STOCHASTIC PROCESS LIMITS FOR TOPOLOGICAL FUNCTIONALS OF GEOMETRIC COMPLEXES

by

Andrew M. Thomas

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Statistics

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Takashi Owada, Chair

Department of Statistics

Dr. Anirban DasGupta

Department of Statistics

Dr. Raghu Pasupathy

Department of Statistics

Dr. Qifan Song

Department of Statistics

Dr. Mark Daniel Ward

Department of Statistics

Approved by:

Dr. Jun Xie

ACKNOWLEDGMENTS

How does one put into a few paragraphs the immensity of gratitude which I feel to those that have helped me along my way? No matter how much I would like to admit it, no man is an island, entire of itself. With respect to my studies, I can thank nobody more highly than Professor Takashi Owada. He has never failed to steer me in the right direction, and I cannot express how fortunate I have been to have him provide support and assistance to me along my mathematical and personal journey.

Coming into my graduate career, I didn't even anticipate studying probability. For that, I have only the Department of Statistics to thank, as I was required to take STAT 519 with Professor Tom Sellke. The next semester I took STAT 532 with Professor Burgess Davis and never looked back. Beyond academics, I would also like to thank the Department of Statistics for supporting me financially over the years, and also for awarding me with the Bilsland Dissertation Fellowship my final semester. Without this fellowship I could not have written and completed this dissertation in the manner in which I wanted.

I would like to thank Professor Raghu Pasupathy for our discussions and his support. Additionally, I would like to thank Professor Anirban DasGupta, for instilling in me a passion for the development and history of probability and statistics, for clever solutions, and for difficult problems. I would also like to thank Professors Mark Daniel Ward and Qifan Song for their assistance over the years.

On a more personal note, I would like to thank my parents Mike and Lisa Thomas, for their unconditional love and support. They gave me all the tools I needed to become the person I wanted to be. Many people say that they have the best parents. Fortunately for me, it's true.

I would like to thank all of the friends I have made in graduate school, who made these years go by so quickly.

I would like to thank my sister Juli, for being the best sister anyone could ask for.

I would like to thank Fern, for being such a good pup.

Finally, I would like to thank the love of my life, Christine, for sticking by me and growing with me these last few years. Thank you for being patient enough to deal with

my idiosyncracies, peccadilloes, and foibles. Thank you for giving me laughter and love and being the best partner possible. You are the light that illuminates my path, a scholar par excellence, and the best thing to have come of out of my graduate school experience.

PREFACE

This dissertation is ultimately a work in the theory of applied probability, imbued with topological data analysis and with implications for statistics. It proceeds in a typical fashion: introducing a problem, informing the reader of the work done on the problem, and then solving the problem. The first two points are the focus of Chapter 1, and the last point takes up Chapter 2–5, of which Chapter 2 is background material necessary for the original research of Chapters 3–5. Chapters 3–5 follow their published versions very closely, with some additional notes to update the reader on the latest developments as well as a few new results and some proofs of technicalities that did not make it to publication.

The Chapters 3–5 form the body of my original research conducted as a graduate student at Purdue University. These chapters proceed chronologically in order of publication. Chapter 3 is concerned with the Betti number process in a probabilistic framework I deem the "traditional" setup. Chapter 4 forms the bridge between Chapters 3 and 5, and is concerned with the Euler characteristic process in the traditional setup. Chapter 5 deals with the Euler characteristic process as well, albeit in an extreme-value theoretic setup. The title of this dissertation, *Stochastic Process Limits for Topological Functionals of Geometric Complexes*, aptly describes each of these chapters.

It was my desire to have this dissertation be fairly self-contained, so we spend a fair bit of time on the concepts in algebraic topology, point processes and stochastic process limits that underpin the results of this dissertation. More than having this work be simply a concatenation of the papers I (co-)authored in graduate school, I wanted the reader to be able to learn from it, and to come away from it with a well-rounded view of the literature and what "stochastic process limits for topological functionals of geometric complexes" encompasses. I tried to strike a balance between accessibility and generality. I hope you enjoy it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\mathcal{X}_n$ Binomial process with intensity nf.

$|S|$ Cardinality of an arbitrary set S.

$\|f\|_\infty$ The essential supremum of the probability density f.

$\mathcal{F}(S)$ The collection of finite subsets of a set S.

$a \vee b$ Maximum of real numbers a and b.

$a \wedge b$ Minimum of real numbers a and b.

$\mathbb{N}$ The set of positive integers $\{1, 2, \dots\}$.

$\mathbb{N}_0$ The set of nonnegative integers $\{0, 1, \dots\}$.

$\bar{\mathbb{N}}_0$ The set $\mathbb{N}_0 \cup \{\infty\}$.

$\mathcal{P}_n$ Poisson process with intensity nf.

Φ_n Poisson process $\mathcal{P}_n$ or binomial process $\mathcal{X}_n$.

2^S The power set—i.e., set of all subsets—of a set S.

(D, J_1) Skorohod space $D[0, \infty)$ equipped with the J_1-topology.

(D, U) Skorohod space $D[0, \infty)$ equipped with the uniform topology.

ω_d Volume of the unit Euclidean ball $B(0, 1) = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.

ABSTRACT

This dissertation establishes limit theory for topological functionals of geometric complexes from a stochastic process viewpoint. Standard filtrations of geometric complexes, such as the Čech and Vietoris-Rips complexes, have a natural parameter $r \geq 0$ which governs the formation of simplices: this is the basis for persistent homology. However, the parameter r may also be considered the time parameter of an appropriate stochastic process which summarizes the evolution of the filtration.

Here we examine the stochastic behavior of two of the foremost classes of topological functionals of such filtrations: the Betti numbers and the Euler characteristic. There are also two distinct setups in which the points underlying the complexes are generated, where the points are distributed randomly in $\mathbb{R}^d$ according to a general density (the traditional setup) and where the points lie in the tail of a heavy-tailed or exponentially-decaying "noise" distribution (the extreme-value theory (EVT) setup).

These results constitute some of the first results combining topological data analysis (TDA) and stochastic process theory. The first collection of results establishes stochastic process limits for Betti numbers of Čech complexes of Poisson and binomial point processes for two specific regimes in the traditional setup: the sparse regime—when the parameter r governing the formation of simplices causes the Betti numbers to concentrate on components of the lowest order; and the critical regime—when the parameter r is of the order $n^{-1/d}$ and the geometric complex becomes highly connected with topological holes of every dimension. The second collection of results establishes a functional strong law of large numbers and a functional central limit theorem for the Euler characteristic of a random geometric complex for the critical regime in the traditional setup. The final collection of results establishes functional strong laws of large numbers for geometric complexes in the EVT setup for the two classes of "noise" densities mentioned above.

1. INTRODUCTION

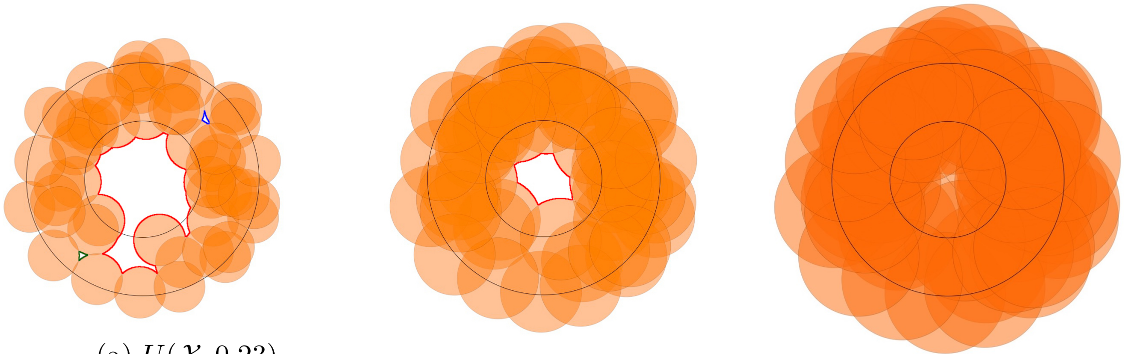
1.1 The parable of the annulus

Suppose we are given a certain number of points from a point cloud $\mathcal{X}$, which is sampled from some subset of d-dimensional Euclidean space $\mathbb{R}^d$. As in so many of the introductions to topological data analysis (TDA), we will offer the *parable of the annulus*. This story starts with a humble annulus A with inner radius 0.5 and outer radius 1 and the point cloud $\mathcal{X} \subset A$.

Figure 1.1. Left: random sample of points $\mathcal{X}$ uniformly distributed on A. Right: disks of small radius added to the point cloud $\mathcal{X}$.

If we look at the the point cloud on the left in Figure 1.1, we see a smattering of points, but in no sense do the union of the points on the left recover the essential property of the annulus—that there is an inner "castle" (and only one castle) that you can only get to by crossing the annulus' "moat". A fundamental question to not only TDA, but data analysis in general, is stated well by Ghrist [39], viz. "how does one assemble discrete points into global structure?". Adding disks of small radii to these points (Figure 1.1 right), yields a union of disks whose shape does no better in approximating the annulus on the right. Nonetheless, these padded points are no longer simply discrete, and some agglomeration has occurred. The question is whether or not there is some radius of disk such that our points satisfy what we will call the *castle & moat condition* of the annulus. If this were to happen, it would give us crucial information about the annulus (moat) from which the points were sampled.

In Figure 1.2 we enlarge the disks around the points from Figure 1.1. Starting from the left in Figure 1.2, the union of the disks don't quite satisfy the castle & moat condition—



(a) $U(\mathcal{X}, 0.23)$

(b) $U(\mathcal{X}, 0.35)$

(c) $U(\mathcal{X}, 0.52)$

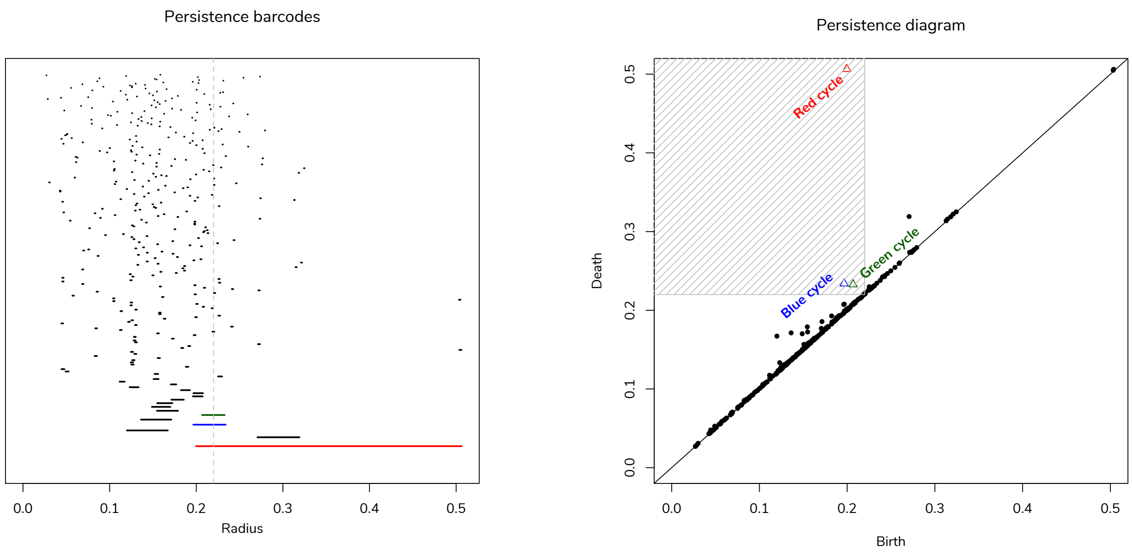
Figure 1.2. Left: disks of medium radius added to $\mathcal{X}$, 3 castles surrounded by a moat. Middle: castle and moat condition satisfied. Right: all moat, no castle.

there are too many castles (outlined in red, green, and blue). In the middle image, we have succeeded and recovered this essential property of the annulus. In the image on the right, the castle has disappeared and been covered by the moat. In mathematical parlance, the castle and moat condition corresponds to the state of affairs where the first *homology* group H_1 of the union of disks is isomorphic to H_1 of the annulus A. The number of "castles" surrounded by moats is represented by β_1, the first Betti number. From left to right in Figure 1.2, we have $\beta_1 = 3, \beta_1 = 1$, and $\beta_1 = 0$, respectively. We must also have a single connected component to truly recover the shape of the annulus (i.e., no disks far from the "moat"). Or in the language of homology, we must have the number of connected (path) components of the union of disks, β_0, equal to 1. We say that the castles that are surrounded by moats are 1-dimensional *cycles*, or 1-cycles for short. This is the terminology that we adopt from now on.

We see that only the middle image in Figure 1.2 captures the homology of the annulus correctly. Now, for the sample $\mathcal{X}$ above, we were able to find a value r such that β_1 of the union of disks of radius r, denoted $U(\mathcal{X}, r)$, is equal to 1. With a minor recalibration of our imagination we can consider $\mathcal{X}$ to be *any* finite random sample in $\mathbb{R}^d$—$\mathcal{X}_n$ if we want to emphasize that $\mathcal{X}$ consists of n points. If the vectors that compose $\mathcal{X}$ admit a probability density f, and the support of f is a manifold M, then $U(\mathcal{X}, r)$ is an estimator of M. Such ideas were predated in the work [28], where $U(\mathcal{X}_n, r_n)$ was used to estimate the support of probability measure on $\mathbb{R}^d$, where $r_n \to 0$ and $nr_n^d \to \infty$, as $n \to \infty$. Of course, many sets in $\mathbb{R}^d$ are manifolds—including a great deal that are identical with a point from the standpoint of homotopy—so this setup has tremendous utility. The pioneering work of [64, 65] established probabilistic bounds on the homology recovery of a manifold, given conditions on r_n, the shape of M, and even the distribution of noise. However, it isn't clear which r is best for allowing $U(\mathcal{X}, r)$ to recover the homology of our humble annulus, much less any manifold. Hence, we may choose to look at how the homology (in particular, β_1) evolves over a range of r—this is the idea of *persistent homology*.

Persistent homology was developed independently by Frosini in 1992 and Robins in 1999 [37, 87]. Algorithms for efficiently computing the persistent homology of $(U(\mathcal{X}, r))_{r\geq 0}$, were developed in [34]. A great summary of the development of persistent homology can be seen in [79]. Persistent homology is an essential tool to capture or estimate a shape, in analogy to how single-linkage clustering captures the evolution of components [43]. Similarly to how a dendrogram is a visual depiction of the evolution of the number of components of $U(\mathcal{X}, r)$ at different thresholds r, there are corresponding notions for persistent homology: persistence barcodes and persistence diagrams. Both of these notions were introduced in [34] and serve as way to track the evolution of the homology as the radii of the disks around the points vary. The first persistent homology PH_1 can be encoded via barcodes, or via a persistence diagram (as seen in Figure 1.3) for the points in $\mathcal{X}$.

There are many excellent introductions to persistent homology, and ones to suit anyone's tastes—be it statistical [96], mathematically rigorous [23, 32], or synoptic [3, 39]. One of the main ideas behind TDA is that longer barcodes (or points far from the diagonal in the persistence diagram) correspond to statistical significant features of an underlying manifold. What is interesting in both images in Figure 1.3 is that although the large castle in the center of the annulus is recovered for large range of r, at any given radius r, the union of balls $U(\mathcal{X}, r)$ contains a panoply of noisy features. This is especially true when the radius is small. Hence, one is lead to wonder: what is the behavior of these extraneous 1-cycles? This question extends to arbitrary k-cycles in $\mathbb{R}^d$. The number of k-cycles for a subset $A \subset \mathbb{R}^d$ is recorded by the kth Betti number β_k, $k \geq 0$—where a k-cycle bounds a $(k+1)$-dimensional object—just like our moat bounds our 2-dimensional castle. In fact $\beta_k(U(\mathcal{X}, r))$



(a) PH_1 for points from Figure 1.1.

(b) PH_1 for points from Figure 1.1.

Figure 1.3. The first persistent homology of $\mathcal{X}$. The 1-cycles are highlighted as in Figure 1.2a. Number of points in hatched region and barcodes that intersect dotted line represents $\beta_k(U(\mathcal{X}, 0.23)) = 3$. Persistent homology computed with **R** package `TDA` [36].

is the number of points in rectangle $[0, r] \times (r, \infty)$ in the kth persistence diagram and the number of bars that intersect the vertical line in the plot of the kth persistence barcodes—see Figure 1.3. A related topological summary called the Euler characteristic χ is equal to the alternating sum of the Betti numbers $\chi = \sum_k (-1)^k \beta_k$. Astonishingly, this is much simpler to compute—as it admits a purely local representation—so we would like to know how the extraneous k-cycles affect χ as well.

Because of the global nature of Betti numbers (and to a greater extent, persistence diagrams) they can be difficult to calculate in practice. Though useful, even calculating Betti numbers can be computationally expensive and often lack analytic formulas [85]. There is another option to summarize the shape of $U(\mathcal{X}, r)$—the Euler characteristic χ. The Euler characteristic is one of the oldest and simplest topological summaries. The Euler characteristic was discovered by Leonhard Euler [84] and was shown to be an invariant of polyhedra, with all polyhedra having Euler characteristic equal to 2. Further development of this concept lead to the proof that the Euler characteristic distinguishes all orientable surfaces from each other, by their genus [44]. A key to computing persistent homology and even homology is the availability of a combinatorial representation of $U(\mathcal{X}, r)$. This representation is called the Čech complex, denoted $\check{C}(\mathcal{X}, r)$. However, in practice the Čech complex is expensive to compute (see [20]) requiring us to look at the intersection of all of the balls in $U(\mathcal{X}, r)$. Another option is the Vietoris-Rips complex $\mathcal{R}(\mathcal{X}, r)$, which adds a *simplex* for each set of points in $\mathcal{X}$ with pairwise distances less than r. Both the Čech and Vietoris-Rips complexes, which we denote agnostically as $\mathcal{K}(\mathcal{X}, r)$, are particular instances of simplicial complexes, which are collections of sets that are closed under inclusion and have a geometric realization in some Euclidean space. Because both the Čech and Vietoris-Rips complexes depend on the configuration of their vertices in space, they are called *geometric complexes*. When the point cloud $\mathcal{X}$ is random, then the complexes $\mathcal{K}(\mathcal{X}, r)$ are called *random geometric complexes*. The beauty of the Euler characteristic is that it can also be represented as the alternating sum of simplex counts of different dimensions, facilitating limits for random geometric complexes that are much more elusive for Betti numbers. This dual nature of the Euler characteristic, as the sum of Betti numbers or simplices, is discussed in Section 2.1.

Returning to the recovery of our annnulus A, it was realized early on that for a radius r such that nr^d is small, or at least not large, though manifold recovery is not possible (see [17, 65]) there is interesting homology even if the support of f is dull, with contractible or trivial topology [86]. Indeed, the behavior of the random Betti numbers, and many other functions of $\mathcal{K}(\mathcal{X}_n, r_n)$ is controlled by nr_n^d. In the right figure of Figure 1.1, $nr_n^2 \approx 0.1$; in Figure 1.2a $nr_n^2 \approx 2.8$; in Figure 1.2b $nr_n^2 \approx 6.5$. The behavior in the *regimes* of nr_n^d where the limit as $n \to \infty$ is 0, finite and positive, or infinite corresponds to the behavior of the Betti numbers seen in the rightmost figure of Figure 1.1, Figure 1.2a and Figure 1.2b. That is, we see sparsity give way to the formation of extraneous features, then to recovery and finally contractibility in Figure 1.2c, where $nr_n^2 \approx 14.3$. Given the fact that topological summaries like Betti numbers have found utility in applied settings as the data on which statistical analyses are performed [48, 90]—it is worthwhile objective to understand the behavior of these quantities for random samples for as wide of a class of distributions as possible, even when homology is *not* recovered and nr_n^d has a limit in $[0, \infty)$. Not only do we want to

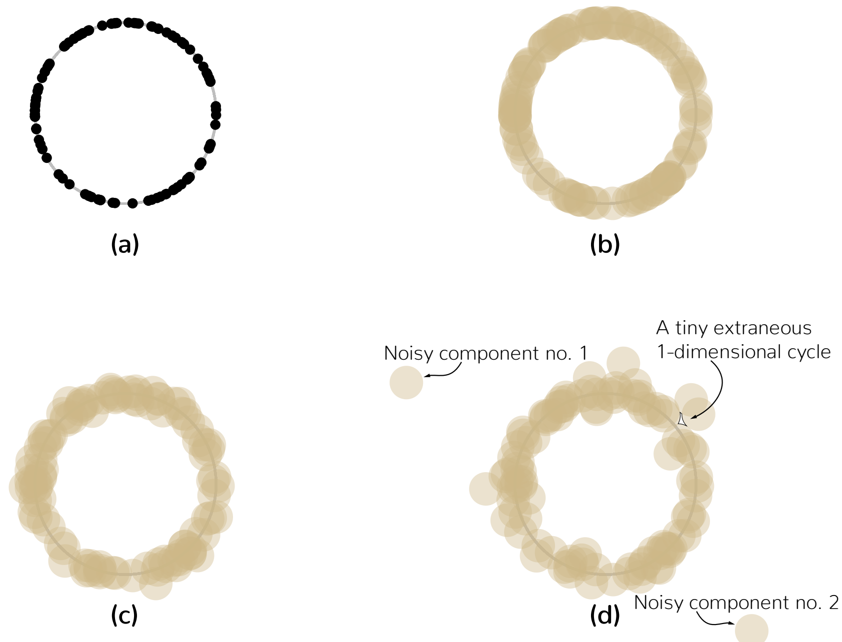


Figure 1.4. (a) A random sample $\mathcal{X}$ of 100 points, uniformly distributed on the unit circle S^1. (b) The union of balls of radius 0.2 around each point in $\mathcal{X}$. (c) The union of balls when $\mathcal{X}$ is perturbed by Gaussian noise. (c) The union of balls when $\mathcal{X}$ is perturbed by Cauchy noise. Noise is applied in the same manner as [64]. Topology is recovered by the union of the balls in cases (b) and (c), but that is not the case if the heavy-tailed Cauchy noise is added.

understand what the behavior of these extraneous k-cycles is like, but it is very difficult to apply traditional statistical methods to persistence diagrams and barcodes: for example, the space of persistence diagrams do not admit a unique "mean" [61]. Though there are other topological summaries such as the persistence landscape [22] which one could apply to point cloud data $\mathcal{X}$, the Betti number process $t \mapsto \beta_k(\mathcal{K}(\mathcal{X}, t))$ and the Euler characteristic process $t \mapsto \chi(\mathcal{K}(\mathcal{X}, t))$ are the focus of this dissertation[1]. Their behavior is a natural extension of the behavior of $\beta_k(\mathcal{K}(\mathcal{X}, r))$ and $\chi(\mathcal{K}(\mathcal{X}, r))$ for single values of r and they extend non-functional summaries that have been investigated before (which we discuss in Section 1.2.1), so there is a strong probabilistic basis for considering them.

[1]↑We adopt t instead of r as the parameter of the Čech and Vietoris-Rips complexes throughout the remainder of this dissertation. This is to emphasize that we are dealing with stochastic processes, which are parametrized by time.

In the eyes of homotopy, the parable of the annulus is the same as the parable of the circle. To satisfy the castle & moat condition for the circle, is to recover the homology of it, in the same sense as for the annulus. The difference of course is that a probability measure on the circle is not absolutely continuous with respect to Lebesgue measure in $\mathbb{R}^2$. In learning a submanifold of $\mathbb{R}^d$, one may not necessarily have a probability distribution that is supported on a subset of $\mathbb{R}^d$ of positive Lebesgue measure. We mention this more general setup because it is the one of the seminal articles [64, 65]—in which sufficient conditions are given for the recovery of the homology of a manifold, conditional upon an understanding of how points are perturbed from the manifold. In [65], given a "nice" manifold, it was shown that one can recover the topology of the manifold by a sufficiently dense random sampling of points if the noise is bounded. In [64] it was shown that the recovery is still possible by a sufficiently dense random sampling of points if the noise is standard multivariate Gaussian and the variance is bounded by a function of the reach and dimension of the manifold[2]. However, if the points on the manifold are perturbed by heavy-tailed noise, the recovery of topology will be severely impacted because of extraneous homological elements generated by this noise. In Figure 1.4, we wish to again satisfy the castle & moat condition of the circle with $U(\mathcal{X}, r)$. In Figure 1.4b and 1.4c, the union of balls satisfy the castle & moat condition, as there is no noise in the case of the former and in the latter case the size of the noise is sufficiently small. However, in Figure 1.4d the noise added to points in S^1 has a heavy-tailed Cauchy distribution. Consequently, $\beta_0 = 3$ and $\beta_1 = 2$. This phenomenon in case (d) raises the question of how the shape of these elements away from the center of S^1 (equivalently the center of the annulus A) may behave in general; this is roughly the idea of what is called *topological crackle*, which was first investigated in [4].

The parable of the annulus does not have an ending. In fact, this dissertation is a continuation of that story; it is a story of the convergence of topological functionals, the homology of noise, stochastic process limits and point processes. All of the results herein apply to the behavior of random balls, and more broadly random complexes, on the annulus A and for general probability densities on $\mathbb{R}^d$. This is not the final word on anything of this, but I hope that it is a positive step towards the story's resolution. In the following section,

[2]↑Portions of the this section are reproduced from [93].

Section 1.2, we will discuss the previous probabilistic results for the topological summaries, in the traditional setup of a probability distribution on $\mathbb{R}^d$ absolutely continuous with respect to Lebesgue measure, as well as for the case of extreme-valued noise, with a subexponential or heavy tail, in the presence of topological crackle—the extreme value theory (EVT) setup. Section 1.2.1 discusses the former and Section 1.2.2 the latter.

1.2 Some limit theory for topological functionals

1.2.1 Traditional setup

The behavior of topological functionals in the next three subsections compose what I deem the "traditional setup" in the study of topological functionals of point processes, and the theory of random geometric complexes. We begin with an exposition on the limiting results for Betti numbers, that lead up to the publication [72]. We follow this up with a treatment on a discussion of the limiting behavior of the Euler characteristic of random geometric complexes that preceded [92] and conclude with the work on stochastic process limits relevant to each of these two studies.

Genesis: Betti numbers

Questions about the Betti numbers of random point clouds (and the union of balls) date back to those posed by Arnold in 1973 [5]. The investigation of the limiting behavior of the union of balls around a stationary Poisson process began even earlier, and was initiated by Gilbert in [41]. However, the study of Betti numbers of random point clouds began in earnest with the work of Robins in the early 2000's [85, 86] and by Kahle in 2011 [49]. In the work of Robins, the expected Betti numbers of the union of balls—whose centers were generated by a homogeneous Poisson process—were studied for centers restricted to a subset of $\mathbb{R}^m$, $m = 2, 3$ of unit area. Before outlining the literature in earnest, let $\beta_{k,n} := \beta_k(\mathcal{K}(\Phi_n, r_n))$, where Φ_n is a point cloud $\mathcal{X}_n$ of n i.i.d points with probability density f or a Poisson process $\mathcal{P}_n$ of Poi(n) i.i.d points (with density f)[3]. These papers, along with the talk given by Persi

[3]↑The particular meaning of Φ_n will depend on the context, and in many cases the limiting behavior of the topological functionals of random geometric complexes does not depend on whether Φ_n is $\mathcal{P}_n$ or $\mathcal{X}_n$.

Diaconis at MSRI in 2006 [29], outlined numerous fruitful ideas which were taken up in [49] to produce rigorous limiting results for expected Betti numbers $\mathbb{E}[\beta_{k,n}]$ for both Čech and Vietoris-Rips complexes (including for a uniform distribution studied in [85]). The study by Kahle and its sequels [51, 52] underpins and motivates much of the research here.

