-dimensional holes drastically varies as compared to the previous regimes. If (R_n) satisfies (2.13), then by definition, $B(0; R_n)$ coincides with the weak core (up to multiplicative factors). Therefore, many random points become highly connected to one another in the area sufficiently close to the weak core. As a result, connected components on i vertices for $i = k+2, k+3, \ldots$ can all contribute to the limit in the CLT. This phenomenon was never observed in the previous regimes.

In order to make the notation for defining the limiting process significantly lighter, we introduce several shorthand notation. First, for $x_i \in \mathbb{R}^d$, $i = 1, \ldots, m$, and $r > 0$,

$$\mathcal{B}(x_1, \ldots, x_m; r) := \bigcup_{i=1}^{m} B(x_i; r).$$

For $i \geq k+2$, $j \geq 1$, and $t \geq 0$, we define an indicator $h_t^{(i,j)} : (\mathbb{R}^d)^i \to \{0, 1\}$ by

(2.14) $h_t^{(i,j)}(\mathcal{Y}) := \mathbf{1}\{\beta_k(\check{C}(\mathcal{Y}; t)) = j, \check{C}(\mathcal{Y}; t) \text{ is connected}\}$

for $\mathcal{Y} = (y_1, \ldots, y_i) \in (\mathbb{R}^d)^i$. Clearly, $h_t^{(k+2,1)}$ coincides with the h_t defined in (2.5). In particular, we write $h^{(i,j)}(\mathcal{Y}) := h_1^{(i,j)}(\mathcal{Y})$.

Furthermore, for $i, i' \geq k+2$, $j, j' \geq 1$, and $t, s \geq 0$, define an indicator $h_{t,s}^{(i,j,i',j')} : (\mathbb{R}^d)^{i+i'} \to \{0, 1\}$ by

$$h_{t,s}^{(i,j,i',j')}(\mathcal{Y}, \mathcal{Y}') = h_t^{(i,j)}(\mathcal{Y}) h_s^{(i',j')}(\mathcal{Y}'), \qquad \mathcal{Y} \in (\mathbb{R}^d)^i, \mathcal{Y}' \in (\mathbb{R}^d)^{i'},$$

and, we set, for $i, i' \geq k+2$, $t, s \geq 0$,

$$D^{(i,i')}(t, s) := \{(x_1, \ldots, x_{i+i'}) \in (\mathbb{R}^d)^{i+i'} :$$
$$\mathcal{B}(x_1, \ldots, x_i; t) \cap \mathcal{B}(x_{i+1}, \ldots, x_{i+i'}; s) \neq \varnothing\}.$$

In the special case $t = s$, we denote $D^{(i,i')}(t) := D^{(i,i')}(t, t)$.

Now, we define stochastic processes $\mathbf{Z}_k^{(i,j)} = (Z_k^{(i,j)}(t), t \geq 0)$ for $i \geq k+2$ and $j \geq 1$, which function as the building blocks for the limiting process in the CLT. First, define, for $i, i' \geq k+2$, $j, j' \geq 1$, $t, s \geq 0$ and $\lambda > 0$,

$$\mu_k^{(i,j,j')}(t, s, \lambda) := s_{d-1} \int_1^\infty \rho^{d-1-\alpha i} \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0, \mathbf{y}) h_s^{(i,j')}(0, \mathbf{y})$$

(2.15)
$$\times e^{-\lambda \rho^{-\alpha}(s \vee t)^d \operatorname{vol}(\mathcal{B}(0, \mathbf{y}; 1))} \, d\mathbf{y} \, d\rho,$$

and

$$\xi_k^{(i,j,i',j')}(t,s,\lambda)$$

$$:= s_{d-1} \int_1^\infty \rho^{d-1-\alpha(i+i')} \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0,\mathbf{y})$$

(2.16)
$$\times \big[\big(\mathbf{1}_{D^{(i,i')}(t,s)}(0,\mathbf{y}) - \mathbf{1}_{D^{(i,i')}((t\vee s)/2)}(0,\mathbf{y}) \big)$$

$$\times e^{-\lambda\rho^{-\alpha} \mathrm{vol}(\mathcal{B}(0,y_1,\dots y_{i-1};t)\cup\mathcal{B}(y_i,\dots,y_{i+i'-1};s))} - \mathbf{1}_{D^{(i,i')}(t,s)}(0,\mathbf{y})$$

$$\times e^{-\lambda\rho^{-\alpha}[\mathrm{vol}(\mathcal{B}(0,y_1,\dots y_{i-1};t))+\mathrm{vol}(\mathcal{B}(y_i,\dots y_{i+i'-1};s))]} \big] \, d\mathbf{y} \, d\rho,$$

where $a \vee b = \max\{a,b\}$ for $a,b \in \mathbb{R}$, and $h_t^{(i,j)}(0,\mathbf{y}) = h_t^{(i,j)}(0,y_1,\dots,y_{i-1})$ with $0, y_1, \dots, y_{i-1} \in \mathbb{R}^d$, etc. These functions are used to formulate the covariance functions of $\mathbf{Z}_k^{(i,j)}$'s. More specifically, for $i \geq k+2$ and $j \geq 1$, we define $\mathbf{Z}_k^{(i,j)}$ as a zero-mean Gaussian process with the covariance function given by

(2.17)
$$\mathrm{Cov}\big(Z_k^{(i,j)}(t), Z_k^{(i,j)}(s)\big) = \frac{\lambda^i}{i!} \mu_k^{(i,j,j)}(t,s,\lambda)$$
$$+ \frac{\lambda^{2i}}{(i!)^2} \xi_k^{(i,j,i,j)}(t,s,\lambda), \qquad t, s \geq 0.$$

For every $i \geq k+2$, there exists $j_0 \geq 1$, which depends on i, such that for all $j \geq j_0$ and $t \geq 0$, $h_t^{(i,j)}$ is identically zero, in which case, (2.17) allows us to take $\mathbf{Z}_k^{(i,j)}$ as a zero process, that is, $Z_k^{(i,j)}(t) \equiv 0$ for all $t \geq 0$. For example, $\mathbf{Z}_k^{(k+2,j)}$ is a zero process for all $j \geq 2$, because it is impossible to form multiple holes by $k+2$ vertices.

In addition, we assume that the processes $(\mathbf{Z}_k^{(i,j)}, i \geq k+2, j \geq 1)$ are dependent on each other in such a way that for $i, i' \geq k+2$, $j, j' \geq 1$, and $t, s \geq 0$,

$$\mathrm{Cov}\big(Z_k^{(i,j)}(t), Z_k^{(i',j')}(s)\big) = \frac{\lambda^i}{i!} \mu_k^{(i,j,j')}(t,s,\lambda)\delta_{i,i'} + \frac{\lambda^{i+i'}}{i!i'!} \xi_k^{(i,j,i',j')}(t,s,\lambda),$$

where $\delta_{i,i'}$ is the Kronecker delta.

We now define a zero-mean Gaussian process by

(2.18)
$$Z_k(t) := \sum_{i=k+2}^\infty \sum_{j\geq 1} j Z_k^{(i,j)}(t), \qquad t \geq 0,$$

which appears in the limiting process in the CLT. As shown in the proof of Theorem 2.8 below, the right-hand side of (2.18) almost surely converges for each $t \geq 0$.

We can rewrite $Z_k(t)$ as

$$Z_k(t) = Z_k^{(k+2,1)}(t) + \sum_{i=k+3}^\infty \sum_{j\geq 1} j Z_k^{(i,j)}(t).$$

Since the covariance function of $\mathbf{Z}_k^{(i,j)}$ involves the indicator function $h_{\cdot}^{(i,j)}$, we can consider the process $\mathbf{Z}_k^{(i,j)}$ as representing the connected components that are on i vertices and possess j holes. In particular, the process $\mathbf{Z}_k^{(k+2,1)}$ represents the connected components on $k+2$ vertices with a single k-dimensional hole. This implies that $\mathbf{Z}_k^{(k+2,1)}$ may share the same property as $\mathbf{Y}_k$ in the last regime in the sense that both processes represent connected components only of the smallest size. In the present regime, however, we cannot ignore the effect of larger components emerging near the weak core and, therefore, many other Gaussian processes, except for $\mathbf{Z}_k^{(k+2,1)}$, will contribute to the limit in the CLT.

Before presenting the main limit theorem, we add a technical assumption that a constant λ in (2.13) is less than $(e\omega_d)^{-1}$, where ω_d is the volume of a unit ball in $\mathbb{R}^d$. The CLT below might hold without any upper bound condition for λ, but this is needed for technical reasons during the proof. Accordingly, the domain of functions in the space C must be restricted to the unit interval $[0, 1]$. The proof of the theorem is deferred to the Appendix.

THEOREM 2.8. *Suppose that (R_n) satisfies*

$$(2.19) \qquad nf(R_ne_1) \to \lambda \in \big(0, (e\omega_d)^{-1}\big) \qquad as\ n \to \infty.$$

Then

$$(2.20) \qquad R_n^{-d/2}\big(\beta_{k,n}(t) - \mathbb{E}\{\beta_{k,n}(t)\}\big) \overset{\text{fidi}}{\Rightarrow} Z_k(t),$$

holds in a finite-dimensional sense for $t \in [0, 1]$, and

$$(2.21) \qquad R_n^{-d/2}\big(L_{k,n}(t) - \mathbb{E}\{L_{k,n}(t)\}\big) \Rightarrow \int_0^t Z_k(s)\,ds \qquad in\ C[0, 1].$$

APPENDIX

In the Appendix, we provide the proofs of Theorems 2.2, 2.5 and 2.8. We first introduce the results known as the "Mecke formula" in order to compute the expectations related to Poisson point processes. Indeed, the Mecke formula applies many times hereafter in the Appendix. In Section A.2, we prove Theorem 2.2, and, subsequently, in Section A.3 we verify Theorem 2.8. We give the proof of Theorem 2.5 in Section A.4, while exploiting many of the results established in the former Section A.3.

Before proceeding to specific subsections, we introduce some useful shorthand notation to save space. For $\mathbf{x} = (x_1, \ldots, x_m) \in (\mathbb{R}^d)^m$, $x \in \mathbb{R}^d$, and $\mathbf{y} = (y_1, \ldots, y_{i-1}) \in (\mathbb{R}^d)^{i-1}$,

$$f(\mathbf{x}) := f(x_1) \cdots f(x_m),$$

$$f(x + \mathbf{y}) := f(x + y_1) \cdots f(x + y_{m-1}),$$

$$h^{(i,j)}(0, \mathbf{y}) := h^{(i,j)}(0, y_1, \ldots, y_{i-1}) \qquad \text{etc.}$$

Denote also by C^* a generic positive constant, which can vary between lines and is independent of n.

A.1. Mecke formula for Poisson point processes.

LEMMA A.1 (Mecke formula for Poisson point processes, [3], see also Section 1.7 in [41]). *Let (X_i) be i.i.d. $\mathbb{R}^d$-valued random variables with common density f. Let $\mathcal{P}_n$ be a Poisson point process on $\mathbb{R}^d$ with intensity nf. Let $u(\mathcal{Y}, \mathcal{X})$ and $v(\mathcal{Y}', \mathcal{X})$ be measurable bounded functions defined for $\mathcal{Y} \in (\mathbb{R}^d)^\ell$, $\mathcal{Y}' \in (\mathbb{R}^d)^m$, and a finite subset $\mathcal{X} \supset \mathcal{Y}, \mathcal{Y}'$ of d-dimensional real vectors. Then*

$$\mathbb{E}\left\{\sum_{\mathcal{Y} \subset \mathcal{P}_n} u(\mathcal{Y}, \mathcal{P}_n)\right\} = \frac{n^\ell}{\ell!}\mathbb{E}\{u(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\},$$

where $\mathcal{Y}'$ is a set of ℓ i.i.d. points in $\mathbb{R}^d$ with density f, independent of $\mathcal{P}_n$. Furthermore,

$$\mathbb{E}\left\{\sum_{\mathcal{Y} \subset \mathcal{P}_n}\sum_{\mathcal{Y}' \subset \mathcal{P}_n, |\mathcal{Y} \cap \mathcal{Y}'|=0} u(\mathcal{Y}, \mathcal{P}_n)v(\mathcal{Y}', \mathcal{P}_n)\right\}$$
$$= \frac{n^{\ell+m}}{\ell!m!}\mathbb{E}\{u(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n)v(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)\},$$

where $\mathcal{Y}_1$ is a set of ℓ i.i.d. points in $\mathbb{R}^d$ and $\mathcal{Y}_2$ is a set of m i.i.d. points in $\mathbb{R}^d$, such that $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are independent, and $\mathcal{Y}_{12} := \mathcal{Y}_1 \cup \mathcal{Y}_2$ is independent of $\mathcal{P}_n$, and $|\mathcal{Y}_1 \cap \mathcal{Y}_2| = 0$, that is, there are no common points between $\mathcal{Y}_1$ and $\mathcal{Y}_2$.