The genesis and the initial proof technique for the study of random geometric complexes comes from the study of random geometric graphs, the pre-eminent study of which is by Penrose [76]. There are a fair number of the results of the Penrose monograph that are relevant to this dissertation. The reasoning behind this is because, in the presence of some sparsity, higher-order components vanish in probability, so results on homology—a global phenomenon—become asymptotically local. To give a taste of what this means, we investigate Chapter 3 of [76], which contains a compendium of important results that mirror the results in this document. Similarly to [16], we detail the parallels between results for Betti numbers of random geometric complexes and their corresponding precursors for random geometric graphs. These parallels are summarized in Table 1.1. Let v_k be the number of vertices such that some function h of a point cloud $\mathcal{X}$ satisfies $h(\mathcal{X}) = 0$ if $\mathcal{X}$ contains no more than v_k points. For Betti numbers in the Čech complex, $v_k = k + 1$; for Betti numbers in the Vietoris-Rips complex $v_k = 2k + 1$; and for an induced (connected) subgraph[4] Γ of order k, $v_k = k - 1$. Before discussing the aforementioned parallels, we define the random geometric graph $G(\Phi_n, r_n)$ where an edge is included in $G(\Phi_n, r_n)$ if the points are within distance r_n of each other and where Φ_n is either the binomial process $\mathcal{X}_n$ or the Poisson process $\mathcal{P}_n$ as with $\beta_{k,n}$. As mentioned above, the behavior of all of these quantities are determined by the average degree of a point in the support of f, i.e. nr_n^d. Namely, if x is in the interior of the support of f, then

$$\mathbb{E}[\Phi_n(B(x, r_n))] = n \int_{B(x,r_n)} f(y)\,\mathrm{d}y \sim \omega_d n r_n^d f(x),$$

by the Lebesgue differentiation theorem, where ω_d is the volume of the Euclidean ball of unit radius in $\mathbb{R}^d$. We adopt terminology for the behavior of nr_n^d that corresponds to that in the

[4] ↑An induced subgraph is a subgraph $H = (V', E')$ of a graph $G = (V, E)$ such that if $e = (v_1, v_2) \in E$ and $v_1, v_2 \in V'$, the edge $e \in E'$.

Table 1.1. Extensions of Penrose results on subgraph counts to Betti numbers

PENROSE CHAPTER 3 ($v_k = k-1$)	BETTI NUMBERS ($v_k = k+1$ OR $2k+1$)
Proposition 3.2 Explicit limit for $\frac{\mathbb{E}[J_n(\Gamma)]}{n^{v_k+1}r_n^{dv_k}}$ $nr_n^d \to 0$.	**Theorems 3.1 & 3.2** [49] Limit for $\frac{\mathbb{E}[\beta_{k,n}]}{n^{v_k+1}r_n^{dv_k}}$ $nr_n^d \to 0$.
Theorem 3.4 (part 1) Poisson weak limit for $G_n(\Gamma)$ $n^{v_k+1}r_n^{dv_k} \to \alpha \in (0,\infty)$.	**Theorems 3.2(*ii*) & 4.2(*ii*)** [52] Poisson weak limit for $\beta_{k,n}$ $n^{v_k+1}r_n^{dv_k} \to \alpha \in (0,\infty)$.
Theorem 3.4 (part 2) CLT for $G_n(\Gamma)$ $nr_n^d \to 0$, $n^{v_k+1}r_n^{dv_k} \to \infty$.	**Theorems[1] 3.2(*iii*) & 4.2(*iii*)** [52] CLT for $\beta_{k,n}$ $nr_n^d \to 0$, $n^{v_k+1}r_n^{dv_k} \to \infty$.
Proposition 3.3 Explicit limit for $\frac{\mathbb{E}[J_n(\Gamma)]}{n}$ $nr_n^d \to \xi \in (0,\infty)$.	**Theorems[2,3] 1.3 & 1.6** [94] Explicit limit for $\frac{\mathbb{E}[\beta_{k,n}]}{n}$ $nr_n^d \to \xi \in (0,\infty)$.
Proposition 3.15 SLLN for $J_n(\Gamma)$ $nr_n^d \to \xi \in (0,\infty)$.	**Theorem[3] 2.1** [42] SLLN for $\beta_{k,n}$ $nr_n^d \to \xi \in (0,\infty)$.

[1] The proof of this result in the Čech complex case contained an error. The proof was remedied in [51], albeit with the restriction that $n^{1+\delta}r_n^d \to 0$ for some $\delta > 0$.

[2] A corollary to Theorem 2.1 of [42], Remark 4.9, establishes the explicit limit for $\mathbb{E}[\beta_{k,n}]/n$ in greater generality than in [94].

[3] The indicated theorem was established only for the Čech complex, where $v_k = k+1$.

literature. The situation when $nr_n^d \to 0$ is called the *sparse* regime, because the "average degree" tends to zero. The limiting behavior for this regime is dictated by $n^{v_k+1}r_n^{dv_k}$. The *critical* (or thermodynamic) regime is when nr_n^d tends to a constant. The *dense* regime is when $nr_n^d \to \infty$. We do not consider the dense regime in this dissertation.

For an induced connected subgraph Γ of order k, we let $G_n(\Gamma), J_n(\Gamma)$ be the number of induced sugraphs (resp. induced subgraph components) isomorphic to Γ in $G(\Phi_n, r_n)$. There are results in the literature that extend those of Penrose that are not listed in the table—because they were proved either in the author's work in [72], or after [72] appeared. We will discuss those in the beginning of Chapter 3.

It is important to realize that the parallels between subgraph counts and Betti numbers start to break down in the critical regime, $nr_n^d \to \xi \in (0, \infty)$ and completely fall apart when $nr_n^d \to \infty$. Indeed, [49] established that when $nr_n^d \geq C \log n$, for a uniform distribution on a compact, convex set with non-empty interior, then

$$\mathbb{P}(\check{C}(\Phi_n, r_n) \text{ is contractible}) \to 1, \quad n \to \infty,$$

where the constant C depends on the dimension d and the Lebesgue measure of the support. This implies that $\beta_k(\check{C}(\Phi_n, r_n)) \xrightarrow{P} 0$, as the homology of a point (besides H_0) is trivial [44]. Clearly this breaks down in the case of $G_n(\Gamma)$ which tends to infinity as $n \to \infty$ (variance and expectation asymptotics are the same, [76, see Proposition 3.7]). However, there are analogues in the random geometric graph case between connectivity and contractibility. In fact, $\mathbb{P}(G(\Phi_n, r_n)$ is connected$) \to 0$, $n \to \infty$ when $nr_n^d \leq C' \log n$ for C' depending on d and points distributed on the unit cube $[0,1]^d$. Connectivity also occurs for $nr_n^d \geq C'' \log n$ ([76, Theorem 13.10]). Additionally relevant to this dissertation are the central limit theorem (CLT) and weak laws for component counts, which can be considered as β_0, the limiting results of which are also seen in Chapter 13 [76].

There are many other additional results on the limiting Betti numbers $\beta_k(\mathcal{K}(\Phi_n, r_n))$ that are relevant and worth discussing. As mentioned at the start of this section, the study of the limiting properties of the union of balls started with [41], which was motivated by problems of transmission in large communication networks and the spread of a contagious disease. The underlying data-generating mechanism for the points was a homogeneous Poisson process in the plane with intensity λ. In this work percolation properties were studied, and so-called continuum percolation (on $\mathbb{R}^d$)—whose theory is masterfully exposited in the monograph [60]—has not only a fascinating theory to go with it, but has applications to the behavior of graph properties of the Čech complex as well—see [12, 49]. More connections will be detailed in the following subsection on the Euler characteristic. Limiting results for Betti numbers of both Čech and Vietoris-Rips complexes of stationary point processes (which include the homogeneous Poisson process), including those which exhibit some degree of spatial dependence, was given in [99][5]. Strong laws for Betti numbers of stationary point processes for the Čech complex were given in [100]. It is tremendously interesting that the strong law of large numbers of [42] has a limit that is defined in terms of the stationary case. In particular, if Φ_n has intensity $n\lambda 1_{[0,1]^d}$ for $\lambda > 0$, [100] showed that if $nr_n^d \to \xi \in (0, \infty)$, and $\hat{\beta}_k(\lambda, \xi)$ is a particular limiting function depending on d, then

$$\frac{\beta_k(\check{C}(\Phi_n, r_n))}{n} \to \hat{\beta}_k(\lambda, \xi), \quad \text{a.s.}. \tag{1.1}$$

The article [42] then established for $\mathcal{X}_n$ n i.i.d. points with density f (subject to very minor conditions), that

$$\frac{\beta_k(\check{C}(\Phi_n, r_n))}{n} \to \int_{\mathbb{R}^d} \hat{\beta}_k(f(x), \xi)\, \mathrm{d}x. \tag{1.2}$$

The antecedent article [94], which followed [100] proved (1.2) for f in a more specialized case where f is *uniformish*—i.e., has compact and convex support[6] with the additional assumptions that $0 < \inf f(x) \leq \sup f(x) < \infty$. These works resolved one of the stated future directions guiding TDA research in [16].

In [100] the authors also gave the first central limit theorem for $\beta_k(\check{C}(\Phi_n, r_n))$ in the critical regime, for Φ_n satisfying similar conditions to those that guarantee (1.1). This was improved upon in [95] for $nr_n^d \to \xi$ with ξ restricted in a similar fashion to what we will be demonstrated in Chapter 3. The study [95] extends the ideas of weak and strong stabilization of functionals [77]. These ideas have fruitful applications in the convergence of Betti numbers, even *persistent* Betti numbers: i.e., the number of cycles that are born before or at time s and die after time t, $t \geq s$ (see [46, 56]). We will have more to say about this at the beginning of Chapter 3. We now discuss the history of the Euler characteristic of random geometric complexes in the traditional setup.

[5]↑Results for $G_n(\Gamma)$ and $J_n(\Gamma)$ in these general cases were given as well.
[6]↑This article included the assumption f was Riemann integrable.

The Euler characteristic

The limiting behavior of the Euler characteristic $\chi_n := \chi(\mathcal{K}(\Phi_n, r_n))$, being the alternating sum of Betti numbers, can often be obtained as a result of the limiting behavior of $\beta_{k,n}$. However, the earliest results on the Euler characteristic of random geometric complexes were established by appealing to either critical points of distance functions (for the Čech complex) or simplex counts (a very general setup). The union of balls $U(\mathcal{X}, r) = \{y \in \mathbb{R}^d : d(x, y) \leq r, \text{ for some } x \in \mathcal{X}\}$, can be represented in terms of sublevel sets of the distance function. Understanding when the homology of $U(\mathcal{X}, r)$ changes based on r can then be assessed by examining the subsets of $\mathcal{X}$ with certain properties, and which determine *critical points*. The importance of this is that due to Morse theory [38], the Euler characteristic of $U(\mathcal{X}, r)$ can be represented as the alternating sum of the number of (Morse) critical points as well.

The questions involving the asymptotics of the number of critical points of Čech complexes were first taken up in [13, 14]. The results of [13, 14] established asymptotics for the first moment of χ_n. Specifically, one has

$$\frac{\mathbb{E}[\chi_n]}{n} \to \begin{cases} 1 & \text{if } nr_n^d \to 0 \\ \sum_{k=0}^{d} (-1)^k \gamma_k(\xi) & \text{if } nr_n^d \to \xi \in (0, \infty) \\ 0 & \text{if } nr_n^d \to \infty, \end{cases} \tag{1.3}$$

where $\gamma_k(\xi)$ can be seen in [14] and depends on f and d in addition to the obvious dependencies. The intuition behind this is that in the sparse regime, there are either n points or approximately n points in Φ_n and the total number of features on $k \geq 2$ points grows at the rate $n^k r_n^{d(k-1)} = o(n)$. Thus, for large samples χ_n simply counts the number of points in the sparse regime. For the dense regime, this result is highly nontrivial. It conveys that the homology of $\check{C}(\Phi_n, r_n)$ grows sublinearly. This entails that because of the amount of space covered by $U(\Phi_n, r_n)$, cycles that are generated quickly disappear. In other words, they become boundaries. The critical regime is especially relevant to Chapter 4 and to a lesser extent, Chapter 5. Because the number of critical points (and simplex counts) of each order grow at the same rate, we see an alternating of sum of limiting functions as well.

At the same time, a publication appeared that found exact formulas for the mean and variance of the Euler characteristic, as well as a concentration inequality for the case when the points are distributed according to a homogeneous Poisson process on the flat torus, and the Čech and Vietoris-Rips complexes are formed with respect to the L^∞ norm [27]. This was also one of the first publications to explicitly consider homology recovery in the stochastic setting. Corollary 4.3.3 can be seen as extending the results of these efforts, finding exact formulas for a class of distributions using the L^∞ norm. The study [17] extended the results of [14] to the case where the points were generated on a smooth, closed m-dimensional manifold M embedded in $\mathbb{R}^d$, for $m < d$. Using critical points, the same result as (1.3) was established (with d replaced by m in the formula), along with other results such as asymptotics for the expected kth Betti numbers in the sparse and critical regimes for the manifold setup, i.e., $nr_n^m \to [0, \infty)$. This study also established groundbreaking results on homology recovery for manifolds, given a sufficiently rich (cf. uniformish) density on M. Additional results for convergence of the Euler characteristic on random geometric complexes have been given. For example, the paper [47] established a multivariate central limit theorem for the intrinsic volumes of a stationary Boolean model composed of general convex grains, including the Euler characteristic; additionally, the earlier book [88] established ergodic theorems for the Euler characteristic of a Boolean model over a stationary and ergodic point process.

It is also important to discuss the connection between the Euler characteristic and percolation. The zeroes of the mean Euler characteristic process (curve) have been shown to provide tight bounds for percolation thresholds for site and bond percolation in two dimensions [63]. Connections between the mean Euler characteristic process and other types of percolation phenomena, specifically *homological percolation*, were studied in [18, 19]. Homological percolation is interested in phenomena such as the radii (or limiting value of nr_n^d) beyond which say the two 1-cycles of the two-dimensional torus appear with probability 1, and below which they are recovered with probability 0. More information on this can be seen in Chapter 4.

Finally, the Euler characteristic has found many applications outside of the literature on random geometric complexes, especially in the arena of random fields [2, 98] and computer vision [83]. Connecting the behavior of sublevel sets of random fields and random geometric complexes (an open problem in [16]), would likely begin by looking at the Euler characteristic.

The stochastic process approach

The novelty of this dissertation is in the fact stochastic process limits are discussed, unlike the treatments of limits for topological functionals discussed in the preceding two sections. We look at processes such as $(\beta_k(\mathcal{K}(\Phi_n, r_n(t))),\ t \geq 0)$ and $(\chi(\mathcal{K}(\Phi_n, r_n(t))),\ t \geq 0)$, where $r_n(t) = s_n t$ and ns_n^d tends to 0, some finite positive constant or infinity. Part of the motivation for this parametrization of the radii of the balls in the geometric complex $\mathcal{K}$ (see Definition 2.1.5) comes from the study in Chapter 4 of [76], which investigated the number of vertices in the random geometric graph $G(\Phi_n, r_n(t))$ with vertex degree at least k, when $ns_n^d = 1$. Furthermore, they established a functional central limit theorem for the processes

$$\left(\sum_{i=1}^{|\Phi_n|} 1\Big\{|\Phi_n \cap B(X_i, r_n(t))| \geq k+1\Big\},\ t \geq 0\right).$$

The convergence of this process in Skorohod space was motivated by the earlier study of functional convergence of the empirical processes of weighted kth nearest-neighbor distances in [75] and $k = 1$ in [8]. Further studies of stochastic process limits for topological functionals brings us to the extreme value theory setup. We refrain from discussing specifics for now, but simply state that functional limit theorems preceding the works seen in this dissertation were proved in [67] for subgraph counts, Betti numbers [68] for i.i.d points clouds in the tail, Betti numbers for data in the tail generated by a moving-average process [69] and convergence of random persistence diagrams [46, 71], from which certain functional results may be derived. Proving weak convergence for quantities such as the Betti number and Euler characteristic processes can lead to insights into the behavior of lifetime sums of persistence diagrams investigated in [45, 68], and often yields more explicit formulas than non-functional results. The results in Chapters 3 and 4 were advanced in the articles [11, 55, 56] for Betti number processes and [57] for the Euler characteristic process, which we shall discuss at the beginning of those respective chapters.

1.2.2 Extreme value theory setup

In Section 1.1, we saw that the castle & moat condition was satisfied in the case where points are perturbed from a circle by Gaussian noise, and gave a reference for [64] which established this result. Of course, much less was known about how homology behaves in the presence of various types of noise distributions—see for example Figure 1.4. Consider the following heavy-tail density on $\mathbb{R}^d$:

$$f(x) = \frac{C}{1 + \|x\|^\alpha}, \quad x \in \mathbb{R}^d \tag{1.4}$$

with $\alpha > d$. Consider a sequence of radii $(R_{k,n})_{n\geq 1}$, where $R_{k,n} = (Cn)^{\frac{1}{\alpha - d/(k+1)}}$ for each $k \geq 0$. Additionally define the sequence of radii $R_n^{(c)} = c_p(n/\log n)^{1/\alpha}$, $n \in \mathbb{N}$, with c_p some constant depending on α, d, and C. Standard arguments from extreme value theory (see [81, 82]) establish that if $\mathcal{X}_n = \{X_1, \dots, X_n\}$ consists of i.i.d with density f, then

$$|\mathcal{X}_n \cap B(0, R_{0,n})^c| \Rightarrow \text{Poi}\left(\frac{s_{d-1}}{\alpha - d}\right).$$

The paper [4] established an annuli structure for Betti numbers for the density f in (1.4), as well as an exponential density. Namely, outside out of $R_{k+1,n}$ (because no k-cycle can form on $k+1$ or fewer points in the Čech complex) the expected kth Betti number of the union of balls— $\mathbb{E}[\beta_k(U(\mathcal{X}_n \cap B(0, R_{k+1,n})^c, 1))]$—tends to a positive finite constant, and for sequences $(R_n)_{n\geq 1}$ such that $R_n/R_{k+1,n} \to 0$ or $R_n/R_{k+1,n} \to \infty$ this same quantity tends to ∞ and 0 respectively. The authors also showed that the union of unit balls with centers in $B(0, R_n^{(c)})$, becomes contractible (which implies it the Betti numbers all become zero). There is no topological crackle for the Gaussian distribution, so that outside of the core $B(0, R_n^{(c)})$ there are hardly any points and the homology becomes trivial.

The ensuing study [70] exhaustively categorized this phenomena for a wide range of features and examined weak convergence rather than simply convergence of expectations.

Broadly stated, what this paper did is rigorously categorize topological crackle for a general functional. We can define topological crackle for various quantities such as Betti numbers, or for subgraph counts. However, here we define the phenomenon in terms of k-simplex counts $S_{k,n}$, which equal the number of subsets of cardinality k of $\mathcal{P}_n$ that have either circumradius (Čech) or diameter (Vietoris-Rips) less than or equal to 1.

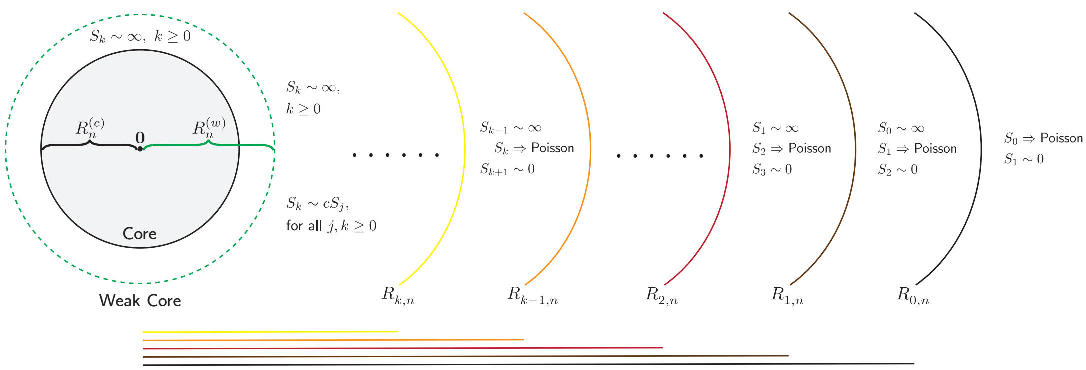


Figure 1.5. Illustration of crackle phenomenon for simplex counts

In [70], the authors established that for each $k \in \mathbb{N}$, the number of k-simplices $S_{k,n}$ whose vertices lie outside of $B(0, R_{k,n})$ converge weakly to a Poisson distribution. Similarly to the situation with the Betti numbers above, $S_{k,n}$ tends to zero for vertices outside of $B(0, R_{k-1,n})$ and tends to infinity for vertices outside of $B(0, R_{k+1,n})$. The crackle phenomenon for simplex counts is summarized in Figure 1.5. Finally, in [67] the authors defined the concept of a *weak core*, a sequence of radii $(R_n^{(w)})_{n\geq 1}$ such that $nf(R_n^{(w)}e_1) \to 1$ for a spherically symmetric density f, where e_1 is the vector in $\mathbb{R}^d$ with 1st element 1 and the rest equal to zero. For both heavy-tailed distributions and distributions with exponentially-decaying tails (e.g. $f(x) = Ce^{-\|x\|^\tau/\tau}$, $0 < \tau \leq 1$), the number of simplices of different dimensions outside of $B(0, R_n^{(w)})$ have the same growth rate, i.e. $S_{k,n} \sim cS_{j,n}$, where c depends on $j, k \geq 0$ along with the density f and the dimension d. In particular, one can gain an intuition for this by noting that $R_n^{(w)} = (Cn)^{1/\alpha}$ for the power-law density in (1.4) so $R_n^{(w)} = \lim_{k\to\infty} R_{k,n}$. This scenario allows us to establish nontrivial limits for the Euler characteristic in this extreme-value theory (EVT) setup, as we do in Chapter 5.

Further results have been established for the limiting behavior of topological functionals in the EVT setup. Owada and Adler [70] proved much more than mentioned above—the authors established Poisson limit theorems for random measures induced by geometric functionals of point clouds with heavy or exponentially-decaying tails. Included in this highly general setup is the vanishing of Betti numbers for the extreme sample clouds of Gaussian distributions. One of the earliest treatments of stochastic process limits for topological functionals in either the EVT and traditional setups appeared in [67], in which stochastic process limits are established for subgraph counts of the family of random geometric graphs outside of an expanding ball, $\Big(G(\mathcal{P}_n \cap B(0, R_n)^c, t),\ t \geq 0\Big)$, with t as the "time" parameter, for $nf(R_n e_1) \to 0$, a finite constant, or infinity. Specifically, Owada [67] established functional central limit theorems in Skorohod space $D[0, \infty)$ of real-valued functions on $[0, \infty)$ that are right-continuous with left-limits for these "extremal" subgraph counts, where the density f has a heavy or exponentially-decaying tail. This article was followed up in [68], with a treatment of Betti number processes of extreme sample clouds with heavy-tailed distributions, for the Čech complex setup. Specifically stochastic process limits (in $D[0, \infty)$ as well as finite-dimensional convergence) for the Betti number processes $\Big(\beta_k\Big(\check{C}(\mathcal{P}_n \cap B(0, R_n)^c, t)\Big),\ t \geq 0\Big)$ were considered In [69], the author considered a different setup where points come from a moving average process, rather than Φ_n, as has been the case with nearly every setup discussed so far (save for stationary setup of [99]). For data from this moving average process, stochastic process limits were proved for the Čech and Vietoris-Rips complex. Finally, [71] discussed point process convergence for persistence diagrams of extreme sample clouds for both heavy-tailed distributions and distributions with exponentially-decaying tails.

1.3 Outline

This dissertation details the limiting properties of topological functionals of binomial and Poisson point processes, in particular the Betti number and Euler characteristic processes. The requisite background material is addressed in Chapter 2 and details the notions in topology, point processes, and stochastic processes necessary for the subsequent chapters that form this dissertation's novel contribution. Chapter 3 discusses the results of [72], i.e., the process-level (finite-dimensional) central limit theorems for the Betti number process. The primary contribution here is an analytic representation of the Betti number process in the critical regime, and both a CLT and functional Poisson limit theorem in the sparse regime (for either Poisson or binomial point processes). These results are restricted to the Čech complex (the union of balls).

Chapters 4 and 5 deal with Euler characteristic process in two distinct setups. Chapter 4 demonstrates a functional strong law of large numbers and functional central limit theorem in the critical regime. Both of these results give fully-fledged functional limit theorems for this topological summary and hinge upon the representation of the Euler characteristic as the alternating sum of simplex counts. Chapter 5 proves functional strong laws of large numbers in the EVT setup, where points are scattered far from the center of a spherically symmetric distribution, and the Euler characteristic is formed from a general simplicial complex based off these points. The underlying process may be Poisson or binomial and the geometric complex need obey only certain regularity conditions.

1.4 Notation

Before continuing, it is worth taking a moment to discuss some of the notation that is used throughout this dissertation. We define m to be Lebesgue measure on d-dimensional Euclidean space $\mathbb{R}^d$. Let $B(x,t)$ be the closed ball of radius $t \geq 0$ around x. Namely, $B(x,t) := \{y \in \mathbb{R}^d : \|x-y\| \leq t\}$ (where $\|\cdot\|$ is any norm on $\mathbb{R}^d$). Unless otherwise specified, we let f be an essentially bounded probability density on $\mathbb{R}^d$. Essentially bounded means that the essential supremum of f, denoted

$$\|f\|_\infty := \inf\{a \in [0,\infty) : m(f^{-1}(a,\infty)) = 0\}\},$$

is finite. Taking $k \geq 0$ we specify m_k to be Lebesgue measure on $(\mathbb{R}^d)^{k+1}$, and

$$C_{f,k} := \frac{1}{(k+2)!}\int_{\mathbb{R}^d} f(x)^{k+2}\,\mathrm{d}x. \tag{1.5}$$

This quantity comes into play for our treatment of process-level Betti numbers, and though $C_{f,-1}$ is well-defined, it is equal to 1. Note that $m = m_0$.

For any real numbers a and b, let $a \vee b := \max\{a, b\}$ be the maximum of a and b and $a \wedge b := \min\{a, b\}$ be the minimum of a and b. Throughout this dissertation, we let C^* denote a generic positive constant that potentially varies across and within the lines. The notation $\sim$ should be clear depending on the context, but variously denotes equivalence between objects x and y when we write $x \sim y$; that a random variable X has a certain distribution F when we write $X \sim F$; and that two real-valued sequences $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ (with $b_n > 0$ for n large enough) obey

$$\lim_{n \to \infty} a_n / b_n = 1,$$

when we write $a_n \sim b_n$. For similarly defined sequences we write $a_n = O(b_n)$ if there exists some $M > 0$ such that $|a_n| \leq M b_n$ for all n large enough. Also, we say $a_n = \Theta(b_n)$ if $a_n \sim c b_n$ for some $c > 0$.

Finally, it is often useful to impose an ordering on $\mathbb{R}^d$ to reduce redundancy in the presence of a symmetric function on $(\mathbb{R}^d)^k$ for some $k \in \mathbb{N}$. Specifically, for $x = (x_1, \ldots, x_d), y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ we say that $x < y$ in the *lexicographic order* if there exists an $i = 1, \ldots, d$ such that $x_j = y_j$ for $j = 1, \ldots, i-1$ and $x_i < y_i$.

2. TOPOLOGY & PROBABILITY

In this chapter we formalize the concepts of the Introduction and give all of the required topological and probabilistic notions. A few of the results here are quoted from other sources, but included here to be self-contained. Those looking for proofs of quoted results will find them in the cited sources. We begin with investigating the topological concepts required, especially those in algebraic topology. We then proceed to the required concepts for point processes (Poisson and binomial) and conclude with the requisite theory for establishing stochastic process limits.

2.1 Concepts in topology

If we consider a topological space as a shape, it helps to break down the shape into simpler pieces that retain the "structure" of the shape. What exactly these simpler pieces are depends on the application. We take the approach of breaking our space up into collections of *simplices* called simplicial complexes. This representation allows to represent a topological space as a (much more complicated) combinatorial object, and hence we are able to compute summaries of the shape via homology, which ends up being the same regardless of whether or not our homology is calculated with singular homology or simplicial homology [44, Theorem 2.27]. Here, we begin by defining the notion of an *abstract simplicial complex* first, and then proceed to specific instances which embed in more familiar Euclidean spaces. The abstractness comes from the fact that it is not necessarily associated with any embedding into Euclidean space.