Moreover, let $w_i(\mathcal{Y})$, $i = 1, 2$ be measurable bounded functions defined for $\mathcal{Y} \in (\mathbb{R}^d)^p$. Then, for every $q \in \{0, \ldots, p\}$,

$$\mathbb{E}\left\{\sum_{\mathcal{Y} \subset \mathcal{P}_n}\sum_{\mathcal{Y}' \subset \mathcal{P}_n, |\mathcal{Y} \cap \mathcal{Y}'|=q} w_1(\mathcal{Y})w_2(\mathcal{Y}')\right\} = \frac{n^{2p-q}}{q!((p-q)!)^2}\mathbb{E}\{w_1(\mathcal{Y}_1)w_2(\mathcal{Y}_2)\},$$

where $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are sets of p i.i.d. points in $\mathbb{R}^d$ with $|\mathcal{Y}_1 \cap \mathcal{Y}_2| = q$.

A.2. Proof of Theorem 2.2. Since (2.10) immediately follows from (2.9) by the continuous mapping theorem, we may prove only (2.9). The proof of (2.9) is divided into two parts. In the first, we show that

$$(A.1) \quad G_{k,n}(t) := \sum_{\mathcal{Y} \subset \mathcal{P}_n} h_t(\mathcal{Y})\mathbf{1}\{m(\mathcal{Y}) \geq R_n\} \Rightarrow V_k(t) \qquad \text{in } D[0, \infty),$$

where $m(x_1, \ldots, x_{k+2}) = \min_{1 \leq i \leq k+2} \|x_i\|$, $x_i \in \mathbb{R}^d$, and, in the second, we prove that the difference between $G_{k,n}(t)$ and $\beta_{k,n}(t)$ vanishes in probability in the space $D[0, \infty)$.

PROOF.   *Part I* We begin with the finite-dimensional weak convergence of (A.1), for which we need to verify

(A.2)
$$\sum_{j=1}^{m} a_j G_{k,n}(t_j) \Rightarrow \sum_{j=1}^{m} a_j V_k(t_j)$$

for every $a_1, \ldots, a_m \in \mathbb{R}$, $t_1, \ldots, t_m \geq 0$, $m \geq 1$.

Let $\sum_{\ell} \epsilon_{v_\ell}$ denote a Poisson random measure on $\mathbb{R}$ with finite mean measure

$$C_k \int_{(\mathbb{R}^d)^{k+1}} \mathbf{1}\left\{ \sum_{j=1}^{m} a_j h_{t_j}(0, \mathbf{y}) \in \cdot \setminus \{0\} \right\} d\mathbf{y}$$

("ϵ" represents the usual Dirac measure). It is then elementary to verify that

$$\sum_{\ell} v_\ell \overset{d}{=} \sum_{j=1}^{m} a_j V_k(t_j).$$

Writing $M_p(\mathbb{R})$ for the space of point measures on $\mathbb{R}$, (A.2) will be complete, provided that we can show the point process convergence

(A.3) $\xi_n := \displaystyle\sum_{\mathcal{Y} \subset \mathcal{P}_n} \mathbf{1}\left\{ \sum_{j=1}^{m} a_j h_{t_j}(\mathcal{Y}) \neq 0, m(\mathcal{Y}) \geq R_n \right\} \epsilon_{(\sum_{j=1}^{m} a_j h_{t_j}(\mathcal{Y}))}$

$\Rightarrow \displaystyle\sum_{\ell} \epsilon_{v_\ell} \qquad$ in $M_p(\mathbb{R})$.

Indeed, since the functional $\widehat{T} : M_p(\mathbb{R}) \to \mathbb{R}$ defined by $\widehat{T}(\sum_{\ell} \epsilon_{z_\ell}) = \sum_{\ell} z_\ell$ is continuous on a set of *finite* point measures, (A.3) implies (A.2) by the continuous mapping theorem.

According to [20] (or use Theorem 2.1 in [40]), in order to establish (A.3), it suffices to prove the following results: as $n \to \infty$,

(A.4) $\mathbb{E}\{\xi_n(A)\} \to \mathbb{E}\left\{ \displaystyle\sum_{\ell} \epsilon_{v_\ell}(A) \right\} \qquad$ for every measurable $A \subset (\mathbb{R}^d)^{k+1}$,

and

$$r_n := \max_{1 \leq \ell \leq k+1} n^{2k+4-\ell} \mathbb{P}\left\{ \sum_{j=1}^{m} a_j h_{t_j}(X_1, \ldots, X_{k+2}) \neq 0, \right.$$

(A.5) $\displaystyle\sum_{j=1}^{m} a_j h_{t_j}(X_1, \ldots, X_\ell, X_{k+3}, \ldots, X_{2k+4-\ell}) \neq 0,$

$$\left. \|X_i\| \geq R_n, i = 1, \ldots, 2k+4-\ell \right\} \to 0.$$

For the proof of (A.4), it follows from the Mecke formula in Lemma A.1 that

$$\mathbb{E}\{\xi_n(A)\} = \frac{n^{k+2}}{(k+2)!} \int_{(\mathbb{R}^d)^{k+2}} f(\mathbf{x})\mathbf{1}\{m(\mathbf{x}) \geq R_n\}$$

$$\times \mathbf{1}\left\{\sum_{j=1}^{m} a_j h_{t_j}(\mathbf{x}) \in A \setminus \{0\}\right\} d\mathbf{x}.$$

Changing the variables $x_1 \leftrightarrow x$, $x_\ell \leftrightarrow x + y_{\ell-1}$, $\ell = 2, \ldots, k+2$, together with the translation invariance of h_{t_j}'s,

$$\mathbb{E}\{\xi_n(A)\} = \frac{n^{k+2}}{(k+2)!} \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{k+1}} f(x) f(x+\mathbf{y})\mathbf{1}\{m(x, x+\mathbf{y}) \geq R_n\}$$

(A.6)

$$\times \mathbf{1}\left\{\sum_{j=1}^{m} a_j h_{t_j}(0, \mathbf{y}) \in A \setminus \{0\}\right\} d\mathbf{y}\,dx.$$

The polar coordinate transform $x \leftrightarrow (r, \theta)$, followed by an additional change of variable $r \leftrightarrow R_n\rho$, yields

$$\mathbb{E}\{\xi_n(A)\} = \frac{n^{k+2} R_{k,n}^d f(R_{k,n}e_1)^{k+2}}{(k+2)!} \int_{S_{d-1}} J(\theta)\,d\theta \int_1^\infty d\rho \int_{(\mathbb{R}^d)^{k+1}} d\mathbf{y}$$

(A.7)

$$\times \rho^{d-1} \frac{f(R_n\rho e_1)}{f(R_n e_1)} \prod_{\ell=1}^{k+1} \frac{f(R_n\|\rho\theta + y_\ell/R_n\|e_1)}{f(R_n e_1)}\mathbf{1}\{\|\rho\theta + y_\ell/R_n\| \geq 1\}$$

$$\times \mathbf{1}\left\{\sum_{j=1}^{m} a_j h_{t_j}(0, \mathbf{y}) \in A \setminus \{0\}\right\},$$

where S_{d-1} is the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$ and $J(\theta)$ is the usual Jacobian, that is,

$$J(\theta) = \sin^{k-2}(\theta_1)\sin^{k-3}(\theta_2)\cdots\sin(\theta_{k-2}).$$

By the regular variation assumption (2.1) of f, we have that for every $\rho \geq 1$, $\theta \in S_{d-1}$, and $y_1, \ldots, y_{k+1} \in \mathbb{R}^d$,

$$\frac{f(R_n\rho e_1)}{f(R_n e_1)} \prod_{\ell=1}^{k+1} \frac{f(R_n\|\rho\theta + y_\ell/R_n\|e_1)}{f(R_n e_1)} \to \rho^{-\alpha(k+2)}, \qquad n \to \infty.$$

Therefore, supposing the dominated convergence theorem is applicable, we can obtain

$$\mathbb{E}\{\xi_n(A)\} \to C_k \int_{(\mathbb{R}^d)^{k+1}} \mathbf{1}\left\{\sum_{j=1}^{m} a_j h_{t_j}(0, \mathbf{y}) \in A \setminus \{0\}\right\} d\mathbf{y}$$

$$= \mathbb{E}\left\{\sum_\ell \epsilon_{v_\ell}(A)\right\}, \qquad n \to \infty.$$

To establish an integrable upper bound, we use the so-called Potter's bound (e.g., Proposition 2.6(ii) in [44]); for every $0 < \xi < \alpha - d$, we have

$$(A.8) \quad \frac{f(R_n \rho e_1)}{f(R_n e_1)} \mathbf{1}\{\rho \geq 1\} \leq (1 + \xi) \rho^{-\alpha + \xi} \mathbf{1}\{\rho \geq 1\},$$

$$(A.9) \quad \prod_{\ell=1}^{k+1} \frac{f(R_n \| \rho\theta + y_\ell/R_n \| e_1)}{f(R_n e_1)} \mathbf{1}\{\| \rho\theta + y_\ell/R_n \| \geq 1\} \leq (1 + \xi)^{k+1}$$

for sufficiently large n. Since $\int_1^\infty \rho^{d-1-\alpha+\xi}\, d\rho < \infty$, the dominated convergence theorem applies as required.

As for the proof of (A.5), proceeding by changing the variables in the same way as the previous argument, we see that as $n \to \infty$,

$$r_n = \max_{1 \leq \ell \leq k+1} \mathcal{O}\big(n^{2k+4-\ell} R_{k,n}^d f(R_{k,n} e_1)^{2k+4-\ell}\big)$$

$$= \max_{1 \leq \ell \leq k+1} \mathcal{O}\big((nf(R_{k,n} e_1))^{k+2-\ell}\big) \to 0.$$

Now, the claim is proved.