Definition 2.1.1. *If $\mathcal{X}$ is a finite set, then $\mathcal{K} = \mathcal{K}(\mathcal{X})$ is an abstract simplicial complex if it is composed of subsets of $\mathcal{X}$ which satisfy*

$$\sigma \in \mathcal{K} \text{ and } \tau \subset \sigma, \text{ then } \tau \in \mathcal{K}.$$

There are various other notions that are necessary to mention in conjunction with an (abstract) simplicial complex $\mathcal{K}$. A subcomplex is a simplicial complex with $\mathcal{K}_0 \subset \mathcal{K}$. We call a k-simplex any element $\sigma \in \mathcal{K}$ such that its cardinality $|\sigma| = k + 1$. The dimension of a simplex σ is defined as $\dim(\sigma) = |\sigma| - 1$. The dimension of a simplicial complex $\mathcal{K}$ is equal to the maximum dimension of the simplices in $\mathcal{K}$, i.e., $\dim(\mathcal{K}) = \max\{\dim(\sigma) : \sigma \in \mathcal{K}\}$. The k-skeleton of $\mathcal{K}$, is the subcomplex of $\mathcal{K}$ consisting of all simplices with dimension less than or equal to k, denoted $\mathcal{K}^{(k)} := \{\sigma \in \mathcal{K} : \dim(\sigma) \leq k\}$. We call the union of all simplices of $\mathcal{K}$ the vertex set and denote it $V(\mathcal{K})$. A simplicial complex $\mathcal{K}$ is *finite* if $|V(\mathcal{K})| < \infty$.

Furthermore, let us define $S_k(\mathcal{K}) := |\{\sigma \in \mathcal{K} : \dim(\sigma) = k\}|$ to be number of k-simplices in $\mathcal{K}$. From this simple concept, we can define the *Euler characteristic* of a simplicial complex $\mathcal{K}$ as

$$\chi(\mathcal{K}) := \sum_{k=0}^{\dim(\mathcal{K})} (-1)^k S_k(\mathcal{K}). \tag{2.1}$$

A final notion to discuss is that of a *geometric realization* of an abstract simplicial complex $\mathcal{K}$. Setting $m = |V(\mathcal{K})|$, the (canonical) geometric realization of $\mathcal{K}$ is denoted $\mathrm{geom}(\mathcal{K})$ and can be specified by a bijection $\varphi : V(\mathcal{K}) \to \{1, \dots, m\}$ such that

$$\mathrm{geom}(\mathcal{K}) = \bigcup_{\sigma \in \mathcal{K}} \mathrm{conv}(\sigma),$$

where $\mathrm{conv}(\sigma)$ is the convex hull of $\{e_{\varphi(v)}\}_{v \in \sigma}$, with each e_i, $i, \dots, m$ representing the ith canonical basis vector of $\mathbb{R}^m$ [24]. Note that there are many distinct geometric realizations of $\mathcal{K}$, e.g. any subset of $\mathbb{R}^m$ homeomorphic to $\mathrm{geom}(\mathcal{K})$. Additionally, each abstract simplicial complex of dimension d has a geometric realization in $\mathbb{R}^{2d+1}$, see [33, p. 53]. If S is a topological space, then a triangulation is a pair $(\mathcal{K}, \phi)$ where $\mathcal{K}$ an (abstract) simplicial complex and $\phi : \mathrm{geom}(\mathcal{K}) \to S$ is a homeomorphism. From here on out we ignore the distinction between the abstract simplicial complex and its geometric realization, and simply use the term "simplicial complex" to refer to either and both. Before continuing, we define a *simplicial isomorphism* of two simplicial complexes.

Definition 2.1.2. *Let $\mathcal{K}$ and $\mathcal{L}$ be simplicial complexes. A simplicial isomorphism is a bijection $\phi : V(\mathcal{K}) \to V(\mathcal{L})$ between the vertex sets of $\mathcal{K}$ and $\mathcal{L}$ such that if $\sigma \in \mathcal{K}$, then $\phi(\sigma) \in \mathcal{L}$.*

Proposition 2.5 in [20] indicates that if $\mathcal{K}$ and $\mathcal{L}$ are isomorphic, then $\mathrm{geom}(\mathcal{K})$ and $\mathrm{geom}(\mathcal{L})$ are homeomorphic.

It is important at this juncture to introduce the concept of *homology*. The idea of homology is that it generalizes the connectivity of a space to higher dimensions (i.e., beyond graph connectivity). We begin by forming a vector space $C_k(\mathcal{K})$ of k-simplices with dimension $S_k(\mathcal{K})$ whose basis elements are the k-simplices of $\mathcal{K}$ and whose elements are the formal sums (called k-chains), $c = \sum_i a_i \sigma_i$, with a_i in some field $\mathbb{F}$ and $\sigma_i \in \mathcal{K}$. Note that i runs through $1, 2, \ldots, S_k(\mathcal{K})$ and that a_i may be zero. To introduce the concept of a boundary map $\partial_k : C_k(\mathcal{K}) \to C_{k-1}(\mathcal{K})$, $k \in \mathbb{N}$, which is a linear transformation between vector spaces, any k-simplex of $\mathcal{K}$ must be given an orientation and we must index each vertex in $V(\mathcal{K})$. We deem the k-simplex $\sigma = [v_0, \ldots, v_k]$, to be the (ordered) k-simplex which consists of the vertices $v_0, \ldots, v_k$. Now we define ∂_k by

$$\partial_k(\sigma) = \sum_i (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_k],$$

where $[v_0, \ldots, \hat{v}_i, \ldots, v_k] = \sigma \setminus \{v_i\}$, such that the previous ordering is preserved. Note that due to the orientation imposed on the simplices of $\mathcal{K}$, we have $\sigma = [v_0, \ldots, v_i, \ldots, v_j, \ldots, v_k] = -[v_0, \ldots, v_j, \ldots, v_i, \ldots, v_k]$, so that σ is equivalent to any even permutation of the vertices which constitute it. We can then define the kth *homology group* as a vector space

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

where $Z_k(\mathcal{K}) = \ker(\partial_k)$ and $B_k(\mathcal{K}) = \operatorname{im}(\partial_{k+1})$. Note that $H_k(\mathcal{K})$ is a quotient vector space and elements $\gamma \in H_k(\mathcal{K})$ are actually equivalence classes. Let $\gamma_1, \gamma_2 \in Z_k(\mathcal{K})$ which are called k-cycles. Then $\gamma_1 \sim \gamma_2$ if and only if $\gamma_1 = \gamma_2 + c$, for some $c \in B_k(\mathcal{K})$. Elements of $B_k(\mathcal{K})$ are called boundaries. In other words, two k cycles are equivalent if they differ by a boundary. If a k-cycle is also a boundary, it is equivalent to 0 in $Z_k(\mathcal{K})$.

Example 2.1.1. Let $\mathbb{Z}_2$ be the field on two elements $\{0, 1\}$, so that $1 + 1 = 0$. Then for the simplicial complex $\mathcal{Q}$ in Figure 2.1, let $\gamma_1 = [a, b] + [b, d] + [d, a]$, i.e. the upper left-hand triangle and $\gamma_2 = [a,b]+[b,c]+[c,d]+[d,a]$, the perimeter of the square. It is straightforward to check that $\gamma_1, \gamma_2 \in Z_1(\mathcal{Q})$. These differ by the 1-cycle

$$\partial_2([b,c,d]) = [b,c] + [c,d] + [d,b].$$

Hence, $\gamma_1 \sim \gamma_2$.

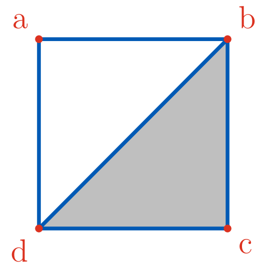


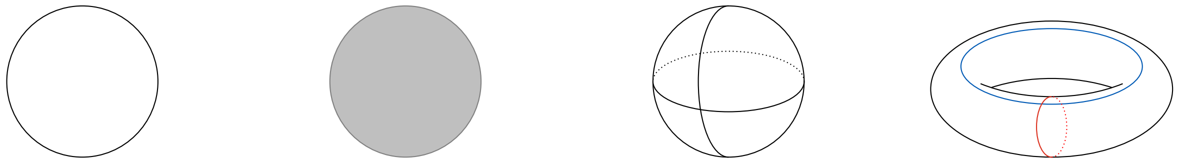
Figure 2.1. The simplicial complex $\mathcal{Q}$ consists of all the red vertices, blue edges, and the grey face $[b,c,d]$.

We denote the dimension of the quotient vector space $H_k(\mathcal{K})$ as $\beta_k(\mathcal{K})$, which is called the kth *Betti number* of $\mathcal{K}$. See Figure 2.2 below for the Betti numbers (and the Euler characteristic) of some common shapes. It is well known that if $(\mathcal{K}, \phi)$ is a triangulation of a topological space S, then $H_k(S)$ is isomorphic to $H_k(\mathcal{K})$ [44, Corollary 2.11]. This implies that $\beta_k(\mathcal{K}) = \beta_k(S)$ for all $k \geq 0$. Furthermore, an algebraic proof as in Theorem 2.44 of [44], yields that $\chi(S) = \chi(\mathcal{K})$, so that

$$\chi(\mathcal{K}) = \sum_{k=0}^{\infty}(-1)^k \beta_k(\mathcal{K}) \quad \text{and} \quad \chi(S) = \sum_{k=0}^{\infty}(-1)^k \beta_k(S).$$

The dual nature of the Euler characteristic, being either combinatorial or homological, can be seen in Figure 2.3.

Before defining the most important versions of simplicial complexes, for any set S let $\mathcal{F}(S)$ be the collection of finite (non-empty) subsets of S. First up is the *Čech complex.*



(a) $\beta_0 = 1$, $\beta_1 = 1$ and $\chi = 1 - 1 = 0$.

(b) $\beta_0 = 1$ and $\chi = 1$.

(c) $\beta_0 = 1$, $\beta_1 = 0$, $\beta_2 = 1$, and $\chi = 1 - 0 + 1 = 2$.

(d) $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 1$, and $\chi = 1 - 2 + 1 = 0$.

Figure 2.2. The object in (a) is the circle $S^1 = \{x \in \mathbb{R}^2 : \|x\| = 1\}$. The object in (b) is the disk $D^2 = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$. The surface in (c) is a 2-sphere or S^2. Finally, (d) is a 2-dimensional torus. Betti numbers not listed for the above shapes are zero.

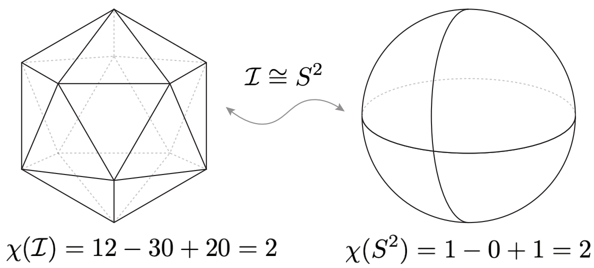


Figure 2.3. The icosahedron is a triangulation of the sphere and the Euler characteristic χ respects this homeomorphism.

Definition 2.1.3 (Čech complex)**.** *Given an $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$, the Čech complex $\check{C}(\mathcal{X}, t)$, $t \geq 0$, is a simplicial complex consisting of the subsets $\sigma \subset \mathcal{X}$ such that*

$$\sigma \in \check{C}(\mathcal{X}, t) \quad \textit{if and only if} \quad \bigcap_{x \in \sigma} B(x, t/2) \neq \emptyset.$$

We now define the *Vietoris-Rips complex*, which contains a simplex for each subset whose points are of pairwise distance at most t.

Definition 2.1.4. *Given an $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$, the Vietoris-Rips complex $\mathcal{R}(\mathcal{X}, t)$, $t \geq 0$, is a simplicial complex consisting of the subsets $\sigma \subset \mathcal{X}$ such that*

$$\sigma \in \mathcal{R}(\mathcal{X}, t) \quad \textit{if and only if} \quad B(x, t/2) \cap B(y, t/2) \neq \emptyset, \;\; x, y \in \sigma.$$

Furthermore, the definitions of both of these simplicial complexes depend only on the metric $\|\cdot\|$ on $\mathbb{R}^d$ and thus may be defined for any metric space, though we restrict our focus throughout to $\mathbb{R}^d$. Note that the above definitions imply that

$$\check{C}(\mathcal{X}, 0) = \mathcal{R}(\mathcal{X}, 0) = \{[x] : x \in \mathcal{X}\}.$$

Because of the fact that the structure of the Čech and Vietoris-Rips complexes depend on the distances between their vertices, or their geometry, we call them *geometric complexes*. In particular, if the point cloud $\mathcal{X}$ is random, then $\check{C}(\mathcal{X}, t)$ and $\mathcal{R}(\mathcal{X}, t)$ are *random* geometric complexes. An illustration of the difference between the Čech and Vietoris-Rips complexes can be seen in Figure 2.4.

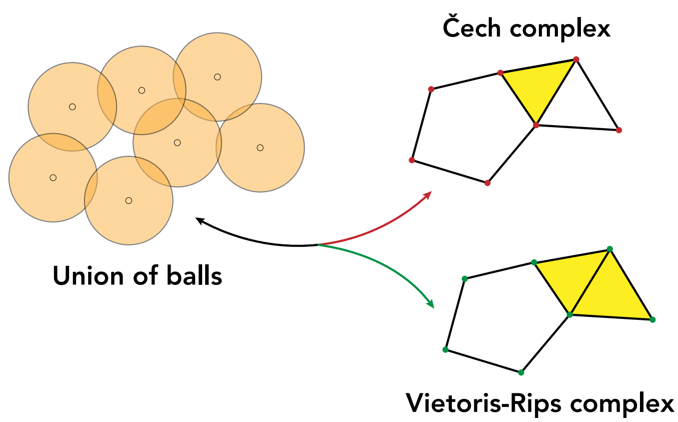


Figure 2.4. Differences between the Čech and Vietoris-Rips complexes

A seminal result about Čech complexes is the Nerve theorem [21]. The Nerve theorem, roughly stated, guarantees that the union of the balls of radius $t/2$ with centers in the finite point cloud $\mathcal{X}$ in $\mathbb{R}^d$ has the same shape (and Betti numbers) as the Čech complex $\check{C}(\mathcal{X}, t)$. We state a simplified version that suits our purposes, taken from [33].

Theorem 2.1.2 (Nerve Theorem)**.** *Given an $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$, the Čech complex $\check{C}(\mathcal{X}, t)$ is homotopy equivalent to $U(\mathcal{X}, t/2)$.*

If $\mathcal{K}$ is a finite simplicial complex, then a *filtration* of simplicial complexes is a (finite) sequence of nested subcomplexes of $\mathcal{K}$, i.e.

$$\mathcal{K}_1 \subset \cdots \subset \mathcal{K}_m = \mathcal{K}.$$

Note that we only consider sequences that are finite in length, though we could easily extend to infinite sequences. We may also index the filtration by $[0, \infty]$, so the filtration is denoted $\{\mathcal{K}_t : t \geq 0\}$, and $\mathcal{K}_t$ is constant on (t_i, t_{i+1}), $0 = t_0 < t_1 < \cdots < t_m < t_{m+1} = \infty$. Now, both $\check{C}(\mathcal{X}) := \{\check{C}(\mathcal{X}, t) : t \geq 0\}$ and $\mathcal{R}(\mathcal{X}) := \{\mathcal{R}(\mathcal{X}, t) : t \geq 0\}$ are filtrations, because $\mathcal{X}$ is finite (see Lemma 2.1.4 below). In fact they are both κ*-filtered simplicial complexes* [23, 46]. That is, there exists a *filtration function* $\kappa : \mathcal{F}(\mathbb{R}^d) \to [0, \infty)$ such that

$$\mathcal{K}(\mathcal{X}, t) = \{\sigma \subset \mathcal{X} : \kappa(\sigma) \leq t\}$$

for each $t \geq 0$, where $\mathcal{K} = \check{C}$ or $\mathcal{R}$. If σ is a simplex in a κ-filtered complex, we deem $\kappa(\sigma)$ to be the *filtration time* of σ. In addition, κ-filtered simplicial complexes are nondecreasing in t, i.e. $\mathcal{K}(\mathcal{X}, s) \subset \mathcal{K}(\mathcal{X}, t)$, for $s \leq t$. We offer a brief proof of the fact that Čech and Vietoris-Rips filtrations are both κ-filtered simplicial complexes.

Proposition 2.1.1. *For any finite $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$, both $\check{C}(\mathcal{X})$ and $\mathcal{R}(\mathcal{X})$ are both κ-filtered simplicial complexes, with filtration functions*

$$\kappa_{\check{C}}(\sigma) = \inf_{x \in \mathbb{R}^d} \max_{y \in \sigma} 2\|x - y\|,$$

and

$$\kappa_{\mathcal{R}}(\sigma) = \max_{x, y \in \sigma} \|x - y\| =: \operatorname{diam}(\sigma).$$

Proof. The case for the Vietoris-Rips complex follows directly from the definition. Clearly, $\check{C}(\mathcal{X}, t) \subset \{\sigma \subset \mathcal{X} : \kappa_{\check{C}}(\sigma) \leq t\}$. Now, suppose $\sigma \subset \mathcal{X}$ satisfies

$$\inf_{x \in \mathbb{R}^d} \max_{y \in \sigma} 2\|x - y\| \leq t.$$

Then for any $n \in \mathbb{N}$ there exists a point $x \in \mathbb{R}^d$ such that $\max_{y\in\sigma}\|x-y\| \leq (t+\frac{1}{n})/2$. Hence,

$$B_n := \bigcap_{y\in\sigma} B\Big(y, (t+\frac{1}{n})/2\Big) \neq \emptyset.$$

Define x_n to be some point in B_n. Then for any N, B_N is compact and there exists a convergent subsequence of $\{x_n\}_{n\geq N}$ such that $x_{n_m} \to x \in B_N$, as $m \to \infty$ (a property of compactness in metric spaces, see Theorem 28.2 [62]). As any subsequence of $\{x_{n_m}\}$ must also converge to x, we have that $x \in B_N$ for all $N \in \mathbb{N}$. Hence, $x \in \cap_{y\in\sigma} B(y, t/2)$ so $\sigma \in \check{C}(\mathcal{X}, t)$. □

Thus, we may index $\check{C}(\mathcal{X})$ and $\mathcal{R}(\mathcal{X})$ by the distinct filtration times of their constituent simplices, instead of by the integers.

An important property of κ-filtered simplicial complexes is that they are *right-continuous*. A filtration $\{\mathcal{K}_t : t \geq 0\}$ is said to be right-continuous if $\cap_{t>s}\mathcal{K}_t = \mathcal{K}_s$ for all $s \geq 0$.

Lemma 2.1.3. *For any finite $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$, the κ-filtered simplicial complex $\{\mathcal{K}(\mathcal{X}, t) : t \geq 0\}$ is right-continuous.*

Proof. It can be shown that $\{\mathcal{K}(\mathcal{X}, t) : t \geq 0\}$ is a filtration as defined above as their are only finitely many simplices and hence finitely many distinct filtration times, which we denote $0 = r_0 < r_1 < \cdots < r_m < r_{m+1} = \infty$. For $s \in (r_i, r_{i+1})$ right-continuity holds as the filtration is constant. For $s = r_i$, $i = 0, \ldots, m$ and $t \in [r_i, r_{i+1})$ we have that $\mathcal{K}(\mathcal{X}, t) = \mathcal{K}(\mathcal{X}, s)$ by definition of an κ-filtered simplicial complex. Hence, $\{\mathcal{K}(\mathcal{X}, t) : t \geq 0\}$ is right-continuous. □

Let $\mathcal{G}$ be the collection of abstract simplicial complexes with vertex set $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$. That is if $\mathcal{K} \in \mathcal{G}$, then $|V(\mathcal{K})| < \infty$. An example of a (simple) element of $\mathcal{G}$ for $d = 2$ is the simplicial complex

$$\mathcal{K} = \{\{(1.2, 5.024)\}, \{(5.833, 6.7)\}, \{(4, \pi)\}, \{(1.2, 3.236), (4, \pi)\}\}, \tag{2.2}$$

which of course contains the empty (-1)-simplex, though we omit it from the representation[1].

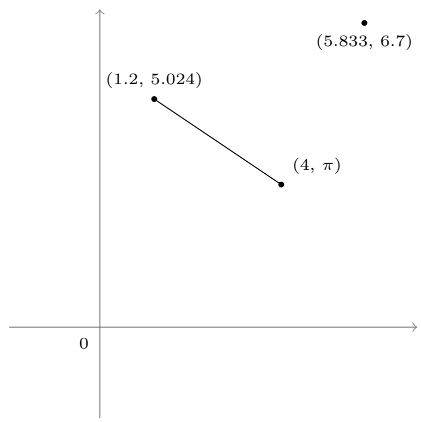


Figure 2.5. Our humble abstract simplicial complex $\mathcal{K}$ from (2.2).

In some instances, only certain properties of the Čech and Vietoris-Rips complexes are essential for proofs, but to obtain limiting results there are more conditions necessary than a generic κ-filtered simplicial complex, so we define an intermediate object called a *simplicial construction function* in Definition 2.1.5. This construction is similar to that of [46].

Definition 2.1.5. *A* simplicial construction function *is a function $\mathcal{K} : \mathcal{F}(\mathbb{R}^d) \times [0, \infty) \to \mathcal{G}$ that assigns a simplicial complex to each finite, non-empty subset $\mathcal{X} \subset \mathbb{R}^d$ and nonnegative real number. Furthermore, $\mathcal{K} = \mathcal{K}(\kappa)$ is induced by a filtration function $\kappa : \mathcal{F}(\mathbb{R}^d) \to [0, \infty)$ so that $\{\mathcal{K}(\mathcal{X}, t) : t \geq 0\}$ is a κ-filtered simplicial complex. Finally, $\mathcal{K}$ and κ are also required to satisfy the following properties for $t \geq 0$:*

1. *For every $x \in \mathbb{R}^d$, $\kappa(\{x\}) = 0$,*

2. *$\mathcal{K}$ is* locally determined*, i.e., there exists a $c > 0$ such that*

$$\sigma \in \mathcal{K}(\mathcal{X}, t) \quad \text{then} \quad \operatorname{diam}(\sigma) \leq ct, \tag{2.3}$$

3. *$\mathcal{K}$ is translation invariant in $\mathcal{X}$, i.e. $\mathcal{K}(\mathcal{X} + z, t)$ is isomorphic to $\mathcal{K}(X, t)$ for any $z \in \mathbb{R}^d$,*

[1] ↑Note that we have written elements of $\mathcal{K}$ as sets, rather than using the bracket notation [ ], as above.

Of course, by virtue of Lemma 2.1.3 a simplicial construction function is right-continuous in t for any fixed $\mathcal{X}$. Condition 1 implies that the 0-simplices of $\mathcal{K}(\mathcal{X}, t)$ are the singletons of $\mathcal{X}$. It is important to also make clear that Condition 2 in Definition 2.1.5 is relative to whichever norm that we place on $\mathbb{R}^d$. Furthermore, Condition 2 implies that $\mathcal{K}(\mathcal{X}, t) \subset \mathcal{R}(\mathcal{X}, ct)$. When a given simplicial construction function $\mathcal{K}$ (along with a filtration function κ) is fixed, define the indicator function

$$h_t^k(\mathcal{Y}) := 1\Big\{\kappa(\mathcal{Y}) \leq t\Big\} 1\Big\{|\mathcal{Y}| = k+1\Big\}, \tag{2.4}$$

for any $\mathcal{Y} \in \mathcal{F}(\mathbb{R}^d)$.

We offer the following "obvious" lemma, which we have alluded to above, to show that $t \mapsto \mathcal{K}(\mathcal{X}, t)$ cannot vary too much for any fixed $\mathcal{X}$.

Lemma 2.1.4. *Let $\mathcal{K}$ be a simplicial construction function and fix an element $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$. Then $t \mapsto \mathcal{K}(\mathcal{X}, t)$ takes only finitely many values in $\mathcal{G}$.*

Proof. Every simplicial complex with vertex set $\mathcal{X}$ is a subset of the power set $2^{\mathcal{X}}$, which has cardinality $2^{|\mathcal{X}|}$. The number of such subsets has cardinality $2^{2^{|\mathcal{X}|}} < \infty$. ☐

The above lemma also implies that for any simplicial construction function $\mathcal{K}$, fixed $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$, and function $F : \mathcal{G} \to \mathbb{R}$, the map $t \mapsto F(\mathcal{K}(\mathcal{X}, t))$ has only finitely many discontinuities and is piecewise constant. It is worth noting that the above discussion is a drastic simplification of the concept of *tameness* of persistence diagrams, a beautiful and theoretical exposition of which can be seen in [25].

We deem any function $F : \mathcal{G} \to \mathbb{R}$ a *topological functional* on $\mathbb{R}^d$. Examples include the Betti numbers β_k and the Euler characteristic χ. Our two most prominent simplicial complexes and filtrations, the Čech and Vietoris-Rips complexes, are examples of right-continuous simplicial construction functions.

We continue with a lemma about the minimal number of vertices required in a simplicial complex to support k-dimensional homology, which is important for Chapter 3.

Lemma 2.1.5. *Let $\mathcal{K}$ be a simplicial complex with $k+2$ vertices and with dimension at most $k+1$. Then $\beta_k(\mathcal{K}) = 1$ if and only if all possible k-simplices are present and there is no $k+1$ simplex.*

Proof. We first prove that if all possible k-simplices are present and there is no $(k+1)$-simplex then $\beta_k(\mathcal{K}) = 1$. Suppose that the vertices of $\mathcal{K}$ are labelled $\sigma = [v_0, \ldots, v_{k+1}]$. As there is no $(k+1)$-simplex, i.e., $\sigma \notin \mathcal{K}$, then $\operatorname{im} \partial_{k+1} = 0$. Therefore we must show that $\ker \partial_k \neq 0$. Now, we have $c = \partial_{k+1}\sigma = \sum_i (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_{k+1}] \in C_k(\mathcal{K})$, by hypothesis. Then by the "fundamental lemma of homology", $\partial_k c = 0$ (see p. 81 in [33]). Thus $\beta_k(\mathcal{K}) > 0$.

We will now show that the above implies that $\dim(\ker \partial_k) = 1$. To do so, it suffices to show that $\{c\}$ spans $\ker \partial_k$, so that $\dim(\ker \partial_k) \leq 1$. This means for any $d \in \ker \partial_k$ there exists an $\alpha \in \mathbb{F}$ such that $\alpha c = d$. Now, $d = \sum_i \alpha_i [v_0, \ldots, \hat{v}_i, \ldots, v_{k+1}]$ by definition of C_k. Thus,

$$\partial(d) = \sum_{i<j} \alpha_i (-1)^j [v_0, \ldots, \hat{v}_i, \ldots, \hat{v}_j, \ldots, v_{k+1}] + \sum_{i>j} \alpha_i (-1)^j [v_0, \ldots, \hat{v}_j, \ldots, \hat{v}_i, \ldots, v_{k+1}],$$

and the second sum is equal to

$$-\sum_{i<j} \alpha_j (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, \hat{v}_j, \ldots, v_{k+1}],$$

so that $\partial(d) = 0$ is equivalent to

$$\sum_{i<j} [\alpha_i (-1)^j - \alpha_j (-1)^i][v_0, \ldots, \hat{v}_i, \ldots, \hat{v}_j, \ldots, v_{k+1}] = 0, \tag{2.5}$$

which can only happen if for every $0 \leq i < j \leq k+1$ we have $\alpha_i(-1)^j = \alpha_j(-1)^i$, by the linear independence of the basis vectors $[v_0, \ldots, \hat{v}_i, \ldots, \hat{v}_j, \ldots, v_{k+1}] \in C_{k-1}$. Thus, $\alpha_i = (-1)^i \alpha_0$ for all i, so if we let $\alpha = \alpha_0$, we have our result.

Now we demonstrate necessity. First $\dim(\mathcal{K}) = k+1$ implies that $\sigma \in \mathcal{K}$ hence $\beta_k(\mathcal{K}) = 0$. Thus $\beta_k(\mathcal{K}) > 0$ implies that $\dim(\mathcal{K}) \leq k$. Suppose that for any fixed index ℓ we have

$[v_0, \ldots, \hat{v}_\ell, \ldots, v_{k+1}] \notin C_k$, so for any $d \in \ker(\partial_k)$ we have $\alpha_\ell = 0$ in (2.5). Then we must have $\alpha_j = 0$ for all $j \neq \ell$ as well. Hence $\ker(\partial_k) = 0$, which is impossible as $\beta_k(\mathcal{K}) \neq 0$. Thus, all k-simplices are present. $\square$

Remark 2.1.6. In light of the above Lemma, it is worth noting that if $\mathcal{K}$ has k or fewer vertices, then $\beta_k(\mathcal{K}) = 0$ necessarily as $\beta_k(\mathcal{K}) \leq S_k(\mathcal{K}) = 0$. If $\mathcal{K}$ has $k+1$ vertices, it has at most one k-simplex. There is one non-zero vector in $\operatorname{im} \partial_k$, hence $\dim(\ker \partial_k) = 0$, by the rank-nullity theorem, so $\beta_k(\mathcal{K}) = 0$ as well.

2.2 Point processes

Though it is often convenient to work with point clouds, we also will work with measures whose support is equal to a given point cloud. The point clouds from which we derive a Čech or Vietoris-Rips complex are taken to be either Poisson or binomial, and inhomogeneous (or nonstationary). As a disclaimer, we note that theory of point processes below exists in enough generality for us to describe it in terms of any complete, separable metric space. However, the setting for this dissertation is entirely Euclidean so the materials that are introduced are kept to this domain. To begin defining these point processes, we need to fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathbb{N}_0$ denote the nonnegative integers and $\bar{\mathbb{N}}_0$ be equal to $\mathbb{N}_0 \cup \{\infty\}$. Equip $\mathbb{R}^d$ with the standard Borel σ-algebra, which we denote $\mathcal{B}(\mathbb{R}^d)$. Let δ_x be the Dirac measure at $x \in \mathbb{R}^d$ so that for any Borel set A we have

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Furthermore, take $\mathbf{N}$ to be set of simple, Radon measures of the form $\sum_{i=1}^{M} \delta_{x_i}$, where $x_i \in \mathbb{R}^d$ for $i = 1, \ldots, M$ and $M \in \bar{\mathbb{N}}_0$. A measure $\mu \in \mathbf{N}$ is called *simple* if $\mu(\{x\}) \leq 1$ for all $x \in \mathbb{R}^d$ and Radon if $\mu(K) < \infty$ for any compact set $K \subset \mathbb{R}^d$. Hence for each measure $\mu \in \mathbf{N}$ the constituent Dirac measures δ_{x_i} are necessarily distinct. The support of a measure $\mu \in \mathbf{N}$, is the set where μ is non-zero. It is denoted $\operatorname{supp}(\mu) := \{x \in \mathbb{R}^d : \mu(\{x\}) > 0\}$. When we would like to define (or emphasize) our measures to have support on a proper Borel subset

E of $\mathbb{R}^d$ for some $d \geq 1$, we write $\mathbf{N}(E)$ for $\mathbf{N}$ and endow E and $\mathbf{N}(E)$ with the appropriate subspace topologies and σ-algebras, which we will define next.

A σ-algebra generated by a subset $C \subset 2^{\mathbb{R}^d}$, is the smallest σ-algebra containing C, denoted $\sigma(C)$. If $\mathcal{T}$ is the set of all open subsets of $\mathbb{R}^d$—in other words $\mathcal{T}$ is the standard topology on $\mathbb{R}^d$—then $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{T})$. We endow $\mathbf{N}$ with the σ-algebra $\mathcal{N}$, where we have

$$C = \{\{\mu \in \mathbf{N} : \mu(B) = k\} : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R}^d)\}$$

so that $\mathcal{N} = \sigma(C)$. That is, $\mathcal{N}$ is the σ-algebra generated by the subsets of measures which take integer values on a given Borel set. We can now state that a *point process* is a measurable mapping $\eta : \Omega \to \mathbf{N}$. The intensity measure μ of a point process is a measure on the (measurable) space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, such that $\mu(B) = \mathbb{E}[\eta(B)]$ for all Borel B. If the intensity measure μ admits a density g, we call g the *intensity* of η. Further information on these concepts can be seen in Chapter 2 of [58]. Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is measurable, then for $\mu = \sum_{i=1}^M \delta_{x_i} \in \mathbf{N}$ we have

$$\mu(f) = \int_{\mathbb{R}^d} f(x)\mu(\mathrm{d}x) = \sum_{i=1}^M f(x_i).$$

In this dissertation, we describe the weak convergence of measures in $\mathbf{N}$. To describe the weak convergence of measures in $\mathbf{N}$, we must specify a means of convergence in $\mathbf{N}$ and in fact we will endow $\mathbf{N}$ with the so-called *vague* topology such that $\mathbf{N}$ is a complete and separable metric space. The vague topology on $\mathbf{N}$ can be specified by the subbasis composed of the sets

$$\{\mu \in \mathbf{N} : s < \mu(f) < t\}$$

where $0 \leq s < t < \infty$ and f is a nonnegative function that vanishes outside of some compact set K [81]. The vague topology makes $\mathbf{N}$ into a Polish space, so that there exists a metric ρ such that $(\mathbf{N}, \rho)$ is a complete, separable metric space (Proposition 3.17 in [81]). The fact that $\mathbf{N}$ becomes a metric space is important because we can apply the continuous mapping theorem, as in Theorem 2.7 of [10]. As a specific case, we see that if $(\eta_n)_{n\geq 1}, \eta$ are point processes and $\eta_n \Rightarrow \eta$ in the typical sense for a metric space we have

$$f(\eta_n) \Rightarrow f(\eta),$$

for $f : \mathbf{N} \to (S, d)$ continuous, where (S, d) is a metric space. This notation will come in handy when it comes to the proof of Theorem 3.6.1.

With this in mind, we can formally define our first point process, the binomial process $\mathcal{X}_n$. Take $X_1, \ldots, X_n$ to be independent and identically distributed random variables on $\mathbb{R}^d$ with distribution F. That is, $\mathbb{P}(X_1 \in A) = F(A)$, for any Borel $A \subset \mathbb{R}^d$. Take m to be d-dimensional Lebesgue measure. Let F be absolutely continuous with respect to Lebesgue measure m, and let the density of F be denoted f. We assume only that f is essentially bounded (see Section 1.4).

The binomial point process $\mathcal{X}_n$ is one of two things: either, a subset of $\mathbb{R}^d$ consisting of the n i.i.d points $\mathcal{X}_n = \{X_1, \ldots, X_n\}$, or a point process on $\mathbb{R}^d$ with intensity nf with a binomial distribution on each Borel set B. In other words, the reason $\mathcal{X}_n$ is called a binomial process, is that

$$\mathbb{P}(\mathcal{X}_n(B) = k) = \mathbb{P}(|\mathcal{X}_n \cap B| = k) = \binom{n}{k} F(B)^k \Big(1 - F(B)\Big)^{n-k},$$

where $\mathcal{X}_n$ is represented as a random element of $\mathbf{N}$ on the left and a point cloud in $\mathbb{R}^d$ on the right. Abusing notation, this means $\mathcal{X}_n(B) = |\mathcal{X}_n \cap B| \sim \mathrm{Bin}\Big(n, F(B)\Big)$. Alternatively, we have

$$\mathcal{X}_n = \sum_{i=1}^{n} \delta_{X_i}.$$

As for the Poisson (point) process $\mathcal{P}_n$, it shares the same duality of form as the binomial process. Let $N_n \sim \mathrm{Poi}(n)$ be a Poisson random variable with parameter n. Then $\mathcal{P}_n$ is either the point cloud $\{X_1, \ldots, X_{N_n}\}$ (where $X_1, X_2, \ldots$ are i.i.d with distribution F) or a point process on $\mathbb{R}^d$ with a Poisson distribution on each Borel set B. Namely,

$$\mathbb{P}(\mathcal{P}_n(B) = k) = \mathbb{P}(|\mathcal{P}_n \cap B| = k) = \frac{e^{-nF(B)}(nF(B))^k}{k!}.$$

Similarly to the binomial process $\mathcal{X}_n$, we have $\mathcal{P}_n(B) = |\mathcal{P}_n \cap B| \sim \mathrm{Poi}\Big(nF(B)\Big)$ and $\mathcal{P}_n = \sum_{i=1}^{N_n} \delta_{X_i}$. An additional property of Poisson processes in general, which is crucial to many of our proofs, is the fact that they are *completely independent*. This implies that for $B_1, \ldots, B_m$ disjoint, the random variables $\mathcal{P}_n(B_1), \ldots, \mathcal{P}_n(B_m)$ are independent. We generically denote Φ_n to be either $\mathcal{X}_n$ or $\mathcal{P}_n$.

Above, we defined a simplicial construction function in terms of the point clouds $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$. Equivalently, we may identify the set of finite measures $\mathbf{N}_{<\infty} \subset \mathbf{N}$ with $\mathcal{F}(\mathbb{R}^d)$ by specifying functions $g_0 : \mathcal{F}(\mathbb{R}^d) \to \mathbf{N}_{<\infty}$ and $g_1 : \mathbf{N}_{<\infty} \to \mathcal{F}(\mathbb{R}^d)$ defined by $g_0(\mathcal{X}) = \sum_{x \in \mathcal{X}} \delta_x$ and $g_1(\mu) = \mathrm{supp}(\mu)$. Clearly, $g_1 \circ g_0$ is the identity on $\mathcal{F}(\mathbb{R}^d)$. By the nature of $\mathbf{N}$, the measure $\mu \in \mathbf{N}_{<\infty}$ admits a representation $\mu = \sum_{i=1}^{M} \delta_{x_i}$, with $M \in \mathbb{N}_0$. Then by μ simple it is clear that $\mu = \sum_{x \in \mathrm{supp}(\mu)} \delta_x$ (we assume that any re-indexing of the Dirac measures of μ are equivalent) so that $g_0 \circ g_1$ is the identity on $\mathbf{N}_{<\infty}$. Therefore, we can define a simplicial construction function $\mathcal{K}$ in terms of point measures, so that $\mathcal{K}$ becomes $\mathcal{K} : \mathbf{N}_{<\infty} \times [0, \infty) \to \mathcal{G}$ defined by $\mathcal{K}(\mu, t) = \mathcal{K}(\mathrm{supp}(\mu), t)$.

Now we will state two important theorems that derive important bounds in terms of the add-one functionals. A functional F of a point process is a measurable mapping $\mathbf{N} \to \mathbb{R}$. The add-one functional $D_x F$, for $x \in \mathbb{R}^d$ is a function from $\mathbf{N}$ to $\mathbb{R}$ defined by

$$D_x F(\mu) = F(\mu + \delta_x) - F(\mu).$$

In terms of a point cloud $\mathcal{X} \subset \mathbb{R}^d$, this can be stated as $D_x F(\mathcal{X}) = F(\mathcal{X} \cup \{x\}) - F(\mathcal{X})$. Let η be a point process and assume that $\mathbb{E}[F(\eta)^2] < \infty$. We will state two inequalities relating to $\mathcal{X}_n$ and $\mathcal{P}_n$ respectively. The first is a version of the Efron-Stein inequality as seen in [91] and used in the proof of Lemma 4.2 in [100].

Lemma 2.2.1 (Efron-Stein inequality)**.** *For any functional F and binomial process $\mathcal{X}_n$ we have*

$$\mathrm{Var}(F(\mathcal{X}_n)) \leq 2n\mathbb{E}\Big[|D_{X_n} F(\mathcal{X}_n \setminus \{X_n\})|^2\Big]$$

Proof. Let X_i', $i = 1, \ldots, n$ be independent of each other and of $\mathcal{X}_n$. We begin by noting that

$$F(\mathcal{X}_n) - F\Big((\mathcal{X}_n \setminus \{X_i\}) \cup \{X_i'\}\Big), \tag{2.6}$$

have the same distribution for each $i = 1, \ldots, n$. The quantities $F(\mathcal{X}_n) - F(\mathcal{X}_n \setminus \{X_i\})$ have the same distribution for each $i = 1, \ldots, n$ as well. Then [91] gives that

$$\mathrm{Var}(F(\mathcal{X}_n)) \leq \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\Big[\Big(F(\mathcal{X}_n) - F\Big((\mathcal{X}_n \setminus \{X_i\}) \cup \{X_i'\}\Big)\Big)^2\Big].$$

By the property (2.6), we have

$$\begin{aligned}
\mathrm{Var}(F(\mathcal{X}_n)) &\leq \frac{n}{2}\mathbb{E}\Big[\Big(F(\mathcal{X}_n) - F\Big((\mathcal{X}_n \setminus \{X_n\}) \cup \{X_n'\}\Big)\Big)^2\Big] \\
&\leq \frac{n}{2}\mathbb{E}\Big[\Big(F(\mathcal{X}_n) - F(\mathcal{X}_n \setminus \{X_n\}) \\
&\qquad + F(\mathcal{X}_n \setminus \{X_n\}) - F\Big((\mathcal{X}_n \setminus \{X_n\}) \cup \{X_n'\}\Big)\Big)^2\Big] \\
&\leq 2n\mathbb{E}\Big[\Big(F(\mathcal{X}_n) - F(\mathcal{X}_n \setminus \{X_n\})\Big)^2\Big].
\end{aligned}$$

□

We also give a statement of Poincaré inequality. The version we use is from [59], albeit modified for our setting.

Lemma 2.2.2 (Poincaré inequality)**.** *For any functional F, the Poisson process $\mathcal{P}_n$ and a random variable X independent of $\mathcal{P}_n$ with density f, we have*

$$\mathrm{Var}(F(\mathcal{P}_n)) \leq n\mathbb{E}[D_X F(\mathcal{P}_n)^2]$$

Before continuing we must define an important concept that captures the difference between the distributions of random variables. The Kolmogorov distance between two (real-valued) random variables W and Z is defined by

$$d_K(W, Z) := \sup_{t\in\mathbb{R}} \Big|\mathbb{P}(W \leq t) - \mathbb{P}(Z \leq t)\Big|.$$

Of course, if $(W_n)_{n\geq 1}$ is a sequence of random variables such that $d_K(W_n, W) \to 0$ as $n \to \infty$, then clearly $W_n \Rightarrow W$.

The primary normal approximation technique that we will use throughout this dissertation is one called Stein's method. The ingenious idea behind Stein's method is that a random variable W has a standard normal distribution if and only if $\mathbb{E}[Wh(W)] = \mathbb{E}[h'(W)]$ for every Lipschitz function $h : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}[h'(Z)] < \infty$, for Z standard normal (i.e. $|h(x) - h(y)| \leq K|x - y|$ for some $K > 0$). A lovely and short proof of this fact can be seen in Lemma 3.2.4 in [74]. This result is applied to solve a difference equation and bound the Kolmogorov distance between W and another random variable Z, as seen in Theorem 2.2.3.

The utility of Theorem 2.2.3 is that subject to certain constraints on the dependence of a number of random variables, their sum obeys a central limit theorem. The dependence as stated below is in terms of a *dependency graph*. Let $(W_i)_{i\in I}$ be a collection of random variables, and I a finite index set. Define a graph $(I, \sim)$ on I in terms of an equivalence relation $\sim$ by including an edge if $i \sim j$ for $i, j \in I$. The graph $(I, \sim)$ is a dependency graph for $(W_i)_{i\in I}$ if for any disjoint subsets I_1 and I_2 of I, we have that $(W_i)_{i\in I_1}$ is independent of $(W_i)_{i\in I_2}$ if $i_1 \not\sim i_2$ for any $i_1 \in I_1$ and $i_2 \in I_2$. Theorem 2.2.3 appeared and was proved in [76] as Theorem 2.4. Recall that the degree of a vertex $i \in I$ is the number of $j \in I$, $j \neq i$ such that $i \sim j$.

Theorem 2.2.3 (Stein's Method [76])**.** *Suppose that $(W_i)_{i\in I}$ is a collection of mean-zero random variables with dependency graph $(I, \sim)$ with maximum degree $D - 1$. Set $W := \sum_{i\in I} W_i$ and suppose that $\mathrm{Var}(W) = \mathbb{E}[W^2] = 1$. Let Z be a standard normal random variable. Then there exists a constant $C > 0$ such that*

$$d_K(W, Z) \leq C\left(\sqrt{D^2 \sum_{i\in I} \mathbb{E}[|W_i|^3]} + \sqrt{D^3 \sum_{i\in I} \mathbb{E}[|W_i|^4]}\right).$$

We offer below a version of Palm theory for U-statistics of Poisson processes that was given in the Appendix of [92]. That result itself was ultimately derived from Lemma 8.1 in [67] and Theorems 1.6, 1.7 in [76]. For simplicity, take $[n] := \{1, \ldots, n\}$.

Lemma 2.2.4 (Palm theory for Poisson processes)**.** *Suppose $\mathcal{P}_n$ is a Poisson point process on $\mathbb{R}^d$ with intensity nf. Further, for every $k_i \in \mathbb{N}_0$, $i = 1, \dots, 4$, let $h_i(\mathcal{Y})$ be a real-valued measurable bounded function defined for $\mathcal{Y} \in (\mathbb{R}^d)^{k_i+1}$. Let $\mathcal{X}_i$ be a collection of $k_i + 1$ i.i.d points with density f. We have the following results:*

(i)

$$\mathbb{E}\Big[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} h_1(\mathcal{Y}_1)\Big] = \frac{n^{k_1+1}}{(k_1+1)!}\, \mathbb{E}[h_1(\mathcal{X}_1)].$$

(ii) For every $\ell \in \{0, \dots, (k_1 \wedge k_2) + 1\}$,

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} h_1(\mathcal{Y}_1) h_2(\mathcal{Y}_2)\, \mathbb{1}\Big\{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = \ell\Big\}\Big] \\
&\quad = \frac{n^{k_1+k_2+2-\ell}}{\ell!(k_1+1-\ell)!(k_2+1-\ell)!}\, \mathbb{E}\Big[h_1(\mathcal{X}_1) h_2(\mathcal{X}_2)\, \mathbb{1}\Big\{|\mathcal{X}_1 \cap \mathcal{X}_2| = \ell\Big\}\Big].
\end{aligned}$$

(iii) For every $\mathbf{b} = (b_{12}, b_{13}, \dots, b_{1234}) \in \mathbb{N}_0^{11}$, we have

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_3 \subset \mathcal{P}_n} h_1(\mathcal{Y}_1) h_2(\mathcal{Y}_2) h_3(\mathcal{Y}_3) \\
&\qquad \times \mathbb{1}\Big\{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = b_{12},\ |\mathcal{Y}_1 \cap \mathcal{Y}_3| = b_{13},\ |\mathcal{Y}_2 \cap \mathcal{Y}_3| = b_{23},\ |\mathcal{Y}_1 \cap \mathcal{Y}_2 \cap \mathcal{Y}_3| = b_{123}\Big\}\Big] \\
&= \frac{n^{k_1+k_2+k_3+3-b_{12}-b_{13}-b_{23}+b_{123}}}{\prod_{\sigma \subset [3],\, \sigma \neq \emptyset} j_\sigma!}\, \mathbb{E}\Big[h_1(\mathcal{X}_1) h_2(\mathcal{X}_2) h_3(\mathcal{X}_3) \\
&\qquad \times \mathbb{1}\Big\{|\mathcal{X}_1 \cap \mathcal{X}_2| = b_{12},\ |\mathcal{X}_1 \cap \mathcal{X}_3| = b_{13},\ |\mathcal{X}_2 \cap \mathcal{X}_3| = b_{23},\ |\mathcal{X}_1 \cap \mathcal{X}_2 \cap \mathcal{X}_3| = b_{123}\Big\}\Big],
\end{aligned}$$

where for a non-empty $\sigma \subset [3]$ and $\mathcal{X}_i \in (\mathbb{R}^d)^{k_i+1}$, $i = 1, \dots, 3$,

$$j_\sigma := \Big|\bigcap_{i \in \sigma} \Big(\mathcal{X}_i \setminus \bigcup_{j \in [3] \setminus \sigma} \mathcal{X}_j\Big)\Big|.$$

(iv) Furthermore, we have

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_3 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_4 \subset \mathcal{P}_n} h_1(\mathcal{Y}_1) h_2(\mathcal{Y}_2) h_3(\mathcal{Y}_3) h_4(\mathcal{Y}_4) \\
&\qquad \times \mathbb{1}\Big\{|\mathcal{Y}_i \cap \mathcal{Y}_j| = b_{ij},\ 1 \le i < j \le 4,\ |\mathcal{Y}_i \cap \mathcal{Y}_j \cap \mathcal{Y}_k| = b_{ijk},\ 1 \le i < j < k \le 4, \\
&\qquad\qquad |\mathcal{Y}_1 \cap \mathcal{Y}_2 \cap \mathcal{Y}_3 \cap \mathcal{Y}_4| = b_{1234}\Big\}\Big]
\end{aligned}$$

$$= \frac{n^{k_1+k_2+k_3+k_4+4-b}}{\prod_{\sigma\subset[4],\,\sigma\neq\emptyset} j_\sigma!}\,\mathbb{E}\Big[h_1(\mathcal{X}_1)h_2(\mathcal{X}_2)h_3(\mathcal{X}_3)h_4(\mathcal{X}_4)\,\mathbb{1}\Big\{|\mathcal{X}_i\cap\mathcal{X}_j| = b_{ij},\ 1\le i<j\le 4,$$

$$|\mathcal{X}_i\cap\mathcal{X}_j\cap\mathcal{X}_k| = b_{ijk},\ 1\le i<j<k\le 4,\ |\mathcal{X}_1\cap\mathcal{X}_2\cap\mathcal{X}_3\cap\mathcal{X}_4| = b_{1234}\Big\}\Big],$$

where

$$b := b_{12}+b_{13}+b_{14}+b_{23}+b_{24}+b_{34}-b_{123}-b_{124}-b_{134}-b_{234}+b_{1234}. \tag{2.7}$$

Similarly to the above, we have for a non-empty $\sigma\subset[4]$ and $\mathcal{X}_i\in(\mathbb{R}^d)^{k_i+1}$, $i=1,\dots,4$, that

$$j_\sigma := \Big|\bigcap_{i\in\sigma}\Big(\mathcal{X}_i\setminus\bigcup_{j\in[4]\setminus\sigma}\mathcal{X}_j\Big)\Big| \tag{2.8}$$

A more specialized version of Palm theory for Poisson processes is also useful when dealing with functions h that depend on the whole point process, and not just those purely local ones like those in Lemma 2.2.4. The lemma below, Lemma 2.2.5, is taken from Kahle and Meckes [52], Theorems 3.11 and 3.12.

Lemma 2.2.5 (Palm theory for Poisson processes, alternate version)**.** *Suppose $\mathcal{P}_n$ is a Poisson point process on $\mathbb{R}^d$ with intensity nf. Further, for every $j\in\mathbb{N}$ and $k_1,\dots,k_j\in\mathbb{N}_0$, let $h_i(\mathcal{Y},\mathcal{X})$, $i=1,\dots,j$ be real-valued measurable bounded function defined for pairs $(\mathcal{Y},\mathcal{X})$ where $\mathcal{Y}\in(\mathbb{R}^d)^{k_i+1}$ and $\mathcal{Y}\subset\mathcal{X}\in\mathcal{F}(\mathbb{R}^d)$. For each $i=1,\dots,j$ let $\mathcal{X}_i$ be a collection of k_i+1 i.i.d points with density f, such that X_i is independent of $\mathcal{P}_n$. We have the following results:*

(i)

$$\mathbb{E}\Big[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n} h_1(\mathcal{Y}_1,\mathcal{P}_n)\Big] = \frac{n^{k_1+1}}{(k_1+1)!}\,\mathbb{E}[h_1(\mathcal{X}_1,\mathcal{X}_1\cup\mathcal{P}_n)].$$

(ii)

$$\mathbb{E}\Bigg[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n}\cdots\sum_{\mathcal{Y}_j\subset\mathcal{P}_n}\Big(\prod_{i=1}^{j} h_i(\mathcal{Y}_i)\mathbb{1}\big\{|\mathcal{Y}_i\cap\mathcal{Y}_\ell| = 0,\ \ell\neq i\big\}\Big)\Bigg]$$

$$= \prod_{i=1}^{j}\frac{n^{k_i+1}}{(k_i+1)!}\mathbb{E}\Big[h_i(\mathcal{X}_i,\mathcal{X}_1\cup\cdots\cup\mathcal{X}_j\cup\mathcal{P}_n)\Big].$$

2.3 Stochastic process limits

As mentioned in the Introduction, the investigation of topological functionals in the literature has focused on either the limit of a sequence of random geometric complexes [14, 27, 42, 49, 51, 52, 94, 95, 99, 100], or the limiting behavior of persistence diagrams of a filtered geometric complex [46, 66, 71]. If we recall the introduction, the "stochastic process" approach could be seen in Chapter 4 of [76] where the author proved a functional central limit theorem for the number of degree k vertices of a random geometric graph. Furthermore, Owada established functional limit theorems (or at least finite-dimensional convergence) for geometric and topological functionals including subgraph counts, Betti numbers and persistence barcode lifetime sums in the EVT setup in [67, 68, 69].

We let $D := D[0,\infty)$ represent the set of all real-valued functions on $[0,\infty)$ that are right-continuous and have left limits. This means that for $x \in D[0,\infty)$ we have that

$$\lim_{t \downarrow s} x(t) = x(s),$$

and that $x(t-) := \lim_{s \uparrow t} x(s)$ exists. For $T \in [0,\infty)$, we let $D[0,T]$ be defined similarly, save for the fact that the domain of the elements in $D[0,T]$ is the compact interval $[0,T]$, as opposed to $[0,\infty)$. Stochastic process limits are highly useful for limiting distributions of pathwise properties. There are two varieties of stochastic process limits that we give in here. Let us define a stochastic process to be a measurable mapping from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(D, \mathcal{D})$, where $\mathcal{D}$ is some σ-algebra. We also call this a random element of D. Let us also equip $(D, \mathcal{D})$ with some metric ρ such that the topology induced by ρ is separable.

Let $C := C[0,\infty)$ be the subset of D consisting of continuous real-valued functions on $[0,\infty)$. We assume for now that C is a measurable subset of D. For our purposes we may take a *Gaussian process* to be a random element $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (D, \mathcal{D})$, such that $\mathbb{P}(X \in C) = 1$ and that $(X(t_1), \dots, X(t_m))$ has a (multivariate) Gaussian distribution for each $t_1, \dots, t_m \in [0,\infty)$. The variance of $X(t)$ may potentially be zero. We will denote a stochastic process on D as $(X(t), t \geq 0)$ and $D[0,T]$ as $(X(t), t \in [0,T])$.

We may now define a *functional central limit theorem.*

Definition 2.3.1. *Let $(X_n)_{n\geq 1}$ and X be mean zero stochastic processes in D. The sequence $(X_n)_{n\geq 1}$ obeys a functional central limit theorem (FCLT) if*

$$(X_n(t),\, t \geq 0) \Rightarrow (X(t),\, t \geq 0),$$

and if X is a mean zero Gaussian process.

Of course, we could prove the above weak convergence using conventional notions of weak convergence for metric spaces. This was the approach taken in the original proof of Donsker's theorem [30], however methods involving the concept of tightness are much more tractable. We will discuss our main mechanism of proving weak convergence after two further definitions—of *finite-dimensional* weak convergence and a *functional strong law of large numbers*—and a brief discussion of the metrics ρ that we employ for D.