Next, we show the tightness of $(G_{k,n}(t), t \geq 0)$ in the space $D[0, \infty)$ equipped with the Skorohod J_1-topology. By Theorem 13.4 in [8], it suffices to show that for every $L > 0$, there exists $B > 0$ such that

$$\mathbb{P}\{\min\{|G_{k,n}(t) - G_{k,n}(s)|, |G_{k,n}(s) - G_{k,n}(r)|\} \geq \lambda\} \leq B\lambda^{-2}(t-r)^2$$

for all $0 \leq r \leq s \leq t \leq L$, $n \geq 1$, and $\lambda > 0$. For typographical ease, define for $n \geq 1$ and $0 \leq s \leq t$,

$$h_{n,t}(\mathcal{Y}) := h_t(\mathcal{Y})\mathbf{1}\{m(\mathcal{Y}) \geq R_{k,n}\}, \qquad \mathcal{Y} \in (\mathbb{R}^d)^{k+2},$$

$$(A.10) \qquad h_{t,s}(\mathcal{Y}) := h_t(\mathcal{Y}) - h_s(\mathcal{Y}), \qquad \mathcal{Y} \in (\mathbb{R}^d)^{k+2},$$

$$h_{t,s}^{\pm}(\mathcal{Y}) := h_t^{\pm}(\mathcal{Y}) - h_s^{\pm}(\mathcal{Y}), \qquad \mathcal{Y} \in (\mathbb{R}^d)^{k+2}.$$

By Markov's inequality, we only have to show that

$$(A.11) \quad \mathbb{E}\Big\{ \sum_{\mathcal{Y} \subset \mathcal{P}_n} \sum_{\mathcal{Y}' \subset \mathcal{P}_n} |h_{n,t}(\mathcal{Y}) - h_{n,s}(\mathcal{Y})||h_{n,s}(\mathcal{Y}') - h_{n,r}(\mathcal{Y}')| \Big\} \leq B(t-r)^2$$

for all $0 \leq r \leq s \leq t \leq L$ and $n \geq 1$. The left-hand side above is clearly equal to

$$\sum_{\ell=0}^{k+2} \mathbb{E}\Big\{ \sum_{\mathcal{Y} \subset \mathcal{P}_n} \sum_{\mathcal{Y}' \subset \mathcal{P}_n, |\mathcal{Y} \cap \mathcal{Y}'| = \ell} |h_{n,t}(\mathcal{Y}) - h_{n,s}(\mathcal{Y})||h_{n,s}(\mathcal{Y}') - h_{n,r}(\mathcal{Y}')| \Big\}$$

$$:= \sum_{\ell=0}^{k+2} \mathbb{E}\{I_{n,\ell}\}.$$

For $\ell = 1, \ldots, k + 2$, the Mecke formula yields

$$\mathbb{E}\{I_{n,\ell}\} = \frac{n^{2k+4-\ell}}{\ell!((k+2-\ell)!)^2} \mathbb{E}\{|h_{n,t}(\mathcal{Y}_1) - h_{n,s}(\mathcal{Y}_1)||h_{n,s}(\mathcal{Y}_2) - h_{n,r}(\mathcal{Y}_2)|\},$$

where $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are sets of $(k + 2)$ i.i.d. points in $\mathbb{R}^d$ sharing ℓ common points, that is, $|\mathcal{Y}_1 \cap \mathcal{Y}_2| = \ell$. By the same change of variables as in (A.6) and (A.7), together with (2.4) and Potter's bound, we eventually have

$$\mathbb{E}\{I_{n,\ell}\} \leq C^* \int_{(\mathbb{R}^d)^{\ell-1}} d\mathbf{y} \int_{(\mathbb{R}^d)^{k+2-\ell}} d\mathbf{z}_2 \int_{(\mathbb{R}^d)^{k+2-\ell}} d\mathbf{z}_1 |h_{t,s}(0, \mathbf{y}, \mathbf{z}_1)||h_{s,r}(0, \mathbf{y}, \mathbf{z}_2)|$$

$$\leq C^* \int_{(\mathbb{R}^d)^{\ell-1}} d\mathbf{y} \int_{(\mathbb{R}^d)^{k+2-\ell}} d\mathbf{z}_2 \int_{(\mathbb{R}^d)^{k+2-\ell}} d\mathbf{z}_1$$

$$(h_{t,s}^+(0, \mathbf{y}, \mathbf{z}_1) h_{s,r}^+(0, \mathbf{y}, \mathbf{z}_2) + h_{t,s}^-(0, \mathbf{y}, \mathbf{z}_1) h_{s,r}^-(0, \mathbf{y}, \mathbf{z}_2)$$

$$+ h_{t,s}^+(0, \mathbf{y}, \mathbf{z}_1) h_{s,r}^-(0, \mathbf{y}, \mathbf{z}_2) + h_{t,s}^-(0, \mathbf{y}, \mathbf{z}_1) h_{s,r}^+(0, \mathbf{y}, \mathbf{z}_2)).$$

Applying Lemma A.2 below, the rightmost term is bounded by $C^*(t - r)^2$, as required.

We need to establish a suitable upper bound for $\mathbb{E}\{I_{n,0}\}$ as well. By the Mecke formula,

$$\mathbb{E}\{I_{n,0}\} = \frac{n^{2k+4}}{((k+2)!)^2} \mathbb{E}\{|h_{n,t}(\mathcal{Y}) - h_{n,s}(\mathcal{Y})|\} \mathbb{E}\{|h_{n,s}(\mathcal{Y}) - h_{n,r}(\mathcal{Y})|\},$$

and the same argument as above can provide an upper bound of the form $C^*(t - r)^2$. Now, we can conclude (A.11).

Part II To complete the proof, one needs to show that

$$(A.12) \qquad \beta_k\big(\check{C}(\mathcal{P}_n \cap B(0; R_{k,n})^c; t)\big) - G_{k,n}(t) \xrightarrow{P} 0 \qquad \text{in } D[0, \infty).$$

To this end, we use obvious inequalities

$$G_{k,n}(t) \leq \beta_k\big(\check{C}(\mathcal{P}_n \cap B(0; R_{k,n})^c; t)\big) \leq G_{k,n}(t) + L_{k,n}(t),$$

where

$$L_{k,n}(t) = \sum_{\mathcal{Y} \subset \mathcal{P}_n} \mathbf{1}\{|\mathcal{Y}| = k + 3, \check{C}(\mathcal{Y}; t) \text{ is connected}\} \times \mathbf{1}\{m(\mathcal{Y}) \geq R_{k,n}\}$$

with $m(x_1, \ldots, x_{k+3}) = \min_{1 \leq i \leq k+3} \|x_i\|$, $x_i \in \mathbb{R}^d$.

We have, for every $T > 0$,

$$\mathbb{E}\Big\{\sup_{0 \leq t \leq T} \big[\beta_k\big(\check{C}(\mathcal{P}_n \cap B(0; R_{k,n})^c; t)\big) - G_{k,n}(t)\big]\Big\}$$

$$\leq \mathbb{E}\Big\{\sup_{0 \leq t \leq T} L_{k,n}(t)\Big\}$$

$$\leq \frac{n^{k+3}}{(k+3)!} \mathbb{P}\{\check{C}(X_1, \dots, X_{k+3}; T) \text{ is connected,}$$

$$\|X_i\| \geq R_{k,n}, i = 1, \dots, k+3\}.$$

The same change of variables as in (A.6) and (A.7), together with Potter's bound, concludes that the rightmost term above turns out to be

$$\mathcal{O}\left(n^{k+3} R_{k,n}^d f(R_{k,n} e_1)^{k+3}\right) = \mathcal{O}\left(n f(R_{k,n} e_1)\right) \to 0 \qquad \text{as } n \to \infty,$$

thus, (A.12) follows. $\square$

LEMMA A.2. *Let* $h_t, h_t^{\pm} : (\mathbb{R}^d)^{k+2} \to \{0, 1\}$ *be indicator functions given in* (2.5) *and* (2.6), *and recall notation* (A.10). *Fix* $L > 0$. *Then we have, for* $\ell \in \{1, \dots, k+2\}$,

$$\int_{(\mathbb{R}^d)^{\ell-1}} d\mathbf{y} \int_{(\mathbb{R}^d)^{k+2-\ell}} d\mathbf{z}_2 \int_{(\mathbb{R}^d)^{k+2-\ell}} d\mathbf{z}_1 \big(h_{t,s}^+(0, \mathbf{y}, \mathbf{z}_1) h_{s,r}^+(0, \mathbf{y}, \mathbf{z}_2)$$

$$+ h_{t,s}^-(0, \mathbf{y}, \mathbf{z}_1) h_{s,r}^-(0, \mathbf{y}, \mathbf{z}_2) + h_{t,s}^+(0, \mathbf{y}, \mathbf{z}_1) h_{s,r}^-(0, \mathbf{y}, \mathbf{z}_2)$$

$$+ h_{t,s}^-(0, \mathbf{y}, \mathbf{z}_1) h_{s,r}^+(0, \mathbf{y}, \mathbf{z}_2) \big) \leq C^*(t-r)^2$$

for all $0 \leq r \leq s \leq t \leq L$.

PROOF. Let $I_1 + I_2 + I_3 + I_4$ denote the triple integral on the left-hand side. It follows from Lemma 7.1 in [39] that $I_i \leq C^*(t-r)^2$ for $i = 1, 2$. The same argument can yield $I_i \leq C^*(t-r)^2$ for $i = 3, 4$ as well. $\square$

A.3. Proof of Theorem 2.8. The goal of this subsection is to complete the proof of Theorem 2.8. The proof is, however, rather long and, therefore, it is divided into several parts.

First, we define for $i \geq k+2$, $j \geq 1$, $t \geq 0$, and $n \geq 1$,

$$h_{n,t}^{(i,j)}(\mathcal{Y}) := h_t^{(i,j)}(\mathcal{Y}) \mathbf{1}\{m(\mathcal{Y}) \geq R_n\}, \qquad \mathcal{Y} \in (\mathbb{R}^d)^i,$$

where $h_t^{(i,j)}$ is given in (2.14), and

$$m(x_1, \dots, x_m) = \min_{1 \leq \ell \leq m} \|x_\ell\|, \qquad x_1, \dots, x_m \in \mathbb{R}^d, m \geq 1.$$

Next, define for $i \geq k+2$, $j \geq 1$, $t \geq 0$, $\mathcal{Y} \in (\mathbb{R}^d)^i$, and a finite subset of d-dimensional real vectors $\mathcal{Z} \supset \mathcal{Y}$

$$g_t^{(i,j)}(\mathcal{Y}, \mathcal{Z}) := h_t^{(i,j)}(\mathcal{Y}) \mathbf{1}\{\check{C}(\mathcal{Y}; t) \text{ is an isolated component of } \check{C}(\mathcal{Z}; t)\}$$

and

$$g_{n,t}^{(i,j)}(\mathcal{Y}, \mathcal{Z}) := g_t^{(i,j)}(\mathcal{Y}, \mathcal{Z}) \mathbf{1}\{m(\mathcal{Y}) \geq R_n\}.$$

Throughout the proof, we rely on a useful representation for the kth Betti number adopted in [33]

$$\beta_{k,n}(t) = \sum_{i=k+2}^{\infty} \sum_{j\geq 1} j \sum_{\mathcal{Y}\subset\mathcal{P}_n} g_{n,t}^{(i,j)}(\mathcal{Y},\mathcal{P}_n).$$

Let $\text{Ann}(K_1, K_2)$ be an annulus of inner radius K_1 and outer radius K_2. For $x_1, \ldots, x_m \in \mathbb{R}^d$, $m \geq 1$, define $\text{Max}(x_1, \ldots, x_m)$ as the function selecting an element with largest distance from the origin; that is, $\text{Max}(x_1, \ldots, x_m) = x_\ell$ if $\|x_\ell\| = \max_{1\leq j\leq m} \|x_j\|$. If multiple x_j's achieve the maximum, we choose an element with the smallest subscript. The following quantity is associated with the kth Betti number and plays an important role in our proof. For $1 \leq K \leq \infty$,

$$\beta_{k,n}(t; K) := \sum_{i=k+2}^{\infty} \sum_{j\geq 1} j \sum_{\mathcal{Y}\subset\mathcal{P}_n} g_{n,t}^{(i,j)}(\mathcal{Y},\mathcal{P}_n)\mathbf{1}\{\text{Max}(\mathcal{Y}) \in \text{Ann}(R_n, K R_n)\}.$$

Clearly, $\beta_{k,n}(t; \infty) = \beta_{k,n}(t)$. Furthermore, we occasionally need a truncated Betti number

$$(A.13) \qquad \beta_{k,n}^{(M)}(t) := \sum_{i=k+2}^{M} \sum_{j\geq 1} j \sum_{\mathcal{Y}\subset\mathcal{P}_n} g_{n,t}^{(i,j)}(\mathcal{Y},\mathcal{P}_n).$$

Analogously, we can also define $\beta_{k,n}^{(M)}(t; K)$ by the truncation.

We start with revealing the asymptotics of the mean and the covariance of the Betti numbers.