Definition 2.3.2. *The stochastic processes $(X_n)_{n\geq 1}$ in D converge in a finite-dimensional sense to a random element X in D if for each $m \in \mathbb{N}$ and $0 \leq t_1 < \cdots < t_m < \infty$ we have*

$$(X_n(t_1), \ldots, X_n(t_m)) \Rightarrow (X(t_1), \ldots, X(t_m)),$$

weakly in $\mathbb{R}^m$. We denote this $(X_n(t),\, t \geq 0) \overset{fidi}{\Rightarrow} (X(t),\, t \geq 0)$.

Definition 2.3.3. *Let $(X_n)_{n\geq 1}$ and X be stochastic processes in D. The sequence $(X_n)_{n\geq 1}$ obeys a functional strong law of large numbers (FSLLN) if*

$$(X_n(t),\, t \geq 0) \to (X(t),\, t \geq 0), \quad \text{a.s.},$$

meaning that $\mathbb{P}(\lim_{n\to\infty} \rho(X_n, X) = 0) = 1$.

There are two associated topologies that we consider on D—the topology U, representing the uniform topology, and J_1, representing the Skorohod J_1-topology. We denote the spaces (D, U) and (D, J_1) when our result is sensitive to the topology. There are many additional topologies that one could equip D with (see chapter 12 in [97]), however U and J_1 suffice for our purposes. The topologies U and J_1 are induced by the metrics u and d respectively, seen below. Before continuing, let us define u_T for $T \geq 0$ by $u_T(x,y) := \|x(t) - y(t)\|_T$ and $\|x\|_T := \sup_{0 \leq t \leq T} |x(t)|$. We begin in earnest by defining the metric $u : D \times D \to [0, \infty)$ by

$$u(x,y) := \sum_{\ell=1}^{\infty} \frac{1}{2^\ell} \Big(u_\ell(x,y) \wedge 1 \Big).$$

It is straightforward to demonstrate that

$$u(x_n, x) \to 0 \quad \text{if and only if} \quad u_T(x_n, x) \to 0, \text{ for all } T \geq 0.$$

It is worthwhile to note that u is the canonical metric on $C := C[0,\infty)$ (see Section 2.4 in [53]). Now, Theorem 16.2 in [10] implies that

$$d(x_n, x) \to 0 \quad \text{if and only if} \quad d_T(x_n, x) \to 0, \text{ for all continuity points } T \text{ of } x,$$

where d is the Skorohod metric defined at (16.4) in [10] and d_T defined for $T \geq 0$ by

$$d_T(x,y) := \inf_{\lambda \in \Lambda_T} \Big\{ \|\lambda - \mathrm{id}\|_T \vee \|x \circ \lambda - y\|_T \Big\},$$

where Λ_T is the collection of strictly increasing, continuous mappings from $[0,T]$ onto itself and id is the identity function on $[0,T]$. If a function $f : (D, J_1) \to (S, \rho)$—where (S,ρ) some metric space—is continuous, this is equivalent to $d(x_n, x) \to 0$ implying that $\rho(f(x_n), f(x)) \to 0$ as $n \to \infty$, for any sequence $\{x_n\}_{n \geq 1}$ in D. As can be readily seen, $u_T(x_n, x) \to 0$ implies that $d_T(x_n, x) \to 0$ for all $T \geq 0$, so that $u(x_n, x) \to 0$ implies $d(x_n, x) \to 0$ as $n \to \infty$. Therefore, if $u(x_n, x) \to 0$ we have $\rho(f(x_n), f(x)) \to 0$ hence $f : (D, U) \to (S, \rho)$ is continuous as well. This means there are more u-continuous functions than d-continuous functions. If measurability is satisfied by a family of random elements $(X_n)_{n \geq 1}$ of D, it is thus more desirable to establish convergence to a random element $X \in D$ via u than with d, as we have a larger collection of continuous functions to apply to each X_n. One quick note about which sets are measurable for (D, U) and (D, J_1)—we cannot just give (D, U) the Borel σ-algebra. We follow the convention of [80] in that we equip both σ-algebras with the σ-algebra $\sigma(\pi_t, t \geq 0)$ generated by the projections $\pi_t : D \to \mathbb{R}$, $t \in [0, \infty)$. For (D, J_1) this σ-algebra and the Borel σ-algebra coincide [10, Theorem 16.6]. In various circumstances we denote (D, J_1) as D or $D[0, \infty)$. We equip $D[0, T]$ with the Borel σ-algebra.

We may now resolve the measurability question of whether C is an element of $\sigma(\pi_t, t \geq 0)$. That is, we show that C is a measurable subset of both (D, J_1) and (D, U). As

$$\begin{aligned} C &= \{x \in D : x(t) = x(t-),\, t \geq 0\} \\ &= \bigcap_{r \in \mathbb{Q} \cap [0,\infty)} \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{x \in D : |x(r) - x(r - 1/n)| \leq 1/k\}, \end{aligned} \tag{2.9}$$

we simply need to show that the set on the right-hand side of (2.9) is measurable with respect to $\sigma(\pi_t, t \geq 0)$. This follows fairly easily by resorting to a similar argument as (2.9) and noting that all projections $x \mapsto (x(t_1), \ldots, x(t_m)) \in \mathbb{R}^m$ are measurable.

In the below theorem, borrowed from [10] (Theorem 16.7) we see that convergence in $D[0, \infty)$ can be reduced to convergence in $D[0, T]$ for a special class of $T > 0$.

Theorem 2.3.1 (Theorem 16.7 in [10])**.** *Let $D[0, \infty)$ and $D[0, T]$ be equipped with the Skorohod J_1 topology for every $T > 0$. Then if*

$$(X_n(t),\, t \in [0, T]) \Rightarrow (X(t),\, t \in [0, T]), \quad \textit{in } D[0, T],$$

as $n \to \infty$, for each $T > 0$ such that $\mathbb{P}(X(T) \neq X(T-)) = 0$, then

$$(X_n(t),\, t \geq 0) \Rightarrow (X(t),\, t \geq 0), \quad \textit{in } D[0, \infty),$$

as well.

For (D, U) the Borel σ algebra and $\sigma(\pi_t, t \geq 0)$ are different (see p. 157 in [10]) and certain familiar processes like the uniform empirical process may not be measurable with respect to the Borel σ-algebra. Nevertheless, we take the approach of using (D, U) for all the functional strong laws of large numbers, because there is no harm in assuming the smaller

σ-algebra (though equal to the Borel σ-algebra for (D, J_1)) and having the greater slate of continuous functions at hand.

We now discuss two primary ways in which we establish functional weak convergence. They are adaptations of the sufficient conditions for convergence in $D[0,1]$ to $D[0,\infty)$ for special cases, from [10].

Theorem 2.3.2 (Adaptation of Theorem 13.5 in [10])**.** *Let $(X_n)_{n\geq 1}$ and X be random elements of (D, J_1), with X having continuous sample paths with probability 1. Then*

$$(X_n(t),\, t \geq 0) \Rightarrow (X(t),\, t \geq 0), \quad n \to \infty$$

if $(X_n(t),\, t \geq 0) \overset{fidi}{\Rightarrow} (X(t),\, t \geq 0)$ and if there exists a nondecreasing, continuous function $F : [0,\infty) \to \mathbb{R}$ such that for every $0 \leq T < \infty$, there exists a $C > 0$ (possibly depending on T) such that

$$\mathbb{E}\Big[|X_n(t_2) - X_n(s)|^2 |X_n(s) - X_n(t_1)|^2\Big] \leq C(F(t_2) - F(t_1))^2, \tag{2.10}$$

for all $0 \leq t_1 \leq s \leq t_2 \leq T$.

Let $T > 0$ and define $D_0 \subset D[0,T]$ to be the set of integer-valued $x \in D[0,T]$. We can apply Theorem 12.6 in [10], stated below as Theorem 2.3.3, if D_0 satisfies certain conditions.

Theorem 2.3.3 (Adaptation of Theorem 12.6 in [10])**.** *Fix $T > 0$. Suppose that a set $E \in \mathcal{B}(D[0,T])$ and that T_0 is a countable dense set in $[0,T]$. Furthermore, suppose that if $(x_n)_{n\geq 1}$ and x are elements of E and $x_n(t) \to x(t)$ for each $t \in T_0$, then $x_n \to x$ with respect to the J_1 topology. Then, if $P(X_n \in E) = P(X \in E) = 1$ for all $n \geq 1$ and*

$$(X_n(t),\, t \in T_0) \overset{fidi}{\Rightarrow} (X(t),\, t \in T_0),$$

we have

$$(X_n(t),\, t \geq 0) \Rightarrow (X(t),\, t \geq 0), \quad \text{in } D[0,T].$$

Lemma 2.3.4. *Fix $T > 0$. The set $D_0 \in \mathcal{B}(D[0,T])$. Furthermore, D_0 satisfies the conditions on E in Theorem 2.3.3 for any such T_0.*

Proof. The proof follows, mutis mutandis, from the argument on p. 137 of [10]. □

In contrast to the above theorems which discuss turning finite-dimensional weak convergence into functional weak convergence, we now discuss a way to turn a pointwise strong law of large numbers into a functional one.

Proposition 2.3.1 (Proposition 4.2 in [92])**.** *Let $(X_n,\, n \in \mathbb{N})$ be a sequence of random elements in D with nondecreasing sample paths. Suppose $\lambda : [0,\infty) \to \mathbb{R}$ is a continuous and nondecreasing function. If we have*

$$X_n(t) \to \lambda(t), \quad n \to \infty, \quad \text{a.s.} \tag{2.11}$$

for every $t \geq 0$, then it follows that

$$\sup_{t\in[0,T]} |X_n(t) - \lambda(t)| \to 0, \quad n \to \infty, \quad \text{a.s.}$$

for every $0 \leq T < \infty$. Hence, it holds that $X_n \to \lambda$ a.s. in (D, U).

Proof. Fix $0 \leq T < \infty$. Note that λ is uniformly continuous on $[0,T]$. Given $\epsilon > 0$, choose $k = k(\epsilon) \in \mathbb{N}$ such that for all $s, t \in [0,T]$,

$$|s - t| \leq 1/k \text{ implies } \Big|\lambda(s) - \lambda(t)\Big| < \epsilon. \tag{2.12}$$

Since $X_n(t)$ and $\lambda(t)$ are both nondecreasing in t, we see that

$$\begin{aligned}
\sup_{t\in[0,T]} \Big|X_n(t) - \lambda(t)\Big| &= \max_{1\leq i\leq k} \sup_{t\in[(i-1)T/k,\, iT/k]} \Big|X_n(t) - \lambda(t)\Big| \\
&\leq \max_{1\leq i\leq k} \left\{ \Big(X_n(iT/k) - \lambda((i-1)T/k)\Big) \vee \Big(\lambda(iT/k) - X_n((i-1)T/k)\Big) \right\} \\
&\leq \max_{1\leq i\leq k} \left\{ \Big(X_n(iT/k) - \lambda(iT/k)\Big) \vee \Big(\lambda((i-1)T/k) - X_n((i-1)T/k)\Big) \right\} + \epsilon
\end{aligned}$$

where the second inequality follows from (2.12). By the SLLN in (2.11), the last expression tends to ϵ almost surely as $n \to \infty$. Since ϵ is arbitrary, we can complete the proof. □

Proposition 2.3.2. *For any fixed $\mu \in \mathbf{N}$, the functions $t \mapsto \beta_k(\mathcal{K}(\mu, t))$ and $t \mapsto \chi(\mathcal{K}(\mu, t))$ are both elements of $D[0, \infty)$ for any right-continuous simplicial construction function $\mathcal{K}$.*

Proof. Lemma 2.1.4 demonstrates that $t \mapsto \mathcal{K}(\mu, t)$ takes only finitely many values. As $\cap_{t>s}\mathcal{K}(\mu, t) = \mathcal{K}(\mu, s)$ there exists some $\delta_1 > 0$ such that $\mathcal{K}(\mu, t) = \mathcal{K}(\mu, s)$ for $s < t < s + \delta_1$. These simplicial complexes are identical, and thus induce identical homeomorphism invariants, so both the above functions are right-continuous. Similarly, there exists an δ_2 such that $t \mapsto \mathcal{K}(\mu, t)$ is constant on $(s - \delta_2, s)$. Hence, by the same argument left limits exist for both $t \mapsto \beta_k(\mathcal{K}(\mu, t))$ and $t \mapsto \chi(\mathcal{K}(\mu, t))$. $\square$

3. LIMIT THEORY FOR BETTI NUMBER PROCESSES

NB: A version of this chapter has been published in *Advances in Applied Probability*, Volume 52, Issue 1, March 2020, pp. 1–31. DOI: https://doi.org/10.1017/apr.2019.50. Reproduced with permission from the Applied Probability Trust.

This section details the limiting properties of the *Betti number process*

$$(\beta_{k,n}(t),\, t \geq 0) := \Big(\beta_k(\check{C}(\Phi_n, r_n(t))),\, t \geq 0\Big),\ \ k \geq 1,$$

in the sparse and critical regimes where $r_n(t) := s_n t$ and $s_n \to 0$, as $n \to \infty$. Recall that the sparse regime corresponds to the situation where $ns_n^d \to 0$ and the complex $\mathcal{K}(\Phi_n, r_n(t))$ contains many connected components consisting of a small number of vertices, that are scattered throughout the space $\mathbb{R}^d$. Furthermore, if the radii s_n of the balls governing the formation of simplices decays to 0 more slowly, i.e., $ns_n^d \to 1$, then $\mathcal{K}(\Phi_n, r_n(t))$ belongs to the critical regime in which percolation occurs (see Penrose [76], chapter 10), and the random geometric complex contains much larger components with topological holes of various dimensions. We begin with a discussion of the required material for stating and detailing the proofs. We then proceed to a discussion of the regimes and their interpretation and follow this up with moment results. Finally, we conclude with three sections: one detailing each of the major theorems contained in this section.

The Betti number process and the material in this chapter were treated by the author and his advisor in the paper [72]. However, the finite-dimensional CLT and Poisson limit theorems for the sparse regime for points from a binomial process—as well as the functional convergence in Theorem 3.6.1—are new. Many of the passages below are reproduced below nearly verbatim from this article. This work results of the aforementioned work were extended in the article by Krebs and Polonik [56] and a similar approach was taken, in a statistical direction, in [11]. In this chapter, we present limit theorems for the Betti number process in the sparse and critical regimes. Further directions for the subject matter in this chapter would be to extend the weak convergence (specifically central limit theorems) to dependent data as in [54] or extend to fully-fledged functional central limit theorems as in [55].

Throughout this chapter we take the norm $\|\cdot\|$ to be Euclidean.

3.1 Setup

In this section we only consider $\mathcal{K} = \check{C}$. An important notion in the entirety of the sparse regime, is that of the order of the smallest component that will support a nontrivial cycle in the kth homology vector space H_k. For the Čech complex to satisfy $\beta_k(\check{C}(\mathcal{X}, t)) > 0$ it is necessary for $\mathcal{X}$ to contain at least $k+2$ points. Lemma 2.1.5 demonstrates that if $|\mathcal{X}| = k+2$ then $\beta_k(\check{C}(\mathcal{X}, t)) > 0$ if and only if all possible k-simplices are present in $\check{C}(\mathcal{X}, t)$ and no $k+1$-simplex is present. In this case $\beta_k(\check{C}(\mathcal{X}, t)) = 1$. For the Vietoris-Rips complex, it is necessary for $\mathcal{X}$ to contain at least $2k+2$ points, for $\beta_k(\mathcal{R}(\mathcal{X}, t)) > 0$ (see Lemma 5.3 [50]). Throughout this section, we focus only on the Čech complex setup, though analogous results for the Vietoris-Rips complex could be achieved by appealing to the arguments for random geometric graphs or that of [69], at least in the sparse regime. The limits for Betti numbers of Vietoris-Rips complexes were addressed in [56] for the critical regime and [52] for the entirety of the sparse regime, though most of the groundwork for the latter paper was laid in Chapter 3 of Penrose's monograph on random geometric graphs [76].

We begin by defining the required notation for this chapter. We start of the notion of an *empty* $k+1$-simplex. An empty $(k+1)$-simplex is a set of $\mathcal{X}$ of $k+2$ points such that $\beta_k(\check{C}(\mathcal{X}, t)) = 1$. Namely for $x_0, \dots, x_k + 1 \in \mathbb{R}^d$, $h_t(x_0, \dots, x_{k+1}) = 1$ if and only if $\check{C}(\mathcal{X}, t)$ contains all possible k-simplices and no $k+1$-simplex. Hence the term "empty". This terminology is taken from [52]. We can decompose $h_t = h_t^+ - h_t^-$, where

$$h_t^+(x_0, \dots, x_{k+1}) := \prod_{i=0}^{k+1} 1\left\{ \bigcap_{j=0, j\neq i}^{k+1} B(x_j, t/2) \neq \emptyset \right\},$$

and

$$h_t^-(x_0, \dots, x_{k+1}) := 1\left\{ \bigcap_{j=0}^{k+1} B(x_j, t/2) \neq \emptyset \right\}.$$

Highly common to proofs in the study of random geometric complexes is decomposing your functional into monotone parts. This is true of every chapter—for each of the Euler characteristic processes—and of course here. Indeed, for any $0 \leq s \leq t < \infty$ we have

$$h_s^{\pm}(x_0, \ldots, x_k) \leq h_t^{\pm}(x_0, \ldots, x_k).$$

This decomposition has important implications for the limiting process throughout the sparse regime. In looking at $\beta_k(\check{C}(\Phi_n, r_n(t)))$ it is useful to decompose this as the sum of components of a certain size. First, recall that $\beta_{k,n}(t) \equiv \beta_k(\check{C}(\Phi_n, r_n(t)))$ and take $G_{k,n}(t)$ to be the number of empty $(k+1)$-simplex components in $\check{C}(\Phi_n, r_n(t))$. Now, for $i \geq k+2$ and $j > 0$ we define $U_{i,j,n}(t)$ to be the number of connected components C of $\check{C}(\Phi_n, r_n(t))$ with cardinality i such that $\beta_k(C) = j$. Thus, we can represent $\beta_{k,n}(t)$ (the kth Betti number of $\check{C}(\Phi_n, r_n(t))$ as

$$\beta_{k,n}(t) := \sum_{i \geq k+2} \sum_{j \geq 0} j U_{i,j,n}(t), \quad t \geq 0. \tag{3.1}$$

The representation in (3.1) was used in [51] to prove a CLT (mentioned above) of a non-functional version of $\beta_{k,n}(t)$ in the sparse regime. We note that $G_{k,n}(t) = U_{k+2,1,n}(t)$ (again by Lemma 2.1.5) so that we have (3.1) is equal to

$$\beta_{k,n}(t) = G_{k,n}(t) + \sum_{i > k+2} \sum_{j > 0} j U_{i,j,n}(t), \quad t \geq 0. \tag{3.2}$$

We denote the righthand side of this sum as $R_{k,n}(t)$. Additionally, it is helpful to introduce further notation for arbitrary finite point sets $\mathcal{Y} \subset \mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$:

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

In particular, denote

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

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

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

where $\mathrm{LMP}(\mathcal{Y})$ is the least point (leftmost point) of the set $\mathcal{Y}$ in the lexicographic ordering on $\mathbb{R}^d$.

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

$$\beta_{k,n,A}(t) = \sum_{i \geq k+2} \sum_{j>0} jU_{i,j,n,A}(t) = G_{k,n,A}(t) + \sum_{i>k+2} \sum_{j>0} jU_{i,j,n,A}(t), \tag{3.3}$$

where, in all functions above, we require the leftmost point of every subset $\mathcal{Y}$ to be an element of A. As with $R_{k,n}(t)$ we let $R_{k,n,A}(t)$ be shorthand for $\sum_{i>k+2} \sum_{j>0} jU_{i,j,n,A}(t)$.

Note that any time we discuss a result having to do with the Čech complex, the result also applies to the homotopy equivalent *alpha complex*, a useful simplicial complex construction for computation due to its significantly fewer simplices (see Chapter 6 in [20]).

3.2 The regimes

Here we introduce the "regimes" of behavior of $\check{C}(\Phi_n, r_n(t))$ in a more rigorous fashion than we have done so far. The behavior of $(\beta_k(\check{C}(\Phi_n, r_n(t))), t \geq 0)$ can be described wholly in terms of $G_{k,n}(t)$ and $R_{k,n}(t)$. We note that $\mathbb{E}[G_{k,n}(t)] = \Theta(n^{k+2}s_n^{d(k+1)})$. This follows standard arguments in the theory of random geometric graphs (for example Proposition 3.2 in [76]) or can be seen explicitly in [51, 52]. Furthermore, $\mathbb{E}[R_{k,n}(t)] = \Theta(n^{k+3}s_n^{d(k+2)})$ for either the sparse or critical regime, as can be seen by the proof of the asymptotic expectation of the first moment of $\beta_{k,n}(t)$ for Proposition 3.3.1 below. Note that in the sparse regime we can take n large enough so that ns_n^d is less than chosen value. Denote

$$\rho_n := n^{k+2}s_n^{d(k+1)}$$

. Clearly, $G_{k,n}(t)$ dominates the contribution to the Betti numbers throughout the sparse regime, whether the process obeys a Poisson or central limit theorem. It is perhaps unremarkable in retrospect that techniques from random geometric graphs find such fruitful applications to the study of *Betti numbers* in the sparse regime. This phenomenon is due to the behavior of $R_{k,n}(t)$, which we shorten in Table 3.1 to $R_{k,n}$ for some $t > 0$ satisfying the conditions of the theorem in that regime. A proof of this phenomenon for the sparse regime can be seen at (3.45).

Table 3.1. Behavior of Betti numbers on higher order ($> k+2$) components.

Regime	Behavior of $R_{k,n}$
$ns_n^d \to 0,\ \rho_n \to 1$	$\mathbb{E}[R_{k,n}] \to 0$
$ns_n^d \to 0,\ \rho_n \to \infty$	$\mathbb{E}[R_{k,n}]/\rho_n \to 0$
$ns_n^d \to 1$	$\mathbb{E}[R_{k,n}] = \Theta(n)$

As you can see, in the sparse regime when the limit of $(\beta_k(\check{C}(\Phi_n, r_n(t))), t \geq 0)$ is Poisson, the higher order components go to zero in probability, without any normalization. When there is a CLT for the process-level Betti number in the sparse regime, the empty $k+1$-simplices dominate the contribution to $\beta_{k,n}(t)$. When in the critical regime, there are components of all orders appearing, and so one needs to restrict how large these components are. In [72] and below, this is done by restricting t to be small, so that connectivity did not get too out of hand, or by restricting the size of the components by truncation. As mentioned before, percolation occurs in the random geometric graph in the critical regime, so either restriction seems to be too severe. This question was resolved in a later study by Krebs and Polonik [56] by using the ideas of stabilization of functionals from [77], in conjunction with more robust topological arguments. However, our CLT for the critical regime still stands as the most general result for a density with unbounded support and contains the most robust analytic representation of the limiting process.

Now, we can adequately go about proving the limit theorems for $(\beta_{k,n}(t),\ t \geq 0)$ in every regime. We proceed in this order: the CLT in the sparse regime in Section 3.4, the CLT in the critical regime in Section 3.5 and finally the functional central limit theorem in the Poisson regime in Section 3.6. Before proceeding the results of these theorems, we need to take a detour to discuss the limiting moments of the process $(\beta_{k,n}(t),\ t \ge 0)$. This is the topic of Section 3.3, which we discuss next.

3.3 Moment results for the Betti number process

We pause before stating the results for all regimes for a word on notation. In the sequel, we write $x + \mathbf{y} = (x + y_1, \dots, x + y_m)$ for $x \in \mathbb{R}^d$ and $\mathbf{y} = (y_1, \dots, y_m) \in (\mathbb{R}^d)^m$ for $m \in \mathbb{N}$. As mentioned above, the first step towards the required central limit theorems in the sparse and critical regimes is to examine the asymptotic moments, which we do in Proposition 3.3.1 and Proposition 3.3.2 for the critical and sparse regimes. We give a cursory treatment of the limiting first moment and covariance limits in the sparse regime, as the techniques are highly similar to those in the critical regime. We begin with the proof of Proposition 3.3.1.

Before proceeding with the proof, let us define the truncated Betti numbers

$$\beta_{k,n,A}^{(M)}(t) := \sum_{i=k+2}^{M} \sum_{j>0} j U_{i,j,n,A}(t), \quad M \in \mathbb{N} \cup \{\infty\} \tag{3.4}$$

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

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

$$\begin{aligned} \eta_{k,A}^{(i,j_1,j_2)}(t_1,t_2) := \int_{(\mathbb{R}^d)^{i-1}} \int_{\mathbb{R}^d} 1\Big\{\check{C}\Big(\{0,\mathbf{y}\}, t_1 \wedge t_2\Big) \text{ is connected}\Big\} \prod_{\ell=1}^{2} b_{j_\ell,t_\ell}(0,\mathbf{y}) \\ \times \exp\bigg(- (t_1 \vee t_2)^d f(x)\, m\Big(\mathcal{B}(\{0,\mathbf{y}\};1)\Big)\bigg) f(x)^i 1_A(x)\, \mathrm{d}x\, \mathrm{d}\mathbf{y}, \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} \nu_{k,A}^{(i_1,i_2,j_1,j_2)}(t_1,t_2) := \int_{\mathbb{R}^d} \mathrm{d}x \int_{(\mathbb{R}^d)^{i_1-1}} \mathrm{d}\mathbf{y}_1 \int_{(\mathbb{R}^d)^{i_2}} \mathrm{d}\mathbf{y}_2\, 1\Big\{\check{C}\Big(\{0,\mathbf{y}_1\}, t_1\Big) \text{ is connected}\Big\} \\ \times 1\Big\{\check{C}\Big(\mathbf{y}_2, t_2\Big) \text{ is connected}\Big\} b_{j_1,t_1}(0,\mathbf{y}_1)\, b_{j_2,t_2}(\mathbf{y}_2) \end{aligned} \tag{3.6}$$

$$\times \Big[\Big(\alpha_{t_1,t_2}\Big(\{0,\mathbf{y}_1\},\mathbf{y}_2\Big) - \alpha_{(t_1\vee t_2)/2}\Big(\{0,\mathbf{y}_1\},\mathbf{y}_2\Big)\Big) e^{-f(x)m\Big(\mathcal{B}(\{0,\mathbf{y}_1\};t_1)\cup\mathcal{B}(\mathbf{y}_2;t_2)\Big)}$$

$$- \alpha_{t_1,t_2}\Big(\{0,\mathbf{y}_1\},\mathbf{y}_2\Big) e^{-f(x)\Big\{m(\mathcal{B}(\{0,\mathbf{y}_1\};t_1))+m(\mathcal{B}(\mathbf{y}_2;t_2))\Big\}} \Big] f(x)^{i_1+i_2} 1_A(x),$$

where

$$\mathcal{B}(\mathcal{X};r) := \bigcup_{y\in\mathcal{X}} B(y;r) \tag{3.7}$$

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

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

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

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

where δ_{i_1,i_2} is again the Kronecker delta and we define $\Phi_{k,A}(t_1,t_2) := \Phi_{k,A}^{(\infty)}(t_1,t_2)$. In the case where $j_1 = j_2 = j$ and $t_1 = t_2 = t$, define $\eta_{k,A}^{(i,j)}(t) := \eta_{k,A}^{(i,j,j)}(t,t)$.

Proposition 3.3.1. *Take Φ_n to be Poisson and let f be an essentially bounded and continuous probability density. Let $ns_n^d = 1$ and $A \subset \mathbb{R}^d$ is open with $m(\partial A) = 0$.*
(i) If $M < \infty$, then for $t, t_1, t_2 > 0$,

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

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

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

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

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

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

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

Part 1: The definition in (3.1), Palm theory for Poisson processes as in Lemma 2.2.5 (i), and the monotone convergence theorem supply that

$$n^{-1}\mathbb{E}[\beta_{k,n}(t)] = \sum_{i=k+2}^{\infty} \sum_{j>0} j \, \frac{n^{i-1}}{i!} \, \mathbb{E}[g_{r_n(t)}^{(i,j)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)], \tag{3.8}$$

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

$$\begin{aligned}
&n^{i-1}\mathbb{E}[g_{r_n(t)}^{(i,j)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)] \qquad (3.9)\\
&= n^{i-1}\mathbb{E}\Big[b_{j,r_n(t)}(\mathcal{X}_i)\mathbb{E}\Big[J_{i,r_n(t)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) \,\Big|\, \mathcal{X}_i\Big]\Big] \\
&= n^{i-1} \int_{(\mathbb{R}^d)^i} 1\Big\{\check{C}(\mathbf{x}, r_n(t)) \text{ is connected}\Big\} \, b_{j,r_n(t)}(\mathbf{x}) \exp\Big(-nI_{r_n(t)}(\mathbf{x})\Big) \prod_{j=1}^{i} f(x_j) \, \mathrm{d}\mathbf{x},
\end{aligned}$$

where

$$I_{r_n(t)}(\mathbf{x}) = I_{r_n(t)}(x_0, \dots, x_{i-1}) = \int_{\mathcal{B}(\mathbf{x}; r_n(t))} f(z) \, \mathrm{d}z.$$

Subsequently we perform the change of variables $x_0 \mapsto x$ and $x_j \mapsto x + s_n y_j$ for $j = 1, \dots, i-1$, to get that (3.9) is equal to

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

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

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

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

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

It then follows from the dominated convergence theorem that

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

It remains to find a summable upper bound for (3.8) to apply the dominated convergence theorem for sums. To this end we use the inequality $j \le \binom{i}{k+1}$ which is the result of the fact that there must be a k-simplex in $\check{C}(\mathcal{X}_i, r_n(t))$ whenever $\beta_k\Big(\check{C}(\mathcal{X}_i, r_n(t))\Big) > 0$. In addition, using an obvious inequality

$$J_{i,r_n(t)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) \le 1\Big\{\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}\Big\}, \tag{3.10}$$

we get that

$$\begin{aligned} n^{-1}\mathbb{E}[\beta_{k,n}(t)] &\le \sum_{i=k+2}^{\infty} \binom{i}{k+1} \frac{n^{i-1}}{i!} \sum_{j=1}^{\binom{i}{k+1}} \mathbb{E}\Big[1\Big\{\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}\Big\}\, b_{j,r_n(t)}(\mathcal{X}_i)\Big] \qquad (3.11) \\ &\le \sum_{i=k+2}^{\infty} \binom{i}{k+1} \frac{n^{i-1}}{i!}\, \mathbb{P}\Big(\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}\Big). \end{aligned}$$

For further analysis we claim that

$$\mathbb{P}\Big(\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}\Big) \le i^{i-2}\Big(r_n(t)^d \|f\|_\infty \omega_d\Big)^{i-1}. \tag{3.12}$$

Indeed this can be derived from

$$
\begin{aligned}
&\mathbb{P}\Big(\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}\Big) \qquad (3.13)\\
&= \int_{(\mathbb{R}^d)^i} 1\Big\{\check{C}(\mathbf{x}, r_n(t)) \text{ is connected}\Big\} \prod_{j=1}^{i} f(x_j)\, \mathrm{d}\mathbf{x}\\
&= r_n(t)^{d(i-1)} \int_{(\mathbb{R}^d)^i} 1\Big\{\check{C}(\{0,\mathbf{y}\},1) \text{ is connected}\Big\}\, f(x) \prod_{j=1}^{i-1} f(x + r_n(t)y_j)\, \mathrm{d}x\, \mathrm{d}\mathbf{y}\\
&\le \Big(r_n(t)^d \|f\|_\infty\Big)^{i-1} \int_{(\mathbb{R}^d)^{i-1}} 1\Big\{\check{C}(\{0,\mathbf{y}\},1) \text{ is connected}\Big\}\, \mathrm{d}\mathbf{y}\\
&\le i^{i-2}\Big(r_n(t)^d \|f\|_\infty \omega_d\Big)^{i-1}.
\end{aligned}
$$

The last inequality comes from the basic fact that there are i^{i-2} spanning trees on i vertices. Combining (3.11), (3.12), and $ns_n^d = 1$ we conclude that

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

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

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

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

The second equality comes from an observation that if $\mathcal{Y}_1 \neq \mathcal{Y}_2$ and the intersection of $\mathcal{Y}_1$ and $\mathcal{Y}_2$ is non-empty, then $\check{C}(\mathcal{Y}_2, r_n(t_2))$ cannot be an isolated component of $\check{C}(\mathcal{P}_n, r_n(t_2))$—so these terms are zero. Appealing to Lemma 2.2.5 (ii), we get that

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

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

Applying (3.8) to each $\mathbb{E}[\beta_{k,n}(t_i)]$, $i = 1, 2$, and utilizing the independence of $\mathcal{X}_{i_1}$ and $\mathcal{X}_{i_2}$, we see that the covariance function can be written as

$$
\mathrm{Cov}(\beta_{k,n}(t_1), \beta_{k,n}(t_2)) = A_{1,n} + A_{2,n}, \tag{3.14}
$$

with

$$
A_{1,n} := \sum_{i=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^i}{i!}\, \mathbb{E}\Big[g_{r_n(t_1)}^{(i,j_1)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i,j_2)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)\Big], \tag{3.15}
$$

$$
\begin{aligned}
A_{2,n} := &\sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^{i_1+i_2}}{i_1! i_2!} \\
&\times \mathbb{E}\Big[g_{r_n(t_1)}^{(i_1,j_1)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i_2,j_2)}(\mathcal{X}_{i_2}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) \\
&\qquad - g_{r_n(t_1)}^{(i_1,j_1)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_{r_n(t_2)}^{(i_2,j_2)}(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}'_n)\Big],
\end{aligned} \tag{3.16}
$$

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

Let us denote the expectation portions of $A_{1,n}$ and $A_{2,n}$ as $E_{1,n}^{(i,\mathbf{j})}$ and $E_{2,n}^{(\mathbf{i},\mathbf{j})}$, with $\mathbf{i} = (\mathbf{i}_1, \mathbf{i}_2)$, and $\mathbf{j} = (\mathbf{j}_1, \mathbf{j}_2)$ respectively. Our goal is to show that $n^{-1}(A_{1,n} + A_{2,n})$ tends to $\Phi_{k,\mathbb{R}^d}(t_1, t_2)$ as $n \to \infty$. For now we shall compute the limits of $n^{i-1} E_{1,n}^{(i,\mathbf{j})}$ and $n^{i_1+i_2-1} E_{2,n}^{(\mathbf{i},\mathbf{j})}$ for each

i, i_1, i_2, j_1, and j_2, while temporarily assuming that the dominated convergence theorem for sums is applicable for both $n^{-1}A_{1,n}$ and $n^{-1}A_{2,n}$. By mirroring the argument from Part 1 with the same change of variables and recalling $t_1 \le t_2$,

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

Hence the assumed dominated convergence theorem for sums concludes that

$$
n^{-1}A_{1,n} \to \sum_{i=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} \frac{j_1 j_2}{i!}\, \eta_{k,\mathbb{R}^d}^{(i,j_1,j_2)}(t_1,t_2) \quad n \to \infty. \tag{3.17}
$$

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

$$
\begin{aligned}
E_{2,n}^{(\mathbf{i},\mathbf{j})} &= \mathbb{E}\Big[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_1} \cup \mathcal{X}_{i_2} \cup \mathcal{P}_n) \\
&\qquad\qquad - g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n)\Big] \\
&\quad + \mathbb{E}\Big[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n)\Big(g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n) - g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n')\Big)\Big] \\
&\quad := B_{1,n} + B_{2,n}.
\end{aligned}
$$

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

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

where $\mathcal{B}(\mathcal{X}; r)$ is defined in (3.7). Hence we have that

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

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

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

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

$$\begin{aligned} E_{2,n}^{(\mathbf{i},\mathbf{j})} &= \mathbb{E}\Big[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n) \\ &\qquad \times \Big(\alpha_{r_n(t_1), r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2}) - \alpha_{r_n(t_2)/2}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})\Big)\Big] \\ &\quad - \mathbb{E}\Big[g_1(\mathcal{X}_{i_1}, \mathcal{X}_{i_1} \cup \mathcal{P}_n) g_2(\mathcal{X}_{i_2}, \mathcal{X}_{i_2} \cup \mathcal{P}_n') \alpha_{r_n(t_1), r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})\Big] \\ &:= C_{1,n} - C_{2,n}. \end{aligned} \tag{3.18}$$

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

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

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

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

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

Therefore,

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

Assuming convergence under summation, we have that

$$
n^{-1}A_{2,n} \to \sum_{i_1=k+2}^{\infty}\sum_{i_2=k+2}^{\infty}\sum_{j_1>0}\sum_{j_2>0} \frac{j_1 j_2}{i_1! i_2!}\, \nu_{k,\mathbb{R}^d}^{(i_1,i_2,j_1,j_2)}(t_1,t_2), \quad n\to\infty. \tag{3.19}
$$

From (3.17) and (3.19), it follows that $n^{-1}(A_{1,n}+A_{2,n}) \to \Phi_{k,\mathbb{R}^d}(t_1,t_2)$ as $n\to\infty$.

Now we would like to show that both $n^{i-1}E_{1,n}^{(i,\mathbf{j})}$ and $n^{i_1+i_2-1}|E_{2,n}^{(\mathbf{i},\mathbf{j})}|$ are bounded by a summable quantity, so that application of the dominated convergence theorem for sums is valid for both $n^{-1}A_{1,n}$ and $n^{-1}A_{2,n}$. Using the bounds (3.10), (3.12), together with $ns_n^d = 1$, we have

$$
\begin{aligned}
n^{-1}A_{1,n} &\le \sum_{i=k+2}^{\infty}\sum_{j_1>0}\sum_{j_2>0} j_1 j_2 \frac{n^{i-1}}{i!}\,\mathbb{E}\Big[1\Big\{\check{C}(\mathcal{X}_i, r_n(t_1)) \text{ is connected}\Big\}\prod_{\ell=1}^{2} b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_i)\Big] \qquad (3.20) \\
&\le \sum_{i=k+2}^{\infty} \binom{i}{k+1}^2 \frac{n^{i-1}}{i!}\,\mathbb{P}\Big(\check{C}(\mathcal{X}_i, r_n(t_1)) \text{ is connected}\Big)
\end{aligned}
$$

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

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

Subsequently we turn our attention to $n^{-1}A_{2,n}$. Returning to (3.18) and using obvious relations

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

we get that

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

By virtue of this bound we have that

$$\begin{aligned} n^{-1}|A_{2,n}| &\leq 3 \sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \sum_{j_1>0} \sum_{j_2>0} j_1 j_2 \frac{n^{i_1+i_2-1}}{i_1!i_2!} \\ &\quad \times \mathbb{E}\Big[\prod_{\ell=1}^{2} 1\Big\{\check{C}(\mathcal{X}_{i_\ell}, r_n(t_2)) \text{ is connected}\Big\}\, b_{j_\ell, r_n(t_\ell)}(\mathcal{X}_{i_\ell})\, \alpha_{r_n(t_2)}(\mathcal{X}_{i_1}, \mathcal{X}_{i_2})\Big] \\ &\leq 3 \sum_{i_1=k+2}^{\infty} \sum_{i_2=k+2}^{\infty} \binom{i_1}{k+1}\binom{i_2}{k+1} \frac{n^{i_1+i_2-1}}{i_1!i_2!} \\ &\quad \times \mathbb{P}\Big(\check{C}(\mathcal{X}_{i_\ell}, r_n(t_2)) \text{ is connected for } \ell = 1,2, \\ &\qquad\qquad \mathcal{B}\big(\mathcal{X}_{i_1}; r_n(t_2)\big) \cap \mathcal{B}\big(\mathcal{X}_{i_2}; r_n(t_2)\big) \neq \emptyset\Big). \end{aligned} \tag{3.21}$$

We claim here that

$$\begin{aligned} &\mathbb{P}\Big(\check{C}(\mathcal{X}_{i_\ell}, r_n(t_2)) \text{ is connected for } \ell = 1,2,\ \mathcal{B}\big(\mathcal{X}_{i_1}; r_n(t_2)\big) \cap \mathcal{B}\big(\mathcal{X}_{i_2}; r_n(t_2)\big) \neq \emptyset\Big) \\ &\leq 2^d i_1^{i_1-1} i_2^{i_2-1}\Big(r_n(t_2)^d \|f\|_\infty \omega_d\Big)^{i_1+i_2-1}. \end{aligned} \tag{3.22}$$

To see this, by the change of variables as in (3.13), we have that

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

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

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

Now we have

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

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

We need to consider a version of the above moment and covariance results for the proof of our limit theorems in the sparse regime. This proof has been abridged owing to the strong similarities with the proof above. Note that the same results hold if Φ_n is either Poisson or binomial, owing to Lemma 3.3 in [52] or Theorem 3.2 in [49] for the expectation, and a "de-Poissonization" argument for the covariance.

Proposition 3.3.2. *Let Φ_n be Poisson or binomial and let f be an essentially bounded and continuous density function. If $ns_n^d \to 0$ and $A \subset \mathbb{R}^d$ is open with $m(\partial A) = 0$, then we have that for $t > 0$,*

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

and for $t_1, t_2 > 0$,

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

where

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

and $\mu_{k,A}(t) := \mu_{k,A}(t, t)$.

Proof. We only discuss the covariance result in the case $A = \mathbb{R}^d$, and for the case when $\Phi_n = \mathcal{P}_n$. The covariance limit proof for Throughout the proof we assume $0 < t_1 \le t_2$. We first derive the same expression as in (3.14) :

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

where $A_{1,n}$ and $A_{2,n}$ are given in (3.15), (3.16) respectively. Observing that $g_{r_n(t)}^{(k+2,j)}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n) = 0$ for all $j \ge 2$ and any $t > 0$, we can split $A_{1,n}$ into two parts, $A_{1,n} = D_{1,n} + D_{2,n}$, where

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

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

Based on this decomposition, we claim that

$$\rho_n^{-1} D_{1,n} \to \mu_{k,\mathbb{R}^d}(t_1, t_2), \quad n \to \infty, \tag{3.23}$$

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

$$\rho_n^{-1} \mathrm{Cov}\Big(G_{k,n}(t_1), G_{k,n}(t_2)\Big) \to \mu_{k,\mathbb{R}^d}(t_1,t_2), \quad n \to \infty;$$

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

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

$$\begin{aligned}
\rho_n^{-1} D_{1,n} &= \frac{\rho_n^{-1} n^{k+2}}{(k+2)!} \mathbb{E}\Big[h_{r_n(t_1)}(\mathcal{X}_{k+2}) h_{r_n(t_2)}(\mathcal{X}_{k+2}) \\
&\qquad\qquad \times \mathbb{E}[J_{k+2,r_n(t_2)}(\mathcal{X}_{k+2}, \mathcal{X}_{k+2} \cup \mathcal{P}_n)\Big| \mathcal{X}_{k+2}]\Big] \\
&= \frac{\rho_n^{-1} n^{k+2}}{(k+2)!} \int_{(\mathbb{R}^d)^{k+2}} h_{r_n(t_1)}(\mathbf{x}) h_{r_n(t_2)}(\mathbf{x}) \exp\Big(-n I_{r_n(t_2)}(\mathbf{x})\Big) \prod_{j=1}^{k+2} f(x_j)\, \mathrm{d}\mathbf{x} \\
&= \frac{1}{(k+2)!} \int_{(\mathbb{R}^d)^{k+1}} \int_{\mathbb{R}^d} h_{t_1}(0,\mathbf{y}) h_{t_2}(0,\mathbf{y}) \exp\Big(-n I_{r_n(t_2)}(x, x+s_n\mathbf{y})\Big) \\
&\qquad\qquad \times f(x) \prod_{j=1}^{k+1} f(x+s_n y_j)\, \mathrm{d}x\, \mathrm{d}\mathbf{y}.
\end{aligned} \tag{3.24}$$

By the continuity of f it holds that $\prod_{j=1}^{k+1} f(x+s_n y_j) \to f(x)^{k+1}$ a.e. as $n \to \infty$. Moreover, the exponential term converges to 1 because we see that

$$n I_{r_n(t_2)}(x, x+s_n\mathbf{y}) \le n s_n^d \|f\|_\infty m\Big(\mathcal{B}\Big(\{0,\mathbf{y}\}; t_2\Big)\Big) \to 0, \quad n \to \infty.$$

Thus (3.23) follows from the dominated convergence theorem.

Next let us turn to the asymptotics of $\rho_n^{-1} D_{2,n}$. Proceeding as in (3.20), while applying (3.10) and (3.12), we have that

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

$$\leq \frac{\left(t_1^d \|f\|_\infty \omega_d\right)^{k+1}}{\left((k+1)!\right)^2} \sum_{i=k+3}^{\infty} b_{i,n},$$

where

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

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

For the evaluation of $n^{-1}|A_{2,n}|$, we apply (3.22) to the right hand side at (3.21). Slightly changing the description of the resulting bound, we obtain

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

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

$\square$

3.4 Central limit theorem in the sparse regime

Throughout this section we assume that $ns_n^d \to 0$ and $\rho_n = n^{k+2} s_n^{d(k+1)} \to \infty$ as $n \to \infty$. The most relevant study to this section is [52], in which the central limit theorem for the sparse regime is discussed. We have extended [52] (with the erratum paper [51]) in two directions. First, we develop the process-level central limit theorem. This highlights the chief contribution of this chapter. Whereas [51, 52], as well as [100] in the ensuing section, treat the “static" topology of random Čech complexes (i.e., no time parameter t involved), the main focus of this paper is “dynamic” topology of the same complex, treating Betti numbers as a stochastic process. Second, our central limit theorem is for the entirety of the sparse regime, without requiring that $s_n = o(n^{-1/d-\delta})$ for some $\delta > 0$ as assumed in [51]. Recall the definition of $C_{f,k}$ as at (1.5).

Before presenting the main result we define the limiting stochastic process

$$\mathcal{G}_k(t) := \int_{(\mathbb{R}^d)^{k+1}} h_t(0, \mathbf{y})\, \mathfrak{G}_k(\mathrm{d}\mathbf{y}), \tag{3.25}$$

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

The significance of the characterization of the process at (3.25) is that if we define

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

then $\mathcal{G}_k^{\pm}(t)$ becomes a time-changed Brownian motion; see Proposition 3.4.1 below. Hence $\mathcal{G}_k(t) = \mathcal{G}_k^{+}(t) - \mathcal{G}_k^{-}(t)$ is a difference of two dependent time-changed Brownian motions, where dependence is due to the same Gaussian random measure $\mathfrak{G}_k$ shared by $\mathcal{G}_k^{+}(t)$ and $\mathcal{G}_k^{-}(t)$. Those wishing to examine this characterization in more detail should refer to [67]. For example, it is proven in [67] that the process $\mathcal{G}_k(t)$ is self-similar with exponent $H = d(k-1)/2$ and is Hölder continuous of any order in $[0, 1/2)$. Hölder continuity of the Euler characteristic process is proved in Theorem 4.4.1 in the next chapter.

Proposition 3.4.1. *The process $\mathcal{G}_k^{\pm}(t)$ can be expressed as*

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

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

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

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

$\square$

Our main result for the sparse regime can be seen below. The is deferred until after the proof of the central limit theorem for the critical regime in Section 3.5, as it is a straightforward variant of the proof for the critical regime. The finite-dimensional CLT for the binomial setup is demonstrated after the proof for the Poisson setup. For the proof we need to examine the asymptotic growth rate of expectations and covariances of $\beta_{k,n}(t)$. The detailed results are presented above in Proposition 3.3.2, where it is seen that the expectation and covariance both grow at the rate ρ_n.

Theorem 3.4.1. *Let $\mathcal{K} = \check{C}$, take Φ_n to be either Poisson or binomial, and suppose $\rho_n \to \infty$ with $ns_n^d \to 0$. Assume that f is an essentially bounded and continuous probability density. Then, we have the following weak convergence in the finite dimensional sense, namely*

$$\Big(\rho_n^{-1/2}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)],\, t \ge 0\Big) \stackrel{fidi}{\Rightarrow} \Big(\mathcal{G}_k(t),\, t \ge 0\Big).$$

Remark 3.4.2. Recently in [66], Owada established almost sure convergence of persistence diagrams (hence persistent Betti numbers) throughout the sparse regime, namely $\rho_n \to [0, \infty]$ with $ns_n^d \to 0$.

3.5 Central limit theorem in the critical regime

We now expand on the results of [100] (see also [95]) by offering an explicit limit of appropriately scaled moments and a central limit theorem for $\beta_{k,n}(t)$. In the critical regime, the connectivity radius s_n is defined to be $s_n = n^{-1/d}$. This sequence decays more slowly than that in the previous section; hence, Čech complexes become highly connected with many topological holes of any dimension $k < d$. More analytically, all terms in the sum (3.1) contribute to the kth Betti number, unlike in the sparse regime. This implies that the k-cycles in the limit could be supported not only on $k+2$ points but also on i points for all possible $i > k+2$.

In [100], the authors established the central limit theorem for the first time for the critical regime (though they referred to it as the "thermodynamic" regime). There are two key differences between that paper and ours. The first is that the Poisson process they consider is homogeneous with unit intensity, restricted to a set B_n such that $m(B_n) = n$. The second difference between the two, and equivalent to the contrast indicated in the sparse regime, is again that [100] treats the static topology of random Čech complexes whereas we treat the dynamic topology. As a consequence, while the weak limit in [100] is a simple Gaussian distribution with unknown variance, our limit is a Gaussian process having structure similar to that of the Betti number (3.1).

The other relevant article to our study is [95], which generalizes [100] to an inhomogeneous Poisson process case, but again only deals with static topology. We would like to emphasize that our proof techniques are significantly different from those in [95, 100]. In fact, our proof is highly analytic in nature, borrowing machinery from [76] and [52], whereas the proofs of [95, 100] rely more on the topological nature of the objects, including weakly/strongly stabilizing properties of Betti numbers, the notion of critical radius of percolation, and the theory of geometric functionals as in [78]; see also Remark 3.5.3. By virtue of our analytic approach, we can fully specify the structure of the limiting Gaussian process as in (3.26) below. This is actually the main objective of this study.

We now define the aforementioned limiting Gaussian process by

$$\mathcal{H}_k(t) = \sum_{i \geq k+2} \sum_{j>0} j \mathcal{H}_k^{(i,j)}(t), \quad t > 0, \tag{3.26}$$

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

$$\mathrm{Cov}\Big(\mathcal{H}_k^{(i_1,j_1)}(t_1), \mathcal{H}_k^{(i_2,j_2)}(t_2)\Big) = \frac{1}{i_1!}\, \eta_{k,\mathbb{R}^d}^{(i_1,j_1,j_2)}(t_1,t_2)\delta_{i_1,i_2} + \frac{1}{i_1! i_2!}\, \nu_{k,\mathbb{R}^d}^{(i_1,i_2,j_1,j_2)}(t_1,t_2). \tag{3.27}$$

Here δ_{i_1,i_2} is the Kronecker delta, and the functions $\eta_{k,\mathbb{R}^d}^{(i_1,j_1,j_2)}$, $\nu_{k,\mathbb{R}^d}^{(i_1,i_2,j_1,j_2)}$ are explicitly defined during the proof of the main theorem (see (3.5) and (3.6)). From (3.27), the covariance of $\mathcal{H}_k^{(i,j)}$ is given by

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

The main point here is that the Betti number (3.1) and the limit (3.26) are represented in a very similar fashion. In fact, the process $U_{i,j,n}(t)$ in (3.1) and $\mathcal{H}_k^{(i,j)}(t)$ in (3.26) both capture the spatial distribution of connected components C with $|C| = i$ and $\beta_k(C) = j$. In particular, $\mathcal{H}_k^{(k+2,1)}(t)$ represents the distribution of components C on $k+2$ points with $\beta_k(C) = 1$ (i.e., components of the smallest size) as does $\mathcal{G}_k(t)$ in the sparse regime. In the present regime however, many of the Gaussian processes in (3.26) beyond $\mathcal{H}_k^{(k+2,1)}(t)$, do contribute to the limit.

As a bit of a technical remark, note that for every $i \geq k+2$, there exists $j_0 > 0$ such that $b_{j,t}(x_1,\dots,x_i) = 0$ for all $j \geq j_0$, $t > 0$, and $(x_1,\dots,x_i) \subset\in \mathbb{R}^d$. In this case,

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

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

It is important to note that the growth rate of the expectation and variance of $\beta_{k,n}(t)$ is of order n—see Proposition 3.3.1. This indicates that the scaling constant for the central limit theorem must be of order $n^{1/2}$. We now give the main result for this section.

Theorem 3.5.1. *Let $\mathcal{K} = \check{C}$, take Φ_n to be Poisson or binomial, and suppose $ns_n^d \to 1$. Assume that f is an essentially bounded and continuous probability density. If $0 < t < (e\|f\|_\infty \omega_d)^{-1/d}$, then we have the following weak convergence in the finite dimensional sense, namely*

$$\left(n^{-1/2}(\beta_{k,n}(t) - \mathbb{E}[\beta_{k,n}(t)]),\, t \geq 0\right) \stackrel{fidi}{\Rightarrow} (\mathcal{H}_k(t),\, t \geq 0).$$

Remark 3.5.2. We will not prove the CLT for the binomial setup for the critical because that is treated in [95] and [56]. Additionally, a proof of de-Poissonization for truncated Betti numbers appeared in [51].

Here we assume $ns_n^d = 1$, but we could generalize it to $ns_n^d \to 1$, $n \to \infty$. Indeed, throughout the proof of Proposition 3.3.1 we will frequently encounter the integral expressions multiplied by $(ns_n^d)^{i-1}$ (e.g., (3.9)). If one assumes $ns_n^d \to 1$, the integral and $(ns_n^d)^{i-1}$ both converge. Thus, without loss of generality we may set $ns_n^d = 1$, so that we do not have to maintain $(ns_n^d)^{i-1}$ outside of the integral.

It was noted in [72] that the "the need for the restriction on t_i's seems to be a delicate issue". Of course, it is, but the restriction was artificial. In the article [100], it was shown that if a Poisson process is homogeneous, the non-functional central limit theorem holds for any fixed $t > 0$. For a "homogeneous" binomial process the value t must be restricted away from an interval in which the occupied and vacant infinite components both exist with probability 1. Additionally, the case of an inhomogeneous Poisson process, [95] has put a restriction on the value of t, despite significant difference in proof techniques with this paper. More specifically, in the notation of Theorem 3.5.1, the result of [95] indicates that t_m must be less than $r_c\|f\|_\infty^{-1/d}$, where r_c is the critical radius for percolation in a Boolean model in $\mathbb{R}^d$ (see [60] for a definition). However, as speculated in [100], this restriction can be removed.

In the aftermath of the publication of the article [72], the article [56] appeared that established *unrestricted* central limit theorems for persistent Betti numbers derived from inhomogeneous Poisson and binomial processes in the critical regime—for densities f that can be uniformly approximated by those that are piecewise uniform. The article [56] advanced the line of research in [46] which gave a CLTs for persistent Betti numbers of homogeneous Poisson processes (again in the critical regime). Both these articles used seminal results from [77] on weak and strong stabilization of functionals of point processes to establish central limit theorems. The main contribution of [56] is the strong stabilization for the "homogeneous" binomial process (n i.i.d points on an expanding hypercube)—removing the restriction mentioned earlier from [100]—and the subsequent uniform stabilization arguments of that allow them to establish finite-dimensional convergence for binomial and Poisson processes with intensity nf, for their special class of densities. An important implication of their results (and particularly Theorem 4.3) is that most of the contribution to $\beta_{k,n}(t)$ in the critical regime, regardless of whether percolation occurs or not, comes from very small loops. This argument is also used fruitfully in [42].