LEMMA A.3. *Suppose that*

$$nf(R_n e_1) \to \lambda \in \left(0, (e\omega_d)^{-1}\right) \qquad \textit{as } n \to \infty.$$

For every $0 \leq t, s \leq 1$ and $1 \leq K \leq \infty$, we have, as $n \to \infty$,

$$R_n^{-d}\mathbb{E}\{\beta_{k,n}(t; K)\} \to \sum_{i=k+2}^{\infty} \sum_{j\geq 1} j\frac{\lambda^i}{i!}\mu_k^{(i,j,j)}(t, t, \lambda; K) \in (0, \infty),$$

$$R_n^{-d}\text{Cov}\{\beta_{k,n}(t; K), \beta_{k,n}(s; K)\}$$

$$\to C_k(t, s; K) := \sum_{i=k+2}^{\infty} \sum_{j,j'\geq 1} jj'\frac{\lambda^i}{i!}\mu_k^{(i,j,j')}(t, s, \lambda; K)$$

$$+ \sum_{i,i'=k+2}^{\infty} \sum_{j,j'\geq 1} jj'\frac{\lambda^{i+i'}}{i!i'!}\xi_k^{(i,j,i',j')}(t, s, \lambda; K) \in (-\infty, \infty)$$

with

$$
\mu_k^{(i,j,j')}(t,s,\lambda;K) := s_{d-1} \int_1^K \rho^{d-1-\alpha i} \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0,\mathbf{y}) h_s^{(i,j')}(0,\mathbf{y})
$$

(A.14)
$$
\times\, e^{-\lambda \rho^{-\alpha}(s\vee t)^d \operatorname{vol}(\mathcal{B}(0,\mathbf{y};1))}\, d\mathbf{y}\, d\rho,
$$

$$
\xi_k^{(i,j,i',j')}(t,s,\lambda;K)
$$
$$
:= s_{d-1} \int_1^K \rho^{d-1-\alpha(i+i')} \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0,\mathbf{y})
$$
$$
\times \Big[\big(\mathbf{1}_{D^{(i,i')}(t,s)}(0,\mathbf{y}) - \mathbf{1}_{D^{(i,i')}((t\vee s)/2)}(0,\mathbf{y})\big)
$$
$$
\times e^{-\lambda\rho^{-\alpha}\operatorname{vol}(\mathcal{B}(0,y_1,...y_{i-1};t)\cup\mathcal{B}(y_i,...,y_{i+i'-1};s))} - \mathbf{1}_{D^{(i,i')}(t,s)}(0,\mathbf{y})
$$
$$
\times e^{-\lambda\rho^{-\alpha}[\operatorname{vol}(\mathcal{B}(0,y_1,...y_{i-1};t))+\operatorname{vol}(\mathcal{B}(y_i,...y_{i+i'-1};s))]}\Big]\, d\mathbf{y}\, d\rho.
$$

In terms of notation (2.15) *and* (2.16), *we have*

$$
\mu_k^{(i,j,j')}(t,s,\lambda) = \mu_k^{(i,j,j')}(t,s,\lambda;\infty),
$$
$$
\xi_k^{(i,j,i',j')}(t,s,\lambda) = \xi_k^{(i,j,i',j')}(t,s,\lambda;\infty).
$$

To prove Lemma A.3, we require the results in Lemmas A.4 and A.5 below, for which we refine the ideas and techniques used in [33] and [32]. Without any loss of generality, we may prove only the case $K = \infty$.

PROOF. By the monotone convergence theorem, together with the Mecke formula in Lemma A.1, we have

$$
R_n^{-d}\mathbb{E}\{\beta_{k,n}(t)\} = \sum_{i=k+2}^{\infty} \sum_{j\geq 1} j R_n^{-d} \frac{n^i}{i!} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\},
$$

where $\mathcal{Y}'$ is a set of i.i.d. points in $\mathbb{R}^d$ with density f, independent of $\mathcal{P}_n$.

It follows from Lemma A.4(i) that

$$
R_n^{-d} n^i \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\} \to \lambda^i \mu_k^{(i,j,j)}(t,t,\lambda).
$$

We need to justify the application of the dominated convergence theorem, for which we apply Lemma A.4(ii), stating that there exists a positive integer $N \in \mathbb{N}_+$ so that for all $i \geq k+2$, $j \geq 1$, and $t \geq 0$,

$$
R_n^{-d} n^i \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\} \leq C^*\big(\lambda(1+\delta)\big)^i \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0,\mathbf{y})\, d\mathbf{y}
$$

for all $n \geq N$, where δ is a positive constant satisfying $\lambda(1+\delta)e\omega_d < 1$.

Appealing to Lemma A.5(i), together with Stirling's formula $i! \geq (i/e)^i$ for sufficiently large i, we have

$$\sum_{i=k+2}^{\infty} \sum_{j \geq 1} j \frac{(\lambda(1+\delta))^i}{i!} \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0, \mathbf{y}) \, d\mathbf{y}$$

$$\leq \sum_{i=k+2}^{\infty} \frac{(\lambda(1+\delta))^i}{i!} \binom{i}{k+2} i^{i-2} (\omega_d)^{i-1}$$

$$\leq C^* \sum_{i=k+2}^{\infty} i^k (\lambda(1+\delta)e\omega_d)^i < \infty.$$

Thus, we can apply the dominated convergence theorem.

Next, we address the computation of the covariance. By the monotone convergence theorem,

$$R_n^{-d} \mathbb{E}\{\beta_{k,n}(t)\beta_{k,n}(s)\}$$

$$= R_n^{-d} \sum_{i=k+2}^{\infty} \sum_{j, j' \geq 1} jj' \mathbb{E}\left\{ \sum_{\mathcal{Y} \subset \mathcal{P}_n} g_{n,t}^{(i,j)}(\mathcal{Y}, \mathcal{P}_n) g_{n,s}^{(i,j')}(\mathcal{Y}, \mathcal{P}_n) \right\}$$

$$+ R_n^{-d} \sum_{i,i'=k+2}^{\infty} \sum_{j, j' \geq 1} jj' \mathbb{E}\left\{ \sum_{\mathcal{Y} \subset \mathcal{P}_n} \sum_{\mathcal{Y}' \subset \mathcal{P}_n, \mathcal{Y}' \neq \mathcal{Y}} g_{n,t}^{(i,j)}(\mathcal{Y}, \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}', \mathcal{P}_n) \right\}$$

$$:= A_n + B_n.$$

The argument similar to that for deriving the limit of $R_n^{-d}\mathbb{E}\{\beta_{k,n}(t)\}$ yields

$$A_n \to \sum_{i=k+2}^{\infty} \sum_{j, j' \geq 1} jj' \frac{\lambda^i}{i!} \mu_k^{(i,j,j')}(t, s, \lambda) \qquad \text{as } n \to \infty.$$

As for B_n, note first that if $\mathcal{Y}$ and $\mathcal{Y}'$ share at least one point,

$$g_{n,t}^{(i,j)}(\mathcal{Y}, \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}', \mathcal{P}_n) = 0.$$

Therefore, it must be that $|\mathcal{Y} \cap \mathcal{Y}'| = 0$ (i.e., no common points exist between $\mathcal{Y}$ and $\mathcal{Y}'$) whenever $\mathcal{Y} \neq \mathcal{Y}'$. It then follows from the Mecke formula that

(A.15)

$$B_n = \sum_{i,i'=k+2}^{\infty} \sum_{j, j' \geq 1} jj' R_n^{-d} \frac{n^{i+i'}}{i!i'!} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)\},$$

where $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are sets of i.i.d. points in $\mathbb{R}^d$ with density f, such that $|\mathcal{Y}_1 \cap \mathcal{Y}_2| = 0$, and $\mathcal{Y}_{12} := \mathcal{Y}_1 \cup \mathcal{Y}_2$ is independent of $\mathcal{P}_n$. Let $\mathcal{P}_n'$ be an independent copy of

$\mathcal{P}_n$, which itself is independent of $\mathcal{Y}_{12}$. Then one more application of the Mecke formula yields

$$R_n^{-d} \mathbb{E}\{\beta_{k,n}(t)\} \mathbb{E}\{\beta_{k,n}(s)\}$$

$$= \sum_{i,i'=k+2} \sum_{j,j' \geq 1} j j' R_n^{-d} \frac{n^{i+i'}}{i! i'!} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}'_n)\}.$$

Combining this with (A.15),

$$B_n - R_n^{-d} \mathbb{E}\{\beta_{k,n}(t)\} \mathbb{E}\{\beta_{k,n}(s)\}$$

$$= \sum_{i,i'=k+2}^{\infty} \sum_{j,j' \geq 1} j j' R_n^{-d} \frac{n^{i+i'}}{i! i'!} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}'_n)\}.$$

By virtue of Lemma A.4(iii), while supposing temporarily that the dominated convergence theorem is applicable, the expression on the right-hand side converges to

$$\sum_{i,i'=k+2}^{\infty} \sum_{j,j' \geq 1} j j' \frac{\lambda^{i+i'}}{i! i'!} \xi_k^{(i,j,i',j')}(t, s, \lambda),$$

and thus, $R_n^{-d} \operatorname{Cov}\{\beta_{k,n}(t), \beta_{k,n}(s)\} \to C_k(t, s; \infty)$, $n \to \infty$ follows, as required.

To establish a summable upper bound, we use Lemma A.4(iv) and Lemma A.5(ii). We have that

$$\sum_{i,i'=k+2}^{\infty} \sum_{j,j' \geq 1} j j' R_n^{-d} \frac{n^{i+i'}}{i! i'!} |\mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}'_n)\}|$$

$$\leq C^* \sum_{i,i'=k+2}^{\infty} \frac{(\lambda(1+\delta))^{i+i'}}{i! i'!} \binom{i}{k+2} \binom{i'}{k+2} i^{i-1} (i')^{i'-1} (\omega_d)^{i+i'-1}$$

$$\leq C^* \left(\sum_{i=k+2}^{\infty} i^{k+1} (\lambda(1+\delta) e \omega_d)^i \right)^2 < \infty.$$

At the last inequality, we used Stirling's formula, that is, $i! \geq (i/e)^i$ for sufficiently large i. □

LEMMA A.4. *Throughout the statements* (i) *and* (ii) *below,* $\mathcal{Y}'$ *denotes a set of i.i.d. points in* $\mathbb{R}^d$ *with density* f, *independent of* $\mathcal{P}_n$:

(i) *For $i \geq k + 2$, $j, j' \geq 1$, and $t, s \geq 0$,*

$$R_n^{-d} n^i \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n) g_{n,s}^{(i,j')}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\} \to \lambda^i \mu_k^{(i,j,j')}(t, s, \lambda),$$

$$n \to \infty.$$

(ii) *There exists a positive integer $N \in \mathbb{N}_+$ such that for all $i \geq k + 2$, $j, j' \geq 1$ and $t, s \geq 0$,*

$$R_n^{-d} n^i \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n) g_{n,s}^{(i,j')}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\}$$

$$\leq C^* (\lambda(1 + \delta))^i \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0, \mathbf{y}) h_s^{(i,j')}(0, \mathbf{y}) \, d\mathbf{y} \qquad \text{for all } n \geq N,$$

where $\delta > 0$ satisfies $\lambda(1 + \delta) e \omega_d < 1$.

Moreover, throughout (iii) *and* (iv) *below, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ denote sets of i.i.d. points in $\mathbb{R}^d$, independent of each other, with density f such that $|\mathcal{Y}_1 \cap \mathcal{Y}_2| = 0$ and $\mathcal{Y}_{12} := \mathcal{Y}_1 \cup \mathcal{Y}_2$ is independent of $\mathcal{P}_n$. Let $\mathcal{P}_n'$ be an independent copy of $\mathcal{P}_n$, which is independent of $\mathcal{Y}_{12}$.*

(iii) *For $i, i' \geq k + 2$, $j, j' \geq 1$, and $t, s \geq 0$,*

$$R_n^{-d} n^{i+i'} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\} \to \lambda^{i+i'} \xi_k^{(i,j,i',j')}(t, s, \lambda)$$

as $n \to \infty$.

(iv) *There exists a positive integer $N \in \mathbb{N}_+$ such that for all $i, i' \geq k + 2$, $j, j' \geq 1$ and $t, s \geq 0$,*

$$\left| R_n^{-d} n^{i+i'} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n) \right.$$

$$\left. - g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\} \right|$$

$$\leq C^* (\lambda(1 + \delta))^{i+i'} \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0, \mathbf{y}) \mathbf{1}_{D^{(i,i')}(t \vee s)}(0, \mathbf{y}) \, d\mathbf{y}$$

for all $n \geq N$, where δ is the same positive constant as in (ii).

PROOF. We start with proving (i). Conditioning on $\mathcal{Y}'$, we have that

$$R_n^{-d} n^i \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n) g_{n,s}^{(i,j')}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\}$$

$$= R_n^{-d} n^i \mathbb{E}\{h_{n,t}^{(i,j)}(\mathcal{Y}') h_{n,s}^{(i,j')}(\mathcal{Y}') \mathbb{P}\{\mathcal{P}_n(\mathcal{B}(\mathcal{Y}'; s \vee t)) = \varnothing | \mathcal{Y}'\}\}$$

$$= R_n^{-d} n^i \int_{(\mathbb{R}^d)^i} f(\mathbf{x}) \mathbf{1}\{m(\mathbf{x}) \geq R_n\} h_t^{(i,j)}(\mathbf{x}) h_s^{(i,j')}(\mathbf{x})$$

$$\times \exp\left\{-n \int_{\mathcal{B}(\mathbf{x}; s \vee t)} f(z) \, dz\right\} d\mathbf{x}.$$

Let J_n denote the last integral. Changing the variables in the same way as in (A.6) and (A.7) yields

$$J_n = \left(nf(R_n e_1)\right)^i \int_{S_{d-1}} J(\theta) \, d\theta \int_1^\infty d\rho \int_{(\mathbb{R}^d)^{i-1}} d\mathbf{y} \rho^{d-1} \frac{f(R_n \rho e_1)}{f(R_n e_1)}$$

$$(A.16) \qquad \times \prod_{\ell=1}^{i-1} \frac{f(R_n \|\rho\theta + y_\ell/R_n\| e_1)}{f(R_n e_1)} \mathbf{1}\{\|\rho\theta + y_\ell/R_n\| \geq 1\}$$

$$\times h_t^{(i,j)}(0,\mathbf{y}) h_s^{(i,j')}(0,\mathbf{y}) \exp\left\{-n \int_{\mathcal{B}(R_n\rho\theta, R_n\rho\theta+\mathbf{y}; s\vee t)} f(z) \, dz\right\},$$

where S_{d-1} is the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$ and $J(\theta)$ is the Jacobian.

By the regular variation assumption (2.1) of f, we have that for every $\rho \geq 1$, $\theta \in S_{d-1}$, and $y_1, \ldots, y_{i-1} \in \mathbb{R}^d$,

$$\frac{f(R_n \rho e_1)}{f(R_n e_1)} \prod_{\ell=1}^{i-1} \frac{f(R_n \|\rho\theta + y_\ell/R_n\| e_1)}{f(R_n e_1)} \to \rho^{-\alpha i}, \qquad n \to \infty.$$

Appealing to Potter's bound as in (A.8) and (A.9), for every $\rho \geq 1$, $\theta \in S_{d-1}$, and $y_1, \ldots, y_{i-1} \in \mathbb{R}^d$,

$$n \int_{\mathcal{B}(R_n\rho\theta, R_n\rho\theta+\mathbf{y}; s\vee t)} f(z) \, dz$$

$$= nf(R_n e_1) \int_{\mathcal{B}(0,\mathbf{y}; s\vee t)} f(R_n \|\rho\theta + z/R_n\| e_1)/f(R_n e_1) \, dz$$

$$\to \lambda\rho^{-\alpha}(s \vee t)^d \operatorname{vol}(\mathcal{B}(0, \mathbf{y}; 1)), \qquad n \to \infty.$$

For an application of the dominated convergence theorem, we employ Potter's bound once again. First, we choose δ, as in the statement of the lemma, so that $\lambda(1+\delta)e\omega_d < 1$, and then, fix $\xi \in (0, \min\{\alpha - d, \delta\})$. Then there exists a positive integer $N_1 \in \mathbb{N}_+$, which is independent of i, such that

$$(A.17) \qquad \frac{f(R_n \rho e_1)}{f(R_n e_1)} \mathbf{1}\{\rho \geq 1\} \leq (1+\xi)\rho^{-\alpha+\xi} \mathbf{1}\{\rho \geq 1\}$$

and

$$(A.18) \qquad \prod_{\ell=1}^{i-1} \frac{f(R_n \|\rho\theta + y_\ell/R_n\| e_1)}{f(R_n e_1)} \mathbf{1}\{\|\rho\theta + y_\ell/R_n\| \geq 1\} \leq (1+\xi)^{i-1}$$

for all $n \geq N_1$. The integrand in (A.16) is now bounded above by $C^*(1 + \xi)^i \rho^{d-1-\alpha+\xi} h_t^{(i,j)}(0,\mathbf{y}) h_s^{(i,j')}(0,\mathbf{y})$, and

$$\int_1^\infty \int_{(\mathbb{R}^d)^{i-1}} \rho^{d-1-\alpha+\xi} h_t^{(i,j)}(0,\mathbf{y}) h_s^{(i,j')}(0,\mathbf{y}) \, d\mathbf{y} \, d\rho < \infty.$$

Therefore, the dominated convergence theorem concludes that $J_n \to \lambda^i \mu^{(i,j,j')} \times (t, s, \lambda)$, $n \to \infty$, as required.

Proof of (ii): Note first that there exists a positive integer $N_2 \in \mathbb{N}_+$ so that

$$(A.19) \qquad (1 + \xi)nf(R_n e_1) \le \lambda(1 + \delta) \qquad \text{for all } n \ge N_2.$$

Because of (A.17) and (A.18), we have, for all $n \ge N := N_1 \vee N_2$,

$$J_n \le \left((1 + \xi)nf(R_n e_1)\right)^i s_{d-1} \int_1^\infty \rho^{d-1-\alpha+\xi} \, d\rho$$

$$\times \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0, \mathbf{y}) h_s^{(i,j')}(0, \mathbf{y}) \, d\mathbf{y}$$

$$\le C^* \left(\lambda(1 + \delta)\right)^i \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0, \mathbf{y}) h_s^{(i,j')}(0, \mathbf{y}) \, d\mathbf{y}.$$

Proof of (iii): First, we write

$$\mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\}$$

$$= \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n)\}$$

$$+ \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n)(g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n) - g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n'))\}$$

$$:= \mathbb{E}\{A_n\} + \mathbb{E}\{B_n\}.$$

Observing that

$$g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) = g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n)\mathbf{1}\{\mathcal{B}(\mathcal{Y}_1; t/2) \cap \mathcal{B}(\mathcal{Y}_2; t/2) = \varnothing\},$$

$$g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n) = g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n)\mathbf{1}\{\mathcal{B}(\mathcal{Y}_1; s/2) \cap \mathcal{B}(\mathcal{Y}_2; s/2) = \varnothing\},$$

one can rewrite $\mathbb{E}\{A_n\}$ as

$$\mathbb{E}\{A_n\} = -\mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n)\mathbf{1}_{D^{(i,i')}((t \vee s)/2)}(\mathcal{Y}_1, \mathcal{Y}_2)\}.$$

Next, we split $\mathbb{E}\{B_n\}$ into two parts:

$$(A.20) \qquad \begin{aligned} \mathbb{E}\{B_n\} &= \mathbb{E}\{\Delta_n \mathbf{1}\{\mathcal{B}(\mathcal{Y}_1; t) \cap \mathcal{B}(\mathcal{Y}_2; s) = \varnothing\}\} \\ &\quad + \mathbb{E}\{\Delta_n \mathbf{1}_{D^{(i,i')}(t,s)}(\mathcal{Y}_1, \mathcal{Y}_2)\}, \end{aligned}$$

where

$$\Delta_n = g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n)(g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n) - g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')).$$

By the spacial independence of the Poisson point process, the first term on the right-hand side of (A.20) equals zero. Rearranging the terms in $\mathbb{E}\{A_n\}$ and $\mathbb{E}\{B_n\}$, we obtain

$$\mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n)g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$
$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n)g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\}$$
$$= \mathbb{E}\{C_n\} - \mathbb{E}\{D_n\},$$

where

$$C_n = g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n)g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n)$$
$$\times (\mathbf{1}_{D^{(i,i')}(t,s)}(\mathcal{Y}_1, \mathcal{Y}_2) - \mathbf{1}_{D^{(i,i')}((t\vee s)/2)}(\mathcal{Y}_1, \mathcal{Y}_2)),$$
$$D_n = g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n)g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\mathbf{1}_{D^{(i,i')}(t,s)}(\mathcal{Y}_1, \mathcal{Y}_2).$$

Conditioning on $\mathcal{Y}_{12}$, we have

$$\mathbb{E}\{C_n\} = \int_{(\mathbb{R}^d)^i} \int_{(\mathbb{R}^d)^{i'}} f(\mathbf{x}_1) f(\mathbf{x}_2) \mathbf{1}\{m(\mathbf{x}_1, \mathbf{x}_2) \geq R_n\} h_t^{(i,j)}(\mathbf{x}_1) h_s^{(i',j')}(\mathbf{x}_2)$$
$$\times (\mathbf{1}_{D^{(i,i')}(t,s)}(\mathbf{x}_1, \mathbf{x}_2) - \mathbf{1}_{D^{(i,i')}((t\vee s)/2)}(\mathbf{x}_1, \mathbf{x}_2))$$
$$\times \exp\left\{-n \int_{\mathcal{B}(\mathbf{x}_1;t) \cup \mathcal{B}(\mathbf{x}_2;s)} f(z)\, dz\right\} d\mathbf{x}_2\, d\mathbf{x}_1.$$

Proceeding as in the proof of (i), while suitably applying Potter's bound, we can obtain, as $n \to \infty$,

$$R_n^{-d} n^{i+i'} \mathbb{E}\{C_n\} \to \lambda^{i+i'} s_{d-1} \int_1^\infty \rho^{d-1-\alpha(i+i')} \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0, \mathbf{y})$$
$$\times (\mathbf{1}_{D^{(i,i')}(t,s)}(0, \mathbf{y}) - \mathbf{1}_{D^{(i,i')}((t\vee s)/2)}(0, \mathbf{y}))$$
$$\times e^{-\lambda \rho^{-\alpha} \operatorname{vol}(\mathcal{B}(0, y_1, \ldots, y_{i-1}; t) \cup \mathcal{B}(y_i, \ldots, y_{i+i'-1}; s))}\, d\mathbf{y}\, d\rho.$$

Similarly, we have

$$R_n^{-d} n^{i+i'} \mathbb{E}\{D_n\}$$
$$\to \lambda^{i+i'} s_{d-1} \int_1^\infty \rho^{d-1-\alpha(i+i')} \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0, \mathbf{y}) \mathbf{1}_{D^{(i,i')}(t,s)}(0, \mathbf{y})$$
$$\times e^{-\lambda \rho^{-\alpha}[\operatorname{vol}(\mathcal{B}(0, y_1, \ldots y_{i-1}; t)) + \operatorname{vol}(\mathcal{B}(y_i, \ldots, y_{i+i'-1}; s))]}\, d\mathbf{y}\, d\rho,$$

and, therefore,

$$R_n^{-d} n^{i+i'} (\mathbb{E}\{C_n\} - \mathbb{E}\{D_n\}) \to \lambda^{i+i'} \xi_k^{(i,j,i',j')}(t, s, \lambda), \qquad n \to \infty.$$