Further research has established functional central limit theorems, with convergence in for persistent Betti numbers generated from Čech and Vietoris-Rips complexes based off homogeneous Poisson processes on cylindrical networks [55]. This uses the extension of [10] to multidimensional parameter spaces that was introduced in [9].

Remark 3.5.3. Although Theorem 3.5.1 imposes a restriction on the range of t_i's, the central limit theorem does hold for every $t > 0$ in the case of the "truncated" Betti number

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

which itself is useful for the approximation arguments in our proof. Owing to the explicit limiting representation of the process-level Betti numbers established here, this certainly has utility in applications (for small t). However, for large t, most of the contribution to the Betti numbers will come from large components, not ones of order less than M, so an approach bounding the diameter of k-cycles, as in [11], may be more tractable for applications.

Before concluding this section we shall exploit Theorem 4.6 in [100] and present the strong law of large numbers of $\beta_{k,n}(t)$. Though the strong law of large numbers has already been proven in [42, 94] for any fixed $t > 0$, we shall state the result to highlight the novelty of our representation of the limit as the sum of contributions to the Betti number for variously sized components.

Corollary 3.5.4. *Under the condition of Theorem 3.5.1, we assume moreover that f has a compact, convex support such that $\inf_{x\in supp(f)} f(x) > 0$. Then we have, as $n \to \infty$,*

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

Proof of Corollary 3.5.4. Theorem 4.6 in [100] verified that

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

almost surely. Combining this with Proposition 3.3.1 (ii) proves the claim. □

With all of the required materials at hand, we may embark on our proofs the central limit theorems for process-level Betti numbers.

Proof of Theorem 3.5.1. We begin by proving the corresponding result for the truncated Betti number in (3.4) for every $M \in \mathbb{N}$, that is,

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

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

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

We now restrict ourselves to the case in which the corresponding left most points belong to a fixed bounded set A. By the Cramér-Wold device [31, p. 176], we need to demonstrate a univariate central limit theorem for $\sum_{i=1}^m a_i\beta_{k,n,A}^{(M)}(t_i)$, where $a_i \in \mathbb{R}$, $m \geq 1$ and $t_i \in [0,\infty)$. There is no loss of generality in assuming that $0 < t_1 < t_2 < \cdots < t_m$, The asymptotic variance of $\sum_{i=1}^m a_i\beta_{k,n,A}^{(M)}(t_i)$ scaled by $n^{-1/2}$ can be derived from Proposition 3.3.1 (i):

$$\begin{aligned}\mathrm{Var}\Big(n^{-1/2}\sum_{i=1}^m a_i\beta_{k,n,A}^{(M)}(t_i)\Big) &= \sum_{i=1}^m\sum_{j=1}^m a_ia_jn^{-1}\mathrm{Cov}(\beta_{k,n,A}^{(M)}(t_i), \beta_{k,n,A}^{(M)}(t_j)) \\ &\to \sum_{i=1}^m\sum_{j=1}^m a_ia_j\Phi_{k,A}^{(M)}(t_i,t_j), \quad n\to\infty.\end{aligned} \tag{3.28}$$

Our proof exploits Stein's normal approximation method for weakly dependent random variables, as in Theorem 2.2.3. We assume the limit in (3.28) is positive as otherwise our proof is trivial. Let $(Q_{j,n},\ j \in \mathbb{N})$ be an enumeration of almost disjoint closed cubes (i.e., their interiors are disjoint) of side length $r_n(t_m)$, such that $\cup_{j\in\mathbb{N}} Q_{j,n} = \mathbb{R}^d$.

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

and

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

so that $\sum_{i=1}^m a_i \beta^{(M)}_{k,n,A}(t_i) = \sum_{j\in V_n} \xi_{j,n}$. We now turn V_n into the vertex set of a *dependency graph* (see Section 2.1 in [76] for the definition) by declaring that for $j, j' \in V_n$, $j \sim j'$ if and only if

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

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

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

Then we have

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

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

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

which itself yields an upper bound of the form

$$\lambda_{j,n} \leq n\|f\|_\infty m\Big(\mathrm{Tube}\Big(Q_{j,n}, Mr_n(t_m)\Big)\Big) =: c. \tag{3.29}$$

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

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

$$\mathbb{E}[|\xi_{j,n}|^\alpha] \leq \left(\sum_{i=1}^m |a_i|\right)^\alpha \mathbb{E}\left[\binom{Y_{j,n}}{k+1}^\alpha\right] \leq \left(\sum_{i=1}^m |a_i|\right)^\alpha \mathbb{E}\left[\binom{Y}{k+1}^\alpha\right] = C^* \tag{3.30}$$

Letting

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

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

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

where we have applied (3.28) for the second inequality.

Now we have by (3.30) that $\mathbb{E}\Big[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^p\Big] \leq C^*$ for $p = 3, 4$, so that

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

From the argument thus far we conclude that

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

which in turn implies

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

for all bounded sets A. The case when A is unbounded can be established by standard approximation arguments nearly identical to those in [52] and [76], so we omit the details and conclude that as $n \to \infty$

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

This is equivalent to

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

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

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

since $\Phi_{k,\mathbb{R}^d}^{(M)}(t_i,t_j) \to \Phi_{k,\mathbb{R}^d}(t_i,t_j)$ as $M \to \infty$. According to Theorem 3.2 in [10] it suffices to show that for every $t > 0$ and $\epsilon > 0$,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}\bigg(\Big|\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t) - \mathbb{E}[\beta_{k,n}(t) - \beta_{k,n}^{(M)}(t)]\Big| > \epsilon n^{1/2}\bigg) = 0. \tag{3.31}$$

By Chebyshev's inequality, the probability in (3.31) is bounded by

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

which itself converges to

$$\frac{1}{\epsilon^2}\sum_{i_1=M+1}^{\infty}\sum_{i_2=M+1}^{\infty}\sum_{j_1>0}\sum_{j_2>0} j_1 j_2\left(\frac{\eta_{k,\mathbb{R}^d}^{(i_1,j_1,j_2)}(t_1,t_2)\delta_{i_1,i_2}}{i_1!}+\frac{\nu_{k,\mathbb{R}^d}^{(i_1,i_2,j_1,j_2)}(t_1,t_2)}{i_1!i_2!}\right), \quad n\to\infty. \tag{3.32}$$

Since $\Phi_{k,\mathbb{R}^d}(t,t)$ is a finite constant, (3.32) goes to 0 as $M\to\infty$. $\square$

We now offer the proof of the CLT in the sparse regime, which is highly similar to the proof above, albeit with "de-Poissonization" at the end, for the binomial setup.

Proof of Theorem 3.4.1. We first establish the central limit theorem for $G_{k,n}(t)$ by proceeding in an almost identical fashion to Theorem 3.5.1. As in the case of the proofs of the limiting first moment and covariance of the Betti number process in Section 3.3, we give only a short argument here. We apply Theorem 2.2.3 once again. As in the Theorem 3.5.1, we require that the left-most point of each subset $\mathcal{Y}\subset\mathcal{P}_n$ to lie in an (open) bounded set $A\subset\mathbb{R}^d$, with $m(\partial A)=0$. Let $V_n, Q_{j,n}$ be defined as in the proof of Theorem 3.5.1 assume that $0<t_1<t_2<\cdots<t_m$. In this case however, we let V_n be the vertex set of a dependency graph by letting $j\sim j'$ if and only if

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

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

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

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

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

It is easy to demonstrate that the Poisson parameter of $Y_{j,n}$ is bounded by cns_n^d for some constant $c > 0$—see (3.29). Letting C^* be a general positive constant as in the proof of Theorem 3.5.1, we get that for $\alpha \in \mathbb{N}$,

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

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

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

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

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

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

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

Thus we have obtained

$$\rho_n^{-1/2}\Big(G_{k,n}(t_i) - \mathbb{E}\Big[G_{k,n}(t_i)\Big],\, i = 1, \dots, m\Big) \Rightarrow \mathcal{N}\Big(0, (\mu_{k,\mathbb{R}^d}(t_i, t_j))_{i,j=1}^m\Big). \tag{3.33}$$

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

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

Therefore (3.33) is equivalent to

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

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

$$\rho_n^{-1/2}\Big(\beta_{k,n}(t) - \mathbb{E}\Big[\beta_{k,n}(t)\Big]\Big) - \rho_n^{-1/2}\Big(G_{k,n}(t) - \mathbb{E}\Big[G_{k,n}(t)\Big]\Big) \xrightarrow{p} 0, \quad n\to\infty.$$

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

$$\mathbb{P}\Big(\rho_n^{-1/2}\Big|R_{k,n}(t) - \mathbb{E}[R_{k,n}(t)]\Big| > \epsilon\Big) \le \frac{1}{\epsilon^2\rho_n}\operatorname{Var}\Big(R_{k,n}(t)\Big) \to 0,$$

where the convergence is a direct consequence of $\rho_n^{-1}D_{2,n}\to 0$ and $\rho_n^{-1}A_{2,n}\to 0$, which were verified in the proof of Proposition 3.3.2.

De-Poissonization for Theorem 3.4.1. For $\mathcal{X}\in\mathcal{F}(\mathbb{R}^d)$ define

$$R_{k,n}^t(\mathcal{X}) := \sum_{i=k+3}^{\infty}\sum_{j>0} j \sum_{\mathcal{Y}\subset\mathcal{X}} g_{r_n(t)}^{(i,j)}(\mathcal{Y},\mathcal{X}).$$

We note that if

$$\frac{\operatorname{Var}(R_{k,n}^t(\mathcal{X}_n))}{n^{k+2}s_n^{d(k+1)}} \to 0, \quad n\to\infty, \tag{3.34}$$

then

$$\frac{R_{k,n}^t(\mathcal{X}_n) - \mathbb{E}[R_{k,n}^t(\mathcal{X}_n)]}{\sqrt{n^{k+2}s_n^{d(k+1)}}} \xrightarrow{P} 0,$$

and proving de-Poissonization for $\sum_{i=1}^m a_i G_{k,n}^{t_i}$ with $G_{k,n}^t(\mathcal{X}) = \sum_{\mathcal{Y}\subset\mathcal{X}} g_{r_n(t)}(\mathcal{Y},\mathcal{X})$, along with an application of Slutsky's theorem finishes the proof for the finite-dimensional CLT for $(\beta_{k,n}(t),\ t\ge 0)$ in the sparse regime, when $\Phi_n=\mathcal{X}_n$. However, the proof for de-Poissonization of $\sum_{i=1}^m a_i G_{k,n}^{t_i}$ is the same as that of Theorem 3.17 in [52] (with minor adjustments to accommodate Cramér-Wold device, such as using the subadditivity of the supremum). Thus, we will only prove (3.34). Note that we will liberally apply the upper bound $\binom{n}{i}\le n^i/i!$ in the sequel.

We can use the Efron-Stein inequality (Lemma 2.2.1) to see that

$$\frac{\mathrm{Var}(R^t_{k,n}(\mathcal{X}_n))}{n^{k+2}s_n^{d(k+1)}} \le \frac{2\mathbb{E}[(R^t_{k,n}(\mathcal{X}_n)) - R^t_{k,n}(\mathcal{X}_{n-1}))^2]}{(ns_n^d)^{k+1}}.$$

From the definition of $R^t_{k,n}$ we have

$$\begin{aligned} R^t_{k,n}(\mathcal{X}_n) - R^t_{k,n}(\mathcal{X}_{n-1}) = &\sum_{i=k+3}^{n} \sum_{j>0} \sum_{\mathcal{Y}\subset\mathcal{X}_{n-1}} j g^{(i,j)}_{r_n(t)}(\mathcal{Y}\cup\{X_n\},\mathcal{X}_n) \\ &+ \sum_{i=k+3}^{n-1} \sum_{j>0} \sum_{\mathcal{Y}\subset\mathcal{X}_{n-1}} j \Big[g^{(i,j)}_{r_n(t)}(\mathcal{Y},\mathcal{X}_n) - g^{(i,j)}_{r_n(t)}(\mathcal{Y},\mathcal{X}_{n-1})\Big]. \end{aligned}$$

We can simplify things further by noting

$$J_{i,r_n(t)}(\mathcal{Y},\mathcal{X}_n) - J_{i,r_n(t)}(\mathcal{Y},\mathcal{X}_{n-1}) = -J_{i,r_n(t)}(\mathcal{Y},\mathcal{X}_{n-1})1\Big\{X_n \in \mathcal{B}(\mathcal{Y}, r_n(t))\Big\},$$

where $\mathcal{B}(\mathcal{Y}, r_n(t))$ defined at (3.7), so that

$$\begin{aligned} R^t_{k,n}(\mathcal{X}_n) - R^t_{k,n}(\mathcal{X}_{n-1}) &= \sum_{i=k+3}^{n} \sum_{j>0} \sum_{\mathcal{Y}\subset\mathcal{X}_{n-1}} j g^{(i,j)}_{r_n(t)}(\mathcal{Y}\cup\{X_n\},\mathcal{X}_n) \\ &\quad - \sum_{i=k+3}^{n-1} \sum_{j>0} \sum_{\mathcal{Y}\subset\mathcal{X}_{n-1}} j g^{(i,j)}_{r_n(t)}(\mathcal{Y},\mathcal{X}_{n-1})1\Big\{X_n \in \mathcal{B}(\mathcal{Y}, r_n(t))\Big\} \\ &=: F_{1,n} - F_{2,n}. \end{aligned}$$

Clearly, we have

$$\Big(R^t_{k,n}(\mathcal{X}_n) - R^t_{k,n}(\mathcal{X}_{n-1})\Big)^2 = F^2_{1,n} - 2F_{1,n}F_{2,n} + F^2_{2,n} \le F^2_{1,n} + F^2_{2,n}.$$

When $\mathcal{Y}_1 \neq \mathcal{Y}_2$,

$$g^{i_1,j_1}_{r_n(t)}(\mathcal{Y}_1\cup\{X_n\},\mathcal{X}_n) g^{i_2,j_2}_{r_n(t)}(\mathcal{Y}_2\cup\{X_n\},\mathcal{X}_n) = 0,$$

for all $i_1, i_2 \ge k+3$ and $j_1, j_2 > 0$, because $\mathcal{Y}_1$ and $\mathcal{Y}_2$ would have to be connected via X_n. Therefore,

$$F^2_{1,n} = \sum_{i=k+3}^{n} \sum_{j>0} \sum_{\mathcal{Y}\subset\mathcal{X}_{n-1}} j^2 g^{(i,j)}_{r_n(t)}(\mathcal{Y}\cup\{X_n\},\mathcal{X}_n).$$

Similar to previous arguments, we see that

$$\begin{aligned}F_{2,n}^2 &= \sum_{i=k+3}^{n-1} \sum_{j>0} \sum_{\mathcal{Y}\subset\mathcal{X}_{n-1}} j^2 g_{r_n(t)}^{(i,j)}(\mathcal{Y},\mathcal{X}_{n-1}) 1\Big\{X_n \in \mathcal{B}(\mathcal{Y}, r_n(t))\Big\} \\ &\quad + \sum_{i_1=k+3}^{n-1} \sum_{i_2=k+3}^{n-1-i_1} \sum_{j_1>0} \sum_{j_2>0} \sum_{\mathcal{Y}_1\subset\mathcal{X}_{n-1}} \sum_{\mathcal{Y}_2\subset\mathcal{X}_{n-1}\setminus\mathcal{Y}_1} j_1 j_2 g_{r_n(t)}^{(i_1,j_1)}(\mathcal{Y}_1,\mathcal{X}_{n-1}) g_{r_n(t)}^{(i_2,j_2)}(\mathcal{Y}_2,\mathcal{X}_{n-1}) \\ &\quad \times 1\Big\{X_n \in \mathcal{B}(\mathcal{Y}_1, r_n(t))\Big\} 1\Big\{X_n \in \mathcal{B}(\mathcal{Y}_2, r_n(t))\Big\}.\end{aligned}$$

Now,

$$\begin{aligned}\frac{\mathbb{E}[F_{1,n}^2]}{(ns_n^d)^{k+1}} &\le \frac{1}{(ns_n^d)^{k+1}} \sum_{i=k+3}^{n} \sum_{j>0} j^2 \binom{n}{i-1} \mathbb{E}[1\Big\{\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}\Big\} b_{j,r_n(t)}(\mathcal{X}_i)] \\ &\le \frac{1}{(ns_n^d)^{k+1}} \sum_{i=k+3}^{\infty} \binom{i}{k+1}^2 \binom{n}{i-1} \mathbb{P}(\check{C}(\mathcal{X}_i, r_n(t)) \text{ is connected}) \\ &\le \frac{1}{[(k+1)!]^2} \sum_{i=k+3}^{\infty} \frac{i^{2k+i}(ns_n^d)^{i-k-2}}{(i-1)!} (t^d \|f\|_\infty \omega_d)^{i-1} \to 0, \quad n\to\infty,\end{aligned}$$

by the argument as at Proposition 6.2. One can establish the same for the expectation of the first term of $F_{2.n}^2$, by noting we choose i points and must have $\check{C}(\mathcal{Y}\cup\{X_n\}, r_n(t))$ connected, because of the term $1\Big\{X_n \in \mathcal{Y} + B(0, r_n(t))\Big\}$. For the second sum in $F_{2,n}^2$ we note that the summand is bounded by

$$\begin{aligned}&\binom{i_1}{k+1}\binom{i_2}{k+1} 1\Big\{\check{C}(\mathcal{Y}_1 \cup \{X_n\}, r_n(t)) \text{ is connected}\Big\} 1\Big\{\check{C}(\mathcal{Y}_2 \cup \{X_n\}, r_n(t)) \text{ is connected}\Big\} \\ &\qquad \times 1\Big\{|\mathcal{Y}_1| = i_1\Big\} 1\Big\{|\mathcal{Y}_2| = i_2\Big\} b_{j_1,r_n(t)}(\mathcal{Y}_1) b_{j_2,r_n(t)}(\mathcal{Y}_2),\end{aligned}$$

hence the whole sum is bounded above by

$$\begin{aligned}&\sum_{i_1=k+3}^{n-1} \sum_{i_2=k+3}^{n-1-i_1} \sum_{\mathcal{Y}_1\subset\mathcal{X}_{n-1}} \sum_{\mathcal{Y}_2\subset\mathcal{X}_{n-1}\setminus\mathcal{Y}_1} \binom{i_1}{k+1}\binom{i_2}{k+1} 1\Big\{\check{C}(\mathcal{Y}_1 \cup \{X_n\}, r_n(t)) \text{ is connected}\Big\} \\ &\qquad \times 1\Big\{\check{C}(\mathcal{Y}_2 \cup \{X_n\}, r_n(t)) \text{ is connected}\Big\} 1\Big\{|\mathcal{Y}_1| = i_1\Big\} 1\Big\{|\mathcal{Y}_2| = i_2\Big\}.\end{aligned}$$

Taking expectations gives an upper bound of

$$\sum_{i_1=k+3}^{n-1}\sum_{i_2=k+3}^{n-1-i_1}\binom{i_1}{k+1}\binom{i_2}{k+1}\binom{n}{i_1+i_2}\binom{i_1+i_2}{i_2}$$
$$\times\,\mathbb{P}\Big(\check{C}(\mathcal{X}_{i_1}\cup\{X_n\},r_n(t))\text{ is connected, }\check{C}(\mathcal{X}_{i_2}\cup\{X_n\},r_n(t))\text{ is connected}\Big),\tag{3.35}$$

where $\mathcal{X}_{i_1}$ and $\mathcal{X}_{i_2}$ are i.i.d sets of points with density f (and disjoint from each other and $\{X_n\}$) with cardinalities i_1 and i_2. We can bound the probability in (3.35), by recalling the bound in (3.22) and seeing that

$$\begin{aligned}
&\mathbb{P}\Big(\check{C}(\mathcal{X}_{i_1}\cup\{X_n\},r_n(t))\text{ is connected, }\check{C}(\mathcal{X}_{i_2}\cup\{X_n\},r_n(t))\text{ is connected}\Big)\\
&\quad=\int_{(\mathbb{R}^d)^{i_1+i_2+1}}1\Big\{\check{C}(\{x_0,x_1,\dots,x_{i_1}\},r_n(t))\text{ is connected}\Big\}\\
&\qquad\qquad\times 1\Big\{\check{C}(\{x_0,x_{i_1+1},\dots,x_{i_1+i_2}\},r_n(t))\text{ is connected}\Big\}\\
&\qquad\qquad\times f(x_0)f(x_1)\dots f(x_{i_1+i_2})\,\mathrm{d}\mathbf{x}\\
&\quad\le (r_n(t)^d\|f\|_\infty)^{i_1+i_2}\int_{(\mathbb{R}^d)^{i_1+i_2+1}}1\Big\{\check{C}(\{0,y_1,\dots,y_{i_1}\},r_n(t))\text{ is connected}\Big\}\\
&\qquad\qquad\times 1\Big\{\check{C}(\{0,y_{i_1+1},\dots,y_{i_1+i_2}\},r_n(t))\text{ is connected}\Big\}\,\mathrm{d}\mathbf{y},
\end{aligned}\tag{3.36}$$

by the standard change of variables. By the same reasoning as after (3.22), we have that (3.36) is bounded by

$$(r_n(t)^d\|f\|_\infty\,\omega_d)^{i_1+i_2}(i_1+1)^{i_1-1}(i_2+1)^{i_2-1}.$$

Therefore (3.35), divided by $(ns_n^d)^{k+1}$ is bounded by

$$\begin{aligned}
&\frac{1}{[(k+1)!]^2(ns_n^d)^{k+1}}\sum_{i_1=k+3}^{\infty}\sum_{i_2=k+3}^{\infty}\frac{(i_1+1)^{i_1+k}(i_2+1)^{i_2+k}}{i_1!i_2!}(nr_n(t)^d\|f\|_\infty\,\omega_d)^{i_1+i_2}\\
&\qquad\left(\frac{1}{(k+1)!}\sum_{i=k+3}^{\infty}\frac{(i+1)^{i+k}}{i!}(ns_n^d)^{i-(k+1)/2}(t^d\|f\|_\infty\,\omega_d)^i\right)^2\to 0,\quad n\to\infty,
\end{aligned}$$

as we can choose n large enough so the dominated convergence theorem applies.

In sum,

$$\frac{\mathrm{Var}(R^t_{k,n}(\mathcal{X}_n))}{n^{k+2}s_n^{d(k+1)}} \to 0, \quad n \to \infty.$$

As mentioned previously, we can de-Poissonize $\sum_{i=1}^m a_i G^{t_i}_{k,n}$ for any $m \in \mathbb{N}$, $a_i \in \mathbb{R}$, $t_i \geq 0$, $i = 1, \dots, m$, with minor adjustments to that of [52].

□

3.6 Poisson functional limit theorem in the sparse regime

Before concluding this paper we shall explore the random topology of Čech complexes when the complex is even more sparse than that in Section 3.4, so that k-cycles hardly ever occur. Then, the kth Betti number no longer follows a central limit theorem. Nevertheless, it does obey a Poisson limit theorem. In terms of the connectivity radii, we assume $\rho_n = n^{k+2}s_n^{d(k+1)} = 1$, equivalently, $s_n = n^{-(k+2)/d(k+1)}$, so that s_n converges to 0 more rapidly than in Section 3.4.

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

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

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

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

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

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

then $\mathcal{V}_k(t) = \mathcal{V}_k^+(t) - \mathcal{V}_k^-(t)$ is the difference of two dependent (time-changed) Poisson processes on $[0, \infty)$. Interestingly, this treatment is analogous to the statement of the Gaussian process limit in Section 3.4, and those wishing a deeper exploration of this in a similar setting should refer to [68]. What is precisely meant by this can be seen in the following proposition.

Proposition 3.6.1. *The process* $\mathcal{V}_k^{\pm}$ *can be expressed as*

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

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

Proof. As with Proposition 3.4.1, we prove only the result for $\mathcal{V}_k^+$, as the proof for $\mathcal{V}_k^-$ is the same. We can see that if $0 = t_0 < t_1 < \cdots < t_k < \infty$ and $\lambda_i > 0$, $i = 1, \ldots, k$, then by the nondecreasingness of h_t^+,

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

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

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

by a change of variable. Hence we have that

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

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

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

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

In what follows we assume $\rho_n = 1$, though we could easily modify this to suppose that $\rho_n \to 1$ as $n \to \infty$. For simplicity in our proofs we assert the former. The main techniques there are those in [26].

Theorem 3.6.1. *Let $\mathcal{K} = \check{C}$, take Φ_n to be either Poisson or binomial, and suppose $\rho_n \to 1$. Assume that f is an essentially bounded and continuous probability density. Then, we have the following weak convergence in Skorohod space, namely*

$$(\beta_{k,n}(t),\, t \geq 0) \Rightarrow (\mathcal{V}_k(t),\, t \geq 0), \quad \text{in } (D, J_1).$$

Proof. We begin by defining

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

and show that

$$(H_{k,n}(t),\, t \geq 0) \stackrel{fidi}{\Rightarrow} \Big(\mathcal{V}_k(t),\, t \geq 0\Big). \tag{3.37}$$

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

$$H_{k,n}(t) - G_{k,n}(t) \xrightarrow{p} 0, \tag{3.38}$$

$$\beta_{k,n}(t) - G_{k,n}(t) \xrightarrow{p} 0. \tag{3.39}$$

Noticing that $(\beta_{k,n}(t),\, t \in [0,T])$ and $(\mathcal{V}_k(t),\, t \in [0,T]) \in D_0$ for all $T > 0$, we see that the conditions of Theorem 2.3.3 hold due to Lemma 2.3.4 so we get the desired convergence in (D, J_1) seen in Theorem 3.6.1.

<u>*Part 1*</u>: For the proof of (3.37), it is sufficient to show that for any $a_1, a_2, \ldots, a_m > 0$, $m \geq 1$,

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

where we suppose that $0 < t_1 < \cdots < t_m < \infty$. We may use positive constants a_i, $i = 1, \ldots, m$ because of the fact that the Laplace transform characterizes a random vector with values in $[0, \infty)^m$. We proceed by using Theorem 3.1 from [26]. First let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a generic probability space on which all objects are defined. Recall that $\mathbf{N}[0, \infty)$ be the set of finite counting measures on $[0,\infty)$ and equip $\mathbf{N}[0,\infty)$ with the vague topology—see Section 2.2. Let us define a point process $\xi_n : \Omega \to \mathbf{N}[0,\infty)$ by

$$\xi_n(\cdot) := \sum_{\mathcal{Y} \subset \Phi_n} \mathbb{1}\Big\{ \sum_{i=1}^m a_i h_{r_n(t_i)}(\mathcal{Y}) > 0 \Big\} \, \delta_{\sum_{i=1}^m a_i h_{r_n(t_i)}(\mathcal{Y})}(\cdot).$$

Additionally let $\zeta : \Omega \to \mathbf{N}[0,\infty)$ denote a Poisson process with mean measure $C_{f,k}\tau_k$ where

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

for any Borel set A in $[0,\infty)$. The rest of Part 1 is devoted to showing that

$$\xi_n \Rightarrow \zeta \ \text{ in } \mathbf{N}[0,\infty). \tag{3.40}$$

According to Theorem 3.1 in [26], the following two conditions suffice for (3.40) when $\Phi_n = \mathcal{P}_n$. Let $\mathbf{L}_n(\cdot) := \mathbb{E}[\xi_n(\cdot)]$ and $\mathbf{M}(\cdot) := \mathbb{E}[\zeta(\cdot)] = C_{f,k}\tau_k(\cdot)$. The first requirement for (3.40) is the convergence in terms of the *total variation distance*:

$$d_{\mathrm{TV}}(\mathbf{L}_n, \mathbf{M}) := \sup_{A \in \mathcal{B}([0,\infty))} \Big| \mathbf{L}_n(A) - \mathbf{M}(A) \Big| \to 0, \quad n \to \infty, \tag{3.41}$$

where $\mathcal{B}([0,\infty))$ is the Borel σ-field over $[0,\infty)$. In addition, the second requirement for (3.40) is

$$\begin{aligned} v_n := \max_{1 \le \ell \le k+1} \int_{(\mathbb{R}^d)^\ell} \bigg(\int_{(\mathbb{R}^d)^{k+2-\ell}} \mathbb{1}\Big\{ \sum_{i=1}^m a_i h_{r_n(t_i)}(x_1, \dots, x_{k+2}) > 0 \Big\} \\ \lambda^{k+2-\ell}\Big(\mathrm{d}(x_{\ell+1}, \dots, x_{k+2})\Big) \bigg)^2 \lambda^\ell\Big(\mathrm{d}(x_1, \dots, x_\ell)\Big) \to 0 \end{aligned} \tag{3.42}$$

as $n \to \infty$, where $\lambda^m = \lambda \otimes \cdots \otimes \lambda$ is a product measure on $\mathbb{R}^m$ with $\lambda(\cdot) = n \int_\cdot f(z)\,\mathrm{d}z$. This condition does not differ between the Poisson and binomial cases. Finally, when $\Phi_n = \mathcal{X}_n$ if also

$$\mathbf{L}_n(\mathbb{R}^d)^2/n \to 0, \quad n \to \infty,$$

then (3.40) holds. As one can see below, $\mathbf{L}_n(\mathbb{R}^d) \to \mathbf{M}(\mathbb{R}^d) < \infty$, $n \to \infty$ for either the Poisson or binomial case, hence $\mathbf{L}_n(\mathbb{R}^d)^2/n \to 0$ for the binomial case in particular.