Proof of (iv): Note first that

$$
\begin{aligned}
&\big| R_n^{-d} n^{i+i'} \mathbb{E}\{ g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n) \\
&\quad - g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}'_n) \} \big| \\
&\quad = \big| R_n^{-d} n^{i+i'} (\mathbb{E}\{C_n\} - \mathbb{E}\{D_n\}) \big| \\
&\quad \leq 2 R_n^{-d} n^{i+i'} \mathbb{E}\{ h_{n,t}^{(i,j)}(\mathcal{Y}_1) h_{n,s}^{(i',j')}(\mathcal{Y}_2) \mathbf{1}_{D^{(i,i')}(t \vee s)}(\mathcal{Y}_1, \mathcal{Y}_2) \}.
\end{aligned}
$$

Changing the variables in the same manner as in (i), the last expression above equals

$$
\begin{aligned}
2\big(n f(R_n e_1)\big)^{i+i'} \int_{S_{d-1}} J(\theta)\, d\theta \int_1^\infty d\rho \int_{(\mathbb{R}^d)^{i+i'-1}} d\mathbf{y}\, \rho^{d-1} \frac{f(R_n \rho e_1)}{f(R_n e_1)} \\
\times \prod_{\ell=1}^{i+i'-1} \frac{f(R_n \|\rho\theta + y_\ell/R_n\| e_1)}{f(R_n e_1)} \mathbf{1}\{ \|\rho\theta + y_\ell/R_n\| \geq 1 \} \\
\times h_{t,s}^{(i,j,i',j')}(0, \mathbf{y}) \mathbf{1}_{D^{(i,i')}(t \vee s)}(0, \mathbf{y}).
\end{aligned}
$$

Using the upper bound (A.17) and

$$
\prod_{\ell=1}^{i+i'-1} \frac{f(R_n \|\rho\theta + y_\ell/R_n\| e_1)}{f(R_n e_1)} \mathbf{1}\{ \|\rho\theta + y_\ell/R_n\| \geq 1 \} \leq (1 + \xi)^{i+i'-1},
$$

and applying (A.19), we can complete the proof. $\square$

LEMMA A.5. *Fix a positive constant $L > 0$:*

(i) *For $i \geq k + 2$, $0 \leq t, s \leq L$,*

$$
\sum_{j,j' \geq 1} jj' \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0, \mathbf{y}) h_s^{(i,j')}(0, \mathbf{y})\, d\mathbf{y} \leq \binom{i}{k+2}^2 i^{i-2}(L^d \omega_d)^{i-1},
$$

where ω_d is a volume of the unit ball in $\mathbb{R}^d$.

(ii) *For $i, i' \geq k + 2$, $0 \leq t, s \leq L$,*

$$
\begin{aligned}
&\sum_{j,j' \geq 1} jj' \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0, \mathbf{y}) \mathbf{1}_{D^{(i,i')}(t \vee s)}(0, \mathbf{y})\, d\mathbf{y} \\
&\qquad\qquad \leq 2^d \binom{i}{k+2} \binom{i'}{k+2} i^{i-1}(i')^{i'-1}(L^d \omega_d)^{i+i'-1}.
\end{aligned}
$$

PROOF. We begin with proving (i). Since every connected component built on a set of i points can contribute to the kth Betti number at most $\binom{i}{k+2}$ times, we

have that

$$\sum_{j,j'\geq 1} jj' \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0,\mathbf{y})h_s^{(i,j')}(0,\mathbf{y})\,d\mathbf{y}$$

$$\leq \binom{i}{k+2}^2 \sum_{j,j'\geq 1} \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0,\mathbf{y})h_s^{(i,j')}(0,\mathbf{y})\,d\mathbf{y}$$

$$\leq \binom{i}{k+2}^2 L^{d(i-1)} \int_{(\mathbb{R}^d)^{i-1}} \mathbf{1}\{\check{C}(0,\mathbf{y};1) \text{ is connected}\}\,d\mathbf{y}.$$

It is well known that there exist i^{i-2} spanning trees on a set of i points, and thus,

$$\int_{(\mathbb{R}^d)^{i-1}} \mathbf{1}\{\check{C}(0,\mathbf{y};1) \text{ is connected}\}\,d\mathbf{y} \leq i^{i-2}(\omega_d)^{i-1}.$$

Now, the claim is proved.

Proof of (ii):

$$\sum_{j,j'\geq 1} jj' \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0,\mathbf{y})\mathbf{1}_{D^{(i,i')}(t\vee s)}(0,\mathbf{y})\,d\mathbf{y}$$

$$\leq \binom{i}{k+2}\binom{i'}{k+2} L^{d(i+i'-1)}$$

$$\times \int_{(\mathbb{R}^d)^{i+i'-1}} \mathbf{1}\{\check{C}(0,y_1,\dots,y_{i-1};1) \text{ is connected,}$$

$$\check{C}(y_i,\dots,y_{i+i'-1};1) \text{ is connected}, \check{C}(0,\mathbf{y};2) \text{ is connected}\}\,d\mathbf{y}.$$

If $\check{C}(0,y_1,\dots,y_{i-1};1)$ is connected, there exist i^{i-2} spanning trees constructed from $\{0,y_1,\dots,y_{i-1}\}$. Similarly, there are $(i')^{i'-2}$ spanning trees built on the points $\{y_i,\dots,y_{i+i'-1}\}$ whenever $\check{C}(y_i,\dots,y_{i+i'-1};1)$ is connected. In addition, if $\check{C}(0,\mathbf{y};2)$ is connected, two sets of points $\{0,y_1,\dots,y_{i-1}\}$ and $\{y_i,\dots,y_{i+i'-1}\}$ must be at a distance of at most 2, implying that $\|y_p - y_q\| \leq 2$ for some $p \in \{0,\dots,i-1\}$ and $q \in \{i,\dots,i+i'-1\}$ (take $y_0 \equiv 0$). Therefore,

$$\int_{(\mathbb{R}^d)^{i+i'-1}} \mathbf{1}\{\check{C}(0,y_1,\dots,y_{i-1};1) \text{ is connected,}$$

$$\check{C}(y_i,\dots,y_{i+i'-1};1) \text{ is connected}, \check{C}(0,\mathbf{y};2) \text{ is connected}\}\,d\mathbf{y}$$

$$\leq i^{i-2}(i')^{i'-2}ii'(\omega_d)^{i+i'-2}2^d\omega_d = 2^d i^{i-1}(i')^{i'-1}(\omega_d)^{i+i'-1}. \qquad \square$$

Subsequently, we establish the CLT for the truncated Betti number (A.13) as well as related lifetime sums, for which, as its limit, we need to define a "truncated" limiting Gaussian process. For $M \geq k+2$, we define

$$Z_k^{(M)}(t) := \sum_{i=k+2}^{M} \sum_{j\geq 1} j Z_k^{(i,j)}(t), \qquad t \geq 0.$$

It is worthwhile noting that for these truncated versions, there is no need to restrict the range of λ as in (2.19). Further, we do not need to restrict the domain of functions in the space C.

LEMMA A.6. *Suppose that*

$$nf(R_n e_1) \to \lambda \in (0, \infty), \qquad n \to \infty.$$

Then, for every $M \geq k + 2$,

(A.21) $$R_n^{-d/2}(\beta_{k,n}^{(M)}(t) - \mathbb{E}\{\beta_{k,n}^{(M)}(t)\}) \overset{\text{fidi}}{\Rightarrow} Z_k^{(M)}(t),$$

and

(A.22) $$R_n^{-d/2} \int_0^t (\beta_{k,n}^{(M)}(s) - \mathbb{E}\{\beta_{k,n}^{(M)}(s)\}) \, ds \Rightarrow \int_0^t Z_k^{(M)}(s) \, ds \qquad in \ C[0, \infty).$$

PROOF. The proofs of (A.21) and (A.22) are very similar up to finite-dimensional weak convergence, so we only verity the latter one. Our proof is closely related to that in Theorem 3.9 in [41]. To prove finite-dimensional weak convergence, we apply the Cramér–Wold device, for which we need to establish the central limit theorem for

$$R_n^{-d/2} \sum_{p=1}^m a_p \int_0^{t_p} (\beta_{k,n}^{(M)}(s) - \mathbb{E}\{\beta_{k,n}^{(M)}(s)\}) \, ds$$

for every $a_1, \ldots, a_m \in \mathbb{R}$, $0 \leq t_1 < \cdots < t_m < \infty$, and $m \geq 1$.

We first decompose this term into two parts in the following manner. For $K \geq 1$, we write

$$\sum_{p=1}^m a_p \int_0^{t_p} \beta_{k,n}^{(M)}(s) \, ds = \sum_{p=1}^m a_p \int_0^{t_p} \beta_{k,n}^{(M)}(s; K) \, ds$$

$$+ \sum_{p=1}^m a_p \int_0^{t_p} \sum_{i=k+2}^M \sum_{j \geq 1} j \sum_{\mathcal{Y} \subset \mathcal{P}_n} g_{n,s}^{(i,j)}(\mathcal{Y}, \mathcal{P}_n)$$

$$\times \mathbf{1}\{\text{Max}(\mathcal{Y}) \in \text{Ann}(K R_n, \infty)\} \, ds$$

$$:= T_n^{(M)}(K) + U_n^{(M)}(K).$$

Define

$$\gamma^{(M)}(K) = \sum_{p=1}^m \sum_{q=1}^m a_p a_q \int_0^{t_p} \int_0^{t_q} C_k^{(M)}(u, v; K) \, du \, dv,$$

where $C_k^{(M)}(u, v; K)$ is a truncated version of $C_k(u, v; K)$ given by

$$C_k^{(M)}(u, v; K) := \sum_{i=k+2}^{M} \sum_{j, j' \geq 1} jj' \frac{\lambda^i}{i!} \mu_k^{(i,j,j')}(u, v, \lambda; K)$$

$$+ \sum_{i,i'=k+2}^{M} \sum_{j, j' \geq 1} jj' \frac{\lambda^{i+i'}}{i! i'!} \xi_k^{(i,j,i',j')}(u, v, \lambda; K).$$

Moreover, $\gamma^{(M)} := \lim_{K \to \infty} \gamma^{(M)}(K)$. It then follows from Lemma A.3 that

$$\gamma^{(M)}(K) = \lim_{n \to \infty} R_n^{-d} \operatorname{Var}\{T_n^{(M)}(K)\}.$$

For the required finite-dimensional weak convergence, we need to show that for every $M \geq k + 2$,

$$R_n^{-d/2} \sum_{p=1}^{m} a_p \int_0^{t_p} (\beta_{k,n}^{(M)}(s) - \mathbb{E}\{\beta_{k,n}^{(M)}(s)\}) \, ds \Rightarrow N(0, \gamma^{(M)}), \qquad n \to \infty.$$

By the standard approximation argument given on page 64 in [41], it suffices to show that for every $K \geq 1$,

$$R_n^{-d/2}(T_n^{(M)}(K) - \mathbb{E}\{T_n^{(M)}(K)\}) \Rightarrow N(0, \gamma^{(M)}(K)), \qquad n \to \infty;$$

equivalently, as $n \to \infty$,

$$\frac{T_n^{(M)}(K) - \mathbb{E}\{T_n^{(M)}(K)\}}{\sqrt{\operatorname{Var}\{T_n^{(M)}(K)\}}} \Rightarrow N(0, 1) \qquad \text{for every } K \geq 1.$$

Let $(Q_\ell, \ell \in \mathbb{N})$ be unit cubes covering $\mathbb{R}^d$. Let

$$V_n := \{\ell \in \mathbb{N} : Q_\ell \cap \operatorname{Ann}(R_n, K R_n) \neq \varnothing\}.$$

Then we see that $|V_n| \leq C^* R_n^d$.

Subsequently, we partition $T_n^{(M)}(K)$ as follows:

$$T_n^{(M)}(K) = \sum_{\ell \in V_n} \sum_{p=1}^{m} a_p \int_0^{t_p} \sum_{i=k+2}^{M} \sum_{j \geq 1} j \sum_{\mathcal{Y} \subset \mathcal{P}_n} g_{n,s}^{(i,j)}(\mathcal{Y}, \mathcal{P}_n)$$

$$\times \mathbf{1}\{\operatorname{Max}(\mathcal{Y}) \in \operatorname{Ann}(R_n, K R_n) \cap Q_\ell\} \, ds$$

$$:= \sum_{\ell \in V_n} \xi_{\ell,n}.$$

We define a relation $\sim$ on a vertex set V_n by $\ell \sim \ell'$ if and only if the distance between Q_ℓ and $Q_{\ell'}$ is less than $2M t_m$. In this case, $(V_n, \sim)$ constitutes a *dependency graph*, that is, for any two vertex sets $I_1, I_2 \subset V_n$ with no edges connecting them, $(\xi_{\ell,n}, \ell \in I_1)$ and $(\xi_{\ell',n}, \ell' \in I_2)$ are independent. By virtue of Stein's method

for normal approximation (see Theorem 2.4 in [41]), the proof will be complete, provided that for $p = 3, 4$,

$$R_n^d \max_{\ell \in V_n} \frac{E|\xi_{\ell,n} - \mathbb{E}\{\xi_{\ell,n}\}|^p}{(\text{Var}\{T_n^{(M)}(K)\})^{p/2}} \to 0, \qquad n \to \infty.$$

For $\ell \in V_n$, we denote by $Z_{\ell,n}$ the number of points in $\mathcal{P}_n$ lying in

$$\text{Tube}(Q_\ell; Mt_m) := \Big\{ x \in \mathbb{R}^d : \inf_{y \in Q_\ell} \|x - y\| \leq Mt_m \Big\}.$$

Clearly, $Z_{\ell,n}$ possesses a Poisson law with mean $n \int_{\text{Tube}(Q_\ell; Mt_m)} f(z)\, dz$. Using Potter's bound, we see that $Z_{\ell,n}$ is stochastically dominated by another Poisson random variable with a constant mean C^*.

Observe that

$$|\xi_{\ell,n}| \leq \sum_{p=1}^m |a_p| t_m \sum_{i=k+2}^M \binom{i}{k+2}^2 \binom{Z_{\ell,n}}{i},$$

and, accordingly, we have

$$\max_{\ell \in V_n} \mathbb{E}|\xi_{\ell,n} - \mathbb{E}\{\xi_{\ell,n}\}|^p \leq C^* \qquad \text{for } p = 3, 4.$$

Therefore, for $p = 3, 4$,

$$R_n^d \max_{\ell \in V_n} \frac{E|\xi_{\ell,n} - \mathbb{E}\{\xi_{\ell,n}\}|^p}{(\text{Var}\{T_n^{(M)}(K)\})^{p/2}} \leq C^* R_n^d (R_n^d)^{-p/2} \to 0, \qquad n \to \infty,$$

which completes the proof of the finite-dimensional weak convergence.

Next, we turn to verifying the tightness of

$$X_n(t) := R_n^{-d/2} \int_0^t \big(\beta_{k,n}^{(M)}(s) - \mathbb{E}\{\beta_{k,n}^{(M)}(s)\}\big)\, ds, \qquad t \geq 0,$$

in the space $C[0, \infty)$. According to Theorem 12.3 in [7], we only have to show that, for any $L > 0$, there exists $B > 0$ such that

$$\mathbb{E}\big\{(X_n(T) - X_n(S))^2\big\} \leq B(T - S)^2$$

for all $0 \leq S \leq T \leq L$ and $n \geq 1$.

We see that

$$\mathbb{E}\big\{(X_n(T) - X_n(S))^2\big\}$$

$$= R_n^{-d} \int_S^T \int_S^T \text{Cov}\{\beta_{k,n}^{(M)}(t), \beta_{k,n}^{(M)}(s)\}\, ds\, dt$$

$$= \int_S^T \int_S^T \sum_{i=k+2}^M \sum_{j,j' \geq 1} jj' R_n^{-d} \frac{n^i}{i!}$$

$$\times \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n) g_{n,s}^{(i,j')}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\}\, ds\, dt$$

$$+ \int_S^T \int_S^T \sum_{i,i'=k+2} \sum_{j,j' \geq 1}^M j j' R_n^{-d} \frac{n^{i+i'}}{i! i'!}$$

$$\times \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\}\, ds\, dt$$

($\mathcal{Y}'$ and $\mathcal{P}_n'$ are defined in the statement of Lemma A.4).

Combining Lemma A.4(ii), (iv) and Lemma A.5(i), (ii), the integrands in the last expression can be bounded above by a positive and finite constant, which does not depend on s, t and n. We now conclude that

$$\mathbb{E}\{(X_n(T) - X_n(S))^2\} \leq C^*(T - S)^2,$$

and thus, the tightness follows. $\square$

PROOF OF THEOREM 2.8.   For the proof of (2.21), by Lemma A.6 and Theorem 3.2 in [8], it suffices to verify that for every $\epsilon > 0$,

(A.23)
$$\lim_{M \to \infty} \limsup_{n \to \infty} \mathbb{P}\bigg\{ \sup_{0 \leq t \leq 1} \bigg| \int_0^t (\beta_{k,n}(s) - \beta_{k,n}^{(M)}(s) - \mathbb{E}\{\beta_{k,n}(s) - \beta_{k,n}^{(M)}(s)\})\, ds \bigg| > \epsilon R_n^{d/2} \bigg\} = 0,$$

and

(A.24)
$$\lim_{M \to \infty} \mathbb{P}\bigg\{ \sup_{0 \leq t \leq 1} \bigg| \int_0^t (Z_k(s) - Z_k^{(M)}(s))\, ds \bigg| > \epsilon \bigg\} = 0.$$

By Chebyshev's inequality, (A.23) immediately follows, provided that

$$\lim_{M \to \infty} \limsup_{n \to \infty} R_n^{-d} \mathbb{E}\bigg\{ \bigg(\int_0^1 |\beta_{k,n}(s) - \beta_{k,n}^{(M)}(s) - \mathbb{E}\{\beta_{k,n}(s) - \beta_{k,n}^{(M)}(s)\}|\, ds \bigg)^2 \bigg\}$$
$$= 0.$$

By the Cauchy–Schwarz inequality, we only have to show that

$$\lim_{M \to \infty} \limsup_{n \to \infty} \int_0^1 \big(R_n^{-d} \operatorname{Var}\{\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t)\} \big)^{1/2}\, dt = 0.$$

One can decompose the integrand as follows:

$$R_n^{-d} \operatorname{Var}\{\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t)\}$$
$$= \sum_{i=M+1}^{\infty} \sum_{j \geq 1} j^2 R_n^{-d} \frac{n^i}{i!} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\}$$

$$+ \sum_{i,i'=M+1}^{\infty} \sum_{j,j' \geq 1} jj' R_n^{-d} \frac{n^{i+i'}}{i!i'!}$$

$$\times \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,t}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,t}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\}$$

($\mathcal{Y}'$ and $\mathcal{P}_n'$ are defined in the statement of Lemma A.4).

Combining Lemma A.4(ii), (iv) and Lemma A.5(i), (ii) proves that this is bounded by

$$C^* \sum_{i=M+1}^{\infty} \frac{(\lambda(1+\delta))^i}{i!} \binom{i}{k+2}^2 i^{i-2} (\omega_d)^{i-1}$$

$$+ C^* \sum_{i,i'=M+1}^{\infty} \frac{(\lambda(1+\delta))^{i+i'}}{i!i'!} \binom{i}{k+2} \binom{i'}{k+2} i^{i-1} (i')^{i'-1} (\omega_d)^{i+i'-1}$$

$$\leq C^* \sum_{i=M+1}^{\infty} i^{2k+2} (\lambda(1+\delta)e\omega_d)^i + C^* \left(\sum_{i=M+1}^{\infty} i^{k+1} (\lambda(1+\delta)e\omega_d)^i \right)^2.$$

Since $0 < \lambda(1+\delta)e\omega_d < 1$, the claim has been proved. Since the proof of (A.24) is almost the same as that of (A.23), we omit it.

A similar (or even easier) argument completes (2.20) as well. $\square$

A.4. Proof of Theorem 2.5. The proof of Theorem 2.5 somewhat parallels that of Theorem 2.8, for which we need to recall the notation of several indicator functions and variants of the Betti numbers defined at the beginning of Section A.3. As in Lemma A.3, we begin with computing the asymptotic mean and covariance of the scaled kth Betti numbers. In the following, let $\rho_n := n^{k+2} R_n^d f(R_n e_1)^{k+2}$.

LEMMA A.7. *For every $t, s \leq 0$ and $1 \leq K \leq \infty$, we have, as $n \to \infty$,*

$$\rho_n^{-1} \mathbb{E}\{\beta_{k,n}(t; K)\} \to \mu_k^{(k+2,1,1)}(t, t, 0; K)/(k+2)! \in (0, \infty),$$

and

$$\rho_n^{-1} \text{Cov}\{\beta_{k,n}(t; K), \beta_{k,n}(s; K)\} \to \mu_k^{(k+2,1,1)}(t, s, 0; K)/(k+2)! \in (0, \infty),$$

where the definition of the limit is given in (A.14).

Recall that, in the last subsection, Lemmas A.4 and A.5 play a crucial role in proving Lemma A.3. In the present subsection, however, one needs to replace Lemma A.4 with Lemma A.8 below in order to show Lemma A.7. Since the proof of Lemma A.8 is analogous to that of Lemma A.4, we omit the proof.

LEMMA A.8. *Throughout the statements* (i) *and* (ii) *below,* $\mathcal{Y}'$ *denotes a set of i.i.d. points in* $\mathbb{R}^d$ *with density* f, *independent of* $\mathcal{P}_n$:

(i) *For* $t, s \geq 0$, *we have, as* $n \to \infty$,

$$\rho_n^{-1} n^{k+2} \mathbb{E}\{g_{n,t}^{(k+2,1)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n) g_{n,s}^{(k+2,1)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\} \to \mu_k^{(k+2,1,1)}(t, s, 0).$$

(ii) *There exists a positive integer* $N \in \mathbb{N}_+$ *such that for all* $i \geq k+2$, $j, j' \geq 1$, *and* $t, s \geq 0$,

$$\rho_n^{-1} n^i \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n) g_{n,s}^{(i,j')}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\}$$

$$\leq C^* \big(2nf(R_n e_1)\big)^{i-(k+2)} \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0, \mathbf{y}) h_s^{(i,j')}(0, \mathbf{y}) \, d\mathbf{y}$$

for all $n \geq N$.

Moreover, $\mathcal{Y}_1$ *and* $\mathcal{Y}_2$ *denote sets of i.i.d. points in* $\mathbb{R}^d$, *independent of each other, with density* f *such that* $|\mathcal{Y}_1 \cap \mathcal{Y}_2| = 0$ *and* $\mathcal{Y}_{12} := \mathcal{Y}_1 \cup \mathcal{Y}_2$ *is independent of* $\mathcal{P}_n$. *Let* $\mathcal{P}_n'$ *be an independent copy of* $\mathcal{P}_n$, *which is independent of* $\mathcal{Y}_{12}$.

(iii) *There exists a positive integer* $N \in \mathbb{N}_+$ *such that for all* $i, i' \geq k+2$, $j, j' \geq 1$, *and* $t, s \geq 0$,

$$\big|\rho_n^{-1} n^{i+i'} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\}\big|$$

$$\leq C^* \big(2nf(R_n e_1)\big)^{i+i'-(k+2)} \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0, \mathbf{y}) \mathbf{1}_{D^{(i,i')}(t \vee s)}(0, \mathbf{y}) \, d\mathbf{y}$$

for all $n \geq N$.

PROOF OF LEMMA A.7. As in the proof of Lemma A.3, we may prove only the case $K = \infty$. Moreover, we compute only the limit of scaled covariance by ρ_n. Proceeding as in the proof of Lemma A.3, one can write

$$\rho_n^{-1} \mathrm{Cov}\{\beta_{k,n}(t), \beta_{k,n}(s)\}$$

$$= \sum_{i=k+2} \sum_{j,j' \geq 1}^{\infty} jj' \rho_n^{-1} \frac{n^i}{i!} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n) g_{n,s}^{(i,j')}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\}$$

$$+ \sum_{i,i'=k+2} \sum_{j,j' \geq 1}^{\infty} jj' \rho_n^{-1} \frac{n^{i+i'}}{i! i'!} \mathbb{E}\{g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)$$

$$- g_{n,t}^{(i,j)}(\mathcal{Y}_1, \mathcal{Y}_1 \cup \mathcal{P}_n) g_{n,s}^{(i',j')}(\mathcal{Y}_2, \mathcal{Y}_2 \cup \mathcal{P}_n')\}.$$

By Lemma A.8(i)–(iii), it now suffices to show that, as $n \to \infty$,

$$A_n := \sum_{i=k+3} \sum_{j,j' \geq 1}^{\infty} jj' \frac{(2nf(R_n e_1))^{i-(k+2)}}{i!} \int_{(\mathbb{R}^d)^{i-1}} h_t^{(i,j)}(0, \mathbf{y}) h_s^{(i,j')}(0, \mathbf{y}) \, d\mathbf{y}$$

$$\to 0,$$

and

$$B_n := \sum_{i,i'=k+2} \sum_{j,j' \geq 1} jj' \frac{(2nf(R_n e_1))^{i+i'-(k+2)}}{i!i'!}$$

$$\times \int_{(\mathbb{R}^d)^{i+i'-1}} h_{t,s}^{(i,j,i',j')}(0,\mathbf{y}) \mathbf{1}_{D^{(i,i')}(t \vee s)}(0,\mathbf{y}) \, d\mathbf{y} \to 0.$$

It follows from Lemma A.5(i) that

$$A_n \leq \sum_{i=k+3}^{\infty} \frac{(2nf(R_n e_1))^{i-(k+2)}}{i!} \binom{i}{k+2}^2 i^{i-2} ((t \vee s)^d \omega_d)^{i-1}$$

$$\leq C^* \sum_{i=k+3}^{\infty} i^{2k+2} (2nf(R_n e_1)(t \vee s)^d e \omega_d)^{i-(k+2)}$$

$$\to 0 \qquad \text{as } n \to \infty,$$

where the last convergence is obtained by $nf(R_n e_1) \to 0$, $n \to \infty$.

Similarly, by Lemma A.5(ii),

$$B_n \leq \sum_{i,i'=k+2}^{\infty} \frac{(2nf(R_n e_1))^{i+i'-(k+2)}}{i!i'!}$$

$$\times 2^d \binom{i}{k+2} \binom{i'}{k+2} i^{i-1} (i')^{i'-1} ((t \vee s)^d \omega_d)^{i+i'-1}$$

$$\leq C^* \sum_{i,i'=k+2}^{\infty} i^{k+1} (i')^{k+1} (2nf(R_n e_1)(t \vee s)^d e \omega_d)^{i+i'-(k+2)}$$

$$\to 0 \qquad \text{as } n \to \infty. \qquad \square$$

The next lemma claims the CLT for the truncated kth Betti number (A.13) and its integral process. The proof is almost the same as that of Lemma A.6 and, therefore, we do not state it here. It is then straightforward to complete the proof of Theorem 2.5 by combining Lemma A.9 and Theorem 3.2 in [8], as in the last subsection.

LEMMA A.9. *For every $M \geq k+2$, we have, as $n \to \infty$,*

$$\rho_n^{-1/2} \big(\beta_{k,n}^{(M)}(t) - \mathbb{E}\{\beta_{k,n}^{(M)}(t)\} \big) \overset{\text{fidi}}{\Rightarrow} Y_k(t),$$

and

$$\rho_n^{-1/2} \int_0^t \big(\beta_{k,n}^{(M)}(s) - \mathbb{E}\{\beta_{k,n}^{(M)}(s)\} \big) \, ds \Rightarrow \int_0^t Y_k(s) \, ds \qquad \text{in } C[0,\infty).$$

REFERENCES

[1] ADLER, R. J., BOBROWSKI, O., BORMAN, M. S., SUBAG, E. and WEINBERGER, S. (2010). Persistent homology for random fields and complexes. In *Borrowing Strength*: *Theory Powering Applications—a Festschrift for Lawrence D. Brown*. *Inst. Math. Stat.* (*IMS*) *Collect.* **6** 124–143. IMS, Beachwood, OH. MR2798515

[2] ADLER, R. J., BOBROWSKI, O. and WEINBERGER, S. (2014). Crackle: The homology of noise. *Discrete Comput. Geom.* **52** 680–704. MR3279544

[3] ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. MR0972770

[4] BALKEMA, G. and EMBRECHTS, P. (2007). *High Risk Scenarios and Extremes*: *A Geometric Approach*. European Mathematical Society (EMS), Zürich. MR2372552

[5] BALKEMA, G., EMBRECHTS, P. and NOLDE, N. (2010). Meta densities and the shape of their sample clouds. *J. Multivariate Anal.* **101** 1738–1754. MR2610743

[6] BALKEMA, G., EMBRECHTS, P. and NOLDE, N. (2013). The shape of asymptotic dependence. In *Prokhorov and Contemporary Probability Theory*. *Springer Proc. Math. Stat.* **33** 43–67. Springer, Heidelberg. MR3070466

[7] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[8] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[9] BJÖRNER, A. (1995). Topological methods. In *Handbook of Combinatorics*, *Vols* 1, 2 1819–1872. Elsevier, Amsterdam. MR1373690

[10] BOBROWSKI, O. and KAHLE, M. (2014). Topology of random simplicial complexes: A survey. In *Algebraic Topology*: *Applications and New Directions*. *Contemp. Math.* **620** 201–221. Amer. Math. Soc., Providence, RI. MR3290093

[11] BOBROWSKI, O., KAHLE, M. and SKRABA, P. (2017). Maximally persistent cycles in random geometric complexes. *Ann. Appl. Probab.* **27** 2032–2060. MR3693519

[12] BOBROWSKI, O. and MUKHERJEE, S. (2015). The topology of probability distributions on manifolds. *Probab. Theory Related Fields* **161** 651–686. MR3334278

[13] BUBENIK, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16** 77–102. MR3317230

[14] CARLSSON, G. (2014). Topological pattern recognition for point cloud data. *Acta Numer.* **23** 289–368. MR3202240

[15] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. MR1932132

[16] DABAGHIAN, Y., MEMOLI, F., FRANK, L. and CARLSSON, G. (2012). A topological paradigm for hippocampal spatial map formation using persistent homology. *PLoS Comput. Biol.* **8** e1002581.

[17] DABROWSKI, A. R., DEHLING, H. G., MIKOSCH, T. and SHARIPOV, O. (2002). Poisson limits for U-statistics. *Stochastic Process. Appl.* **99** 137–157. MR1894255

[18] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An Introduction*. Springer, New York. MR2234156

[19] DE SILVA, V. and GHRIST, R. (2007). Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.* **7** 339–358. MR2308949

[20] DECREUSEFOND, L., SCHULTE, M. and THÄLE, C. (2016). Functional Poisson approximation in Kantorovich–Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Probab.* **44** 2147–2197. MR3502603

[21] DUY, T. K., HIRAOKA, Y. and SHIRAI, T. (2016). Limit theorems for persistence diagrams. Available at arXiv:1612.08371.

[22] EDELSBRUNNER, H. and HARER, J. L. (2010). *Computational Topology*: *An Introduction*. Amer. Math. Soc., Providence, RI. MR2572029

[23] EDELSBRUNNER, H., LETSCHER, D. and ZOMORODIAN, A. (2002). Topological persistence and simplification. *Discrete Comput. Geom.* **28** 511–533. MR1949898

[24] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events*: *For Insurance and Finance*. *Applications of Mathematics* (*New York*) **33**. Springer, Berlin. MR1458613

[25] FASY, B. T., LECCI, F., RINALDO, A., WASSERMAN, L., BALAKRISHNAN, S. and SINGH, A. (2014). Confidence sets for persistence diagrams. *Ann. Statist.* **42** 2301–2339. MR3269981

[26] GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed. Krieger, Melbourne, FL. MR0936631

[27] GHRIST, R. (2008). Barcodes: The persistent topology of data. *Bull. Amer. Math. Soc.* (*N.S.*) **45** 61–75. MR2358377

[28] GHRIST, R. (2014). *Elementary Applied Topology*, Createspace.

[29] HATCHER, A. (2002). *Algebraic Topology*. Cambridge Univ. Press, Cambridge. MR1867354

[30] HIRAOKA, Y. and SHIRAI, T. (2017). Minimum spanning acycle and lifetime of persistent homology in the Linial–Meshulam process. *Random Structures Algorithms* **51** 315–340. MR3683365

[31] KAHLE, M. (2011). Random geometric complexes. *Discrete Comput. Geom.* **45** 553–573. MR2770552

[32] KAHLE, M. and MECKES, E. (2013). Limit theorems for Betti numbers of random simplicial complexes. *Homology, Homotopy Appl.* **15** 343–374. MR3079211

[33] KAHLE, M. and MECKES, E. (2016). Erratum to "Limit theorems for Betti numbers of random simplicial complexes." *Homology, Homotopy Appl.* **18** 129–142. MR3485340

[34] KUSANO, G., FUKUMIZU, K. and HIRAOKA, Y. (2016). Persistence weighted Gaussian kernel for topological data analysis. In *ICML'16 Proceedings of the 33rd International Conference on International Conference on Machine Learning*, Vol. **48**, pp. 2004–2013.

[35] LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York–Berlin. MR0691492

[36] MARTIN, S., THOMPSON, A., COUTSIAS, E. A. and WATSON, J. (2010). Topology of cyclo-octane energy landscape. *J. Chem. Phys.* **132** 234115.

[37] MUNKRES, J. R. (1984). *Elements of Algebraic Topology*, 1st ed. Benjamin-Cummings, San Francisco, CA.

[38] NIYOGI, P., SMALE, S. and WEINBERGER, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419–441. MR2383768

[39] OWADA, T. (2017). Functional central limit theorem for subgraph counting processes. *Electron. J. Probab.* **22** Paper No. 17, 38. MR3622887

[40] OWADA, T. and ADLER, R. J. (2017). Limit theorems for point processes under geometric constraints (and topological crackle). *Ann. Probab.* **45** 2004–2055. MR3650420

[41] PENROSE, M. (2003). *Random Geometric Graphs. Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198

[42] PORT, A., GHEORGHITA, I., GUTH, D., CLARK, J. M., LIANG, C., DASU, S. and MARCOLLI, M. (2015). Persistent topology of syntax. Available at arXiv:1507.05134.

[43] RESNICK, S. I. (1987). *Extreme Values*, *Regular Variation*, *and Point Processes*. *Applied Probability. A Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810

[44] RESNICK, S. I. (2007). *Heavy-Tail Phenomena*: *Probabilistic and Statistical Modeling*. Springer, New York. MR2271424

[45] SCHULTE, M. and THÄLE, C. (2012). The scaling limit of Poisson-driven order statistics with applications in geometric probability. *Stochastic Process*. *Appl*. **122** 4096–4120. MR2971726

[46] VICK, J. W. (1994). *Homology Theory*: *An Introduction to Algebraic Topology*, 2nd ed. *Graduate Texts in Mathematics* **145**. Springer, New York. MR1254439

[47] YOGESHWARAN, D. and ADLER, R. J. (2015). On the topology of random complexes built over stationary point processes. *Ann*. *Appl*. *Probab.* **25** 3338–3380. MR3404638

[48] YOGESHWARAN, D., SUBAG, E. and ADLER, R. J. (2017). Random geometric complexes in the thermodynamic regime. *Probab. Theory Related Fields* **167** 107–142. MR3602843

[49] ZOMORODIAN, A. and CARLSSON, G. (2005). Computing persistent homology. *Discrete Comput*. *Geom*. **33** 249–274. MR2121296

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907
USA
E-MAIL: owada@purdue.edu