Let us now return to (3.41) and present its proof here. As usual, we have assumed $0 < t_1 < t_2 < \cdots < t_m$. Let us consider the Poisson case first. Then, for any $A \in \mathcal{B}([0,\infty))$ we have from Palm theory, the change of variables $x_1 = x$, $x_i = x + s_n y_{i-1}$ for $i = 2, \dots, k+2$, and $\rho_n = 1$ that

$$\begin{aligned}\mathbf{L}_n(A) &= \frac{n^{k+2}}{(k+2)!} \int_{(\mathbb{R}^d)^{k+2}} \mathbb{1}\Big\{ \sum_{i=1}^m a_i h_{r_n(t_i)}(\mathbf{x}) \in A \setminus \{0\} \Big\} \prod_{j=1}^{k+2} f(x_j)\, \mathrm{d}\mathbf{x} \qquad (3.43)\\ &= \frac{1}{(k+2)!} \int_{(\mathbb{R}^d)^{k+2}} \mathbb{1}\Big\{ \sum_{i=1}^m a_i h_{t_i}(0,\mathbf{y}) \in A \setminus \{0\} \Big\} f(x) \prod_{j=1}^{k+1} f(x + s_n y_j)\, \mathrm{d}x\, \mathrm{d}\mathbf{y}.\end{aligned}$$

Therefore,

$$\begin{aligned}\Big| \mathbf{L}_n(A) - \mathbf{M}(A) \Big| \le \frac{1}{(k+2)!} \int_{\mathbb{R}^{d(k+2)}} &\mathbb{1}\Big\{ \sum_{i=1}^m a_i h_{t_i}(0,\mathbf{y}) \in A \setminus \{0\} \Big\} \\ &\times f(x) \Big| \prod_{j=1}^{k+1} f(x + s_n y_j) - f(x)^{k+1} \Big| \,\mathrm{d}x\, \mathrm{d}\mathbf{y}. \qquad (3.44)\end{aligned}$$

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

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

We have by continuity of f that $\Big| \prod_{j=1}^{k+1} f(x + s_n y_j) - f(x)^{k+1} \Big|$ converges to 0 a.e. as $n \to \infty$ and is bounded by $2\|f\|_\infty^{k+1} < \infty$. So the dominated convergence theorem applies to get $\Big| \mathbf{L}_n(A) - \mathbf{M}(A) \Big| \to 0$ as $n \to \infty$. Since this convergence holds uniformly for all $A \in \mathcal{B}([0,\infty))$, we have now established (3.41) for the Poisson case.

We need only make a minor modification for the binomial setup. In this case, $\mathbf{L}_n(A)$ is the same as in (3.43), save for the difference in the constant outside the integral: $\binom{n}{k+2}$ versus $n^{k+2}/(k+2)!$. Let us denote the quantity at (3.43) by $\mathbf{L}'_n(A)$. Then

$$\Big|\mathbf{L}_n(A) - \mathbf{M}(A)\Big| \le \Big|\mathbf{L}_n(A) - \mathbf{L}'_n(A)\Big| + \Big|\mathbf{L}'_n(A) - \mathbf{M}(A)\Big|,$$

and the right term disappears by the argument after (3.44). The term $\Big|\mathbf{L}_n(A) - \mathbf{L}'_n(A)\Big|$ goes to zero as well, after noting the integral terms are the same (as well as bounded uniformly for all Borel sets A) all that remains to note is that

$$|n(n-1)\cdots(n-k-1) - n^{k+2}|s_n^{d(k+1)} \to 0, \quad n \to \infty,$$

which gives $\Big|\mathbf{L}_n(A) - \mathbf{L}'_n(A)\Big| \to 0$, $n \to \infty$ uniformly for all Borel $A \subset [0, \infty)$.

Next we turn to proving (3.42). First we can immediately see that

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

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

$$\begin{aligned} v_n \le \|f\|_\infty^{2k+3-\ell} \max_{1 \le \ell \le k+1} n^{2k+4-\ell} s_n^{d(2k+3-\ell)} \int_{\mathbb{R}^{d(2k+3-\ell)}} & 1\Big\{\sum_{i=1}^m a_i h_{t_i}(0, y_1, \dots, y_{k+1}) > 0\Big\} \\ & \times 1\Big\{\sum_{i=1}^m a_i h_{t_i}(0, y_1, \dots, y_{\ell-1}, y_{k+2}, \dots, y_{2k+3-\ell}) > 0\Big\}\,\mathrm{d}\mathbf{y}. \end{aligned}$$

Obviously the above integral is finite, and

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

by the assumption $\rho_n = 1$. So $v_n \to 0$ follows and (3.42) is obtained.

Part 2: Define the map $\widehat{T} : \mathbf{N}[0,\infty) \to [0,\infty)$ by $\widehat{T}(\sum_n \delta_{x_n}) = \sum_n x_n$. This map is continuous because it is defined on the space of *finite* counting measures. Applying the continuous mapping theorem to (3.40) gives $\widehat{T}(\xi_n) \Rightarrow \widehat{T}(\zeta)$. Equivalently, we have

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

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

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

where $Y_1, Y_2, \dots$ are i.i.d with common distribution $\tau_k(\cdot)/\tau_k([0,\infty))$ and M_n is Poisson distributed with parameter $C_{f,k}\tau_k([0,\infty))$—this is possible due to Proposition 3.8 in [58], for example. Further, $(Y_i)_{i\geq 1}$ and M_n are independent. On one hand, it follows from the Laplace functional of a Poisson process (see Theorem 5.1 in [82]) that for every $\lambda > 0$,

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

On the other hand it is straightforward to compute that

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

implying $\widehat{T}(\zeta) \stackrel{d}{=} \sum_{i=1}^m a_i \mathcal{V}_k(t_i)$—via the equality of the Laplace transforms—as required.

Part 3: It remains to show (3.38) and (3.39). As for (3.38), we know from (3.23) with $\rho_n = 1$, that

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

Since the exponential term in (3.24) converges to 1 without affecting the value of the limit, it must be that the $\mathbb{E}[H_{k,n}(t)]$ and $\mathbb{E}[G_{k,n}(t)]$ have the same limit. That is,

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

and thus, Markov's inequality gives (3.38).

Finally we turn our attention to (3.39). By Markov's inequality, it suffices to show that $\mathbb{E}[R_{k,n}(t)] \to 0$ as $n \to \infty$. Mimicking the derivation of (3.11) with $\rho_n = 1$, we get that

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

using $\binom{n}{i} \le n^i/i!$ for the binomial case. Recalling the bound in (3.12), we have

$$\mathbb{E}[R_{k,n}(t)] \le \frac{\Big(t^d \|f\|_\infty \omega_d\Big)^{k+1}}{(k+1)!} \sum_{i=k+3}^{\infty} \frac{i^{i-2}}{(i-k-1)!} \Big(n r_n(t)^d \|f\|_\infty \omega_d\Big)^{i-(k+2)} \to 0 \tag{3.45}$$

as $n \to \infty$. □

4. LIMIT THEORY FOR EULER CHARACTERISTIC PROCESSES

NB: A version of this chapter has been published in *Advances in Applied Probability*, Volume 53, Issue 1, March 2021, pp. 57–80. DOI: https://doi.org/10.1017/apr.2020.46. Reproduced with permission from the Applied Probability Trust.

Here we discuss a quantity that had received less of a treatment in the literature of random geometric complexes—the Euler characteristic process. The relevant stochastic process here is

$$(\chi_n(t), \geq 0) := \Big(\chi(\mathcal{K}(\Phi_n, r_n(t))),\, t \geq 0\Big), \tag{4.1}$$

and we let Φ_n be Poisson as in the previous chapter. For the functional strong law of large numbers, we let $\mathcal{K}$ be any right-continuous simplicial construction function, but take $\mathcal{K} = \mathcal{R}$ for the FCLT. The first section details the necessary background material and as in the previous chapter we proceed with a description of moment results. Section 4.3 handles the functional strong law of large numbers for (4.1) and Section 4.4 handles the functional central limit theorem. All the limiting results in this chapter take place in the critical regime, where $ns_n^d \to 1$. Recall that the behavior of the Euler characteristic is less interesting in the sparse regime because points, or 0-simplices, dominate all other features. Furthermore, crucial results have been established for the dense regime—see the subsection on the Euler characteristic in the Introduction.

Significant portions of the work in this chapter were reproduced from the article [92] by the author. In Remark 4.4.2 we discuss the extensions of the author's work in [92] that were made in [57] and the connections they have to our results. Unlike the previous section, the norm $\|\cdot\|$ may be taken to be arbitrary here, with the Čech complex coinciding with the Vietoris-Rips complex for the L^∞ norm[1].

---

[1] ↑A proof for this can be found at Lemma *VII* in [40].

4.1 Setup

At this point, most of the relevant notation has been introduced. However, there are few things that should be clarified for the results in this chapter. Recall the definition of the indicator function h_t^k seen at (2.4). This determines if a subset of $k+1$ points in $\mathbb{R}^d$ forms a k-simplex with respect to the simplicial complex $\mathcal{K}(\mathcal{X}, t)$, for some point cloud $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$ and simplicial construction function $\mathcal{K}$. With this in mind, we define $S_{k,n}(t) := S_k\Big(\mathcal{K}(\mathcal{P}_n, r_n(t))\Big)$, where S_k is defined in Section 2.1. Hence, we may define our Euler characteristic process as

$$\chi_n(t) := \sum_{k=0}^{\infty} (-1)^k S_{k,n}(t) = \sum_{k=0}^{\infty} (-1)^k \sum_{\mathcal{Y} \subset \mathcal{P}_n} h_{r_n(t)}^k, \quad t \geq 0. \tag{4.2}$$

Notice that (4.2) is almost surely a finite sum because the cardinality of $\mathcal{P}_n$, denoted as $|\mathcal{P}_n|$, is finite a.s. and $S_{k,n}(t) \equiv 0$ for all $k \geq |\mathcal{P}_n|$. Furthermore, for a Borel subset A of $\mathbb{R}^d$, define a restriction of the Euler characteristic to A by

$$\chi_{n,A}(t) := \sum_{k=0}^{\infty} (-1)^k \sum_{\mathcal{Y} \subset \mathcal{P}_n} h_{r_n(t)}^k(\mathcal{Y}) 1\Big\{\text{LMP}(\mathcal{Y}) \in A\Big\}, \tag{4.3}$$

as we did with the Betti number process at (3.3), where LMP again represents the least (leftmost) point in the lexicographic ordering on $\mathbb{R}^d$. Similarly, this will be used in applying Stein's method for normal approximation to demonstrate finite-dimensional convergence of the normalized version of (4.2). Clearly, $\chi_{n,\mathbb{R}^d}(t) = \chi_n(t)$.

4.2 Moment results for the Euler characteristic process

In order to obtain a clear picture of our limit theorems, it would be beneficial to start with some results on asymptotic moments of χ_n. Define for $k_1, k_2 \in \mathbb{N}_0$, $t, s \geq 0$, and a Borel subset A of $\mathbb{R}^d$,

$$\Psi_{k_1,k_2,A}(t,s) := \sum_{j=1}^{(k_1 \wedge k_2)+1} \psi_{j,k_1,k_2,A}(t,s),$$

and

$$\psi_{j,k_1,k_2,A}(t,s) := \frac{\int_A f(x)^{k_1+k_2+2-j}\,\mathrm{d}x}{j!(k_1+1-j)!(k_2+1-j)!}$$
$$\times \int_{(\mathbb{R}^d)^{k_1+k_2+1-j}} h_t^{k_1}(0,y_1,\dots,y_{k_1}) h_s^{k_2}(0,y_1,\dots,y_{j-1},y_{k_1+1},\dots,y_{k_1+k_2+1-j})\,\mathrm{d}\mathbf{y}.$$

Here we set $h_t^k(0,y_1,\dots,y_k)=1$ if $k=0$, so that $\Psi_{0,0,A}(t,s)=\psi_{1,0,0,A}(t,s)=\int_A f(x)dx$. In the sequel, we write $\Psi_{k_1,k_2}(t,s) := \Psi_{k_1,k_2,\mathbb{R}^d}(t,s)$ with $\psi_{j,k_1,k_2}(t,s) := \psi_{j,k_1,k_2,\mathbb{R}^d}(t,s)$. Finally, shorten $\psi_{k+1,k,k,A}(t,t)$ to $\psi_{k,A}(t)$ and write $\psi_k(t) := \psi_{k,\mathbb{R}^d}(t)$.

Proposition 4.2.1. *For $t,s\ge 0$, $A\subset\mathbb{R}^d$ open with $m(\partial A)=0$, we have*

$$n^{-1}\mathbb{E}[\chi_{n,A}(t)] \to \sum_{k=0}^{\infty}(-1)^k\psi_{k,A}(t), \quad n\to\infty, \tag{4.4}$$
$$n^{-1}\mathrm{Cov}\Big(\chi_{n,A}(t),\chi_{n,A}(s)\Big) \to \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}(-1)^{k_1+k_2}\Psi_{k_1,k_2,A}(t,s), \quad n\to\infty, \tag{4.5}$$

so that both of the right hand sides are convergent for every such $A\subset\mathbb{R}^d$.

Before detailing the proof of these results, we need a lemma. We assume that $ns_n^d=1$, though a straightforward argument would generalize it to case where $ns_n^d\to 1$.

Lemma 4.2.1. *For any right-continuous simplicial construction function $\mathcal{K}$ and h_t^k as defined at (2.4) we have,*
(i) For $t\ge 0$, that

$$\frac{n^k}{(k+1)!}\,\mathbb{E}\left[h^k_{r_n(t)}(\mathcal{X}_{k+1})\right] \to \psi_k(t), \quad n\to\infty.$$

(ii) For all $n\in\mathbb{N}$,

$$n^k\mathbb{E}\left[h^k_{r_n(t)}(\mathcal{X}_{k+1})\right] \le (a_t)^k,$$

where

$$a_t := (ct)^d\omega_d\|f\|_\infty. \tag{4.6}$$

In the following, let $\mathcal{X}_1, \mathcal{X}_2$ be i.i.d sets of k_1+1, k_2+1 points with density f, that potentially intersect. Then,

(iii) For $1 \le j \le (k_1 \wedge k_2)+1$, $k_1, k_2 \in \mathbb{N}_0$, and $t, s \ge 0$,

$$\frac{n^{k_1+k_2+1-j}}{j!(k_1+1-j)!(k_2+1-j)!}\, \mathbb{E}\Big[h^{k_1}_{r_n(t)}(\mathcal{X}_1)h^{k_2}_{r_n(s)}(\mathcal{X}_2)1\Big\{|\mathcal{X}_1 \cap \mathcal{X}_2| = j\Big\}\Big] \to \psi_{j,k_1,k_2}(t,s)$$

as $n \to \infty$.

(iv) For all $n \in \mathbb{N}$,

$$n^{k_1+k_2+1-j}\mathbb{E}\Big[h^{k_1}_{r_n(t)}(\mathcal{X}_1)h^{k_2}_{r_n(s)}(\mathcal{X}_2)1\Big\{|\mathcal{X}_1 \cap \mathcal{X}_2| = j\Big\}\Big] \le (a_{t\vee s})^{k_1+k_2+1-j}.$$

Proof. We shall prove (iii) and (iv) only, since (i) and (ii) can be established by a similar and simpler argument. Making change of variables $x_0 = x$ and $x_i = x + s_n y_i$, $i = 1, \dots, k_1 + k_2 + 1 - j$, the left hand side of (iii) equals

$$\begin{aligned}
&\frac{n^{k_1+k_2+1-j}}{j!(k_1+1-j)!(k_2+1-j)!} \int_{(\mathbb{R}^d)^{k_1+k_2+2-j}} h^{k_1}_{r_n(t)}(x_0, \dots, x_{k_1}) \\
&\qquad\qquad \times h^{k_2}_{r_n(s)}(x_0, \dots, x_{j-1}, x_{k_1+1}, \dots, x_{k_1+k_2+1-j}) \prod_{i=1}^{k_1+k_2+2-j} f(x_i)\, \mathrm{d}\mathbf{x} \\
&= \frac{(ns_n^d)^{k_1+k_2+1-j}}{j!(k_1+1-j)!(k_2+1-j)!} \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{k_1+k_2+1-j}} h^{k_1}_t(0, y_1 \dots, y_{k_1}) \qquad (4.7) \\
&\quad \times h^{k_2}_s(0, y_1, \dots, y_{j-1}, y_{k_1+1}, \dots, y_{k_1+k_2+1-j}) f(x) \prod_{i=1}^{k_1+k_2+1-j} f(x + s_n y_i)\, \mathrm{d}\mathbf{y}\, \mathrm{d}x.
\end{aligned}$$

Recall that $ns_n^d = 1$ and note that $\prod_{i=1}^{k_1+k_2+1-j} f(x + s_n y_i) \to f(x)^{k_1+k_2+1-j}$, $n \to \infty$, holds under the integral sign because of the Lebesgue differentiation theorem. Thus, (4.7) converges to $\psi_{j,k_1,k_2}(t,s)$ as $n \to \infty$.

Now let us turn to proving statement (iv). Without loss of generality, we may assume $s \le t$. Performing the same change of variables as in (iii), the left hand side of (iv) is bounded by

$$\Big(\|f\|_\infty\Big)^{k_1+k_2+1-j} \int_{(\mathbb{R}^d)^{k_1+k_2+1-j}} h^{k_1}_t(0, y_1 \dots, y_{k_1}) h^{k_2}_s(0, y_1, \dots, y_{j-1}, y_{k_1+1}, \dots, y_{k_1+k_2+1-j})\, \mathrm{d}\mathbf{y}. \tag{4.8}$$

By the definition of the indicators $h_t^{k_1}$, $h_s^{k_2}$, each of the y_i's in (4.8) must be distance at most t from the origin. Therefore, (4.8) can be bounded by

$$\Big(\|f\|_\infty\Big)^{k_1+k_2+1-j} m\Big(B(0,ct)\Big)^{k_1+k_2+1-j} = (a_t)^{k_1+k_2+1-j}.$$

$\square$

Proof of Proposition 4.2.1. We only prove (4.5) as the proof techniques for (4.4) are very similar to (4.5). Specifically, we shall make use of *(ii)*, *(iii)*, and *(iv)* of Lemma 4.2.1. We start by writing

$$n^{-1}\mathrm{Cov}\Big(\chi_n(t), \chi_n(s)\Big) = n^{-1}\mathbb{E}\Bigg[\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}(-1)^{k_1+k_2}S_{k_1,n}(t)S_{k_2,n}(s)\Bigg] - n^{-1}\mathbb{E}\Bigg[\sum_{k=0}^{\infty}(-1)^k S_{k,n}(t)\Bigg]\mathbb{E}\Bigg[\sum_{k=0}^{\infty}(-1)^k S_{k,n}(t)\Bigg]. \tag{4.9}$$

Next, Palm theory for Poisson processes, Lemma 2.2.4*(ii)*, along with the bounds given in Lemma 4.2.1*(ii)* and *(iv)*, yields that

$$\begin{aligned}
&\mathbb{E}\Big[S_{k_1,n}(t)S_{k_2,n}(s)\Big] \\
&= \sum_{j=0}^{(k_1\wedge k_2)+1}\mathbb{E}\Bigg[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n}\sum_{\mathcal{Y}_2\subset\mathcal{P}_n} h_{r_n(t)}^{k_1}(\mathcal{Y}_1)h_{r_n(s)}^{k_2}(\mathcal{Y}_2)\, 1\Big\{|\mathcal{Y}_1\cap\mathcal{Y}_2| = j\Big\}\Bigg] \\
&= \frac{n^{k_1+k_2+2}}{(k_1+1)!(k_2+1)!}\,\mathbb{E}\Big[h_{r_n(t)}^{k_1}(\mathcal{X}_1)\Big]\mathbb{E}\Big[h_{r_n(s)}^{k_2}(\mathcal{X}_2)\Big] \\
&\quad + \sum_{j=1}^{(k_1\wedge k_2)+1}\frac{n^{k_1+k_2+2-j}}{j!(k_1+1-j)!(k_2+1-j)!}\,\mathbb{E}\Big[h_{r_n(t)}^{k_1}(\mathcal{X}_1)h_{r_n(s)}^{k_2}(\mathcal{X}_2)1\Big\{|\mathcal{X}_1\cap\mathcal{X}_2| = j\Big\}\Big] \\
&\leq \frac{n^2(a_t)^{k_1}(a_s)^{k_2}}{(k_1+1)!(k_2+1)!} + \sum_{j=1}^{(k_1\wedge k_2)+1}\frac{n(a_{t\vee s})^{k_1+k_2+1-j}}{j!(k_1+1-j)!(k_2+1-j)!},
\end{aligned}$$

where $\mathcal{X}_1, \mathcal{X}_2$ defined as in Lemma 4.2.1. Here it is straightforward to see that

$$\sum_{k=0}^{\infty}\frac{(a_t)^k}{(k+1)!} < e^{a_t} < \infty, \quad \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{j=1}^{(k_1\wedge k_2)+1}\frac{(a_{t\vee s})^{k_1+k_2+1-j}}{j!(k_1+1-j)!(k_2+1-j)!} < 2e^{3a_{t\vee s}} < \infty.$$

So Fubini's theorem is applicable to the first term in (4.9). Repeating the same argument for the second term of (4.9), one can get

$$n^{-1}\mathrm{Cov}\Big(\chi_n(t), \chi_n(s)\Big) = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}(-1)^{k_1+k_2}\sum_{j=1}^{(k_1\wedge k_2)+1}\frac{n^{k_1+k_2+1-j}}{j!(k_1+1-j)!(k_2+1-j)!}$$
$$\times\, \mathbb{E}\Big[h_{r_n(t)}^{k_1}(\mathcal{X}_1)h_{r_n(s)}^{k_2}(\mathcal{X}_2)1\Big\{|\mathcal{X}_1\cap\mathcal{X}_2| = j\Big\}\Big].$$

By virtue of Lemma 4.2.1 (*iii*) and (*iv*), the dominated convergence theorem can conclude that the last expression converges to $\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}(-1)^{k_1+k_2}\Psi_{k_1,k_2}(t,s)$ as required. $\square$

4.3 Functional strong law of large numbers

The main result of this section is to establish (and prove) the FSLLN for the process $(\chi_n(t)/n,\, t \ge 0)$ in (D, U). We then give an analytic expression for the limiting function for a few special cases, and detail a connection to a phenomenon called homological percolation. We now introduce the FSLLN.

Theorem 4.3.1. *Let $\mathcal{K}$ be any right-continuous simplicial construction function, take Φ_n to be Poisson, and suppose $ns_n^d \to 1$. Assume that f is an essentially bounded probability density. Then, as $n \to \infty$,*

$$\left(\frac{\chi_n(t)}{n},\, t \ge 0\right) \to \left(\sum_{k=0}^{\infty}(-1)^k\psi_k(t),\, t\ge 0\right), \quad \text{a.s. in } (D, U). \tag{4.10}$$

Remark 4.3.2. See Examples 5.2.2 and 5.3.3 in Chapter 5 in for applications of the continuous mapping theorem to the convergence established in Theorem 4.3.1.

Consider $\mathcal{K} = \mathcal{R}$ equipped with the L^p norm, $1 \le p \le \infty$. That is $\sigma \in \mathcal{R}(\mathcal{X}, t)$ if and only if $\sigma \subset \mathcal{X}$ and $\max_{x,y\in\sigma}\|x-y\|_p \le t$, where

$$\|x\|_p := \left(\sum_{i=1}^{d} x_i^p\right)^{1/p},$$

for finite p and $\|x\|_\infty = \max_{1\le i\le d}|x_i|$ for $p=\infty$. When $p=\infty$ we have an exact analytic formula for the limiting function of $(\frac{\chi_n(t)}{n},\, t\ge 0)$ in Theorem 4.3.1.

Corollary 4.3.3. *Let X be a random variable with density f as in Theorem 4.3.1. If $p = \infty$ then for any $d \geq 1$ and $t \geq 0$, we have*

$$\sum_{k=0}^{\infty}(-1)^k \psi_k(t) = \sum_{k=0}^{\infty}(-t^d)^k \mathbb{E}[f(X)^k]\frac{(k+1)^d}{(k+1)!},$$

and T_p, $p \geq 0$ is the p^{th} Touchard "polynomial",

$$T_p(x) := e^{-x}\sum_{k=0}^{\infty}\frac{x^k k^p}{k!},$$

which is a polynomial if $p \in \mathbb{N}$. Moreover, if f is the density of a uniform distribution on Borel set $A \subset \mathbb{R}^d$ of finite, non-zero Lebesgue measure, then

$$\sum_{k=0}^{\infty}(-1)^k \psi_k(t) = -\frac{m(A)e^{-t^d/m(A)}}{t^d}T_d\Big(-t^d/m(A)\Big), \tag{4.11}$$

or if f is the density of any nondegenerate multivariate Gaussian distribution $N(\mu, \Sigma)$ with normalization constant $C = (2\pi)^{d/2}\sqrt{\det(\Sigma)}$, then

$$\sum_{k=0}^{\infty}(-1)^k \psi_k(t) = -\frac{Ce^{-t^d/C}}{t^d}T_{d/2}(-t^d/C). \tag{4.12}$$

Proof. It suffices to find what

$$\int_{(\mathbb{R}^d)^k} h_1^k(0, y_1, \ldots, y_k)\,\mathrm{d}y_1 \ldots \mathrm{d}y_k,$$

evaluates to in the case of $p = \infty$, as

$$\begin{aligned}\psi_k(t) &:= \frac{\int_{\mathbb{R}^d} f(x)^{k+1}\,\mathrm{d}x}{(k+1)!}\int_{(\mathbb{R}^d)^k} h_t^k(0, y_1, \ldots, y_k)\,\mathrm{d}y_1 \ldots \mathrm{d}y_k \\ &= \frac{t^{dk}}{(k+1)!}\int_{(\mathbb{R}^d)^k} h_1^k(0, y_1, \ldots, y_k)\,\mathrm{d}y_1 \ldots \mathrm{d}y_k,\end{aligned}$$

after the change of variable $y_i \mapsto t y_i$. As we can represent

$$h_1^k(0, y_1, \ldots, y_k) = \prod_{1 \le i < j \le k} 1\Big\{\big\|y_i - y_j\big\| \le 1\Big\},$$

when $y_i \in [-1,1]^d$ for each $i = 1, \ldots, k$, then we have

$$\begin{aligned}
\int_{(\mathbb{R}^d)^k} h_1^k(0, y_1, \ldots, y_k)\, \mathrm{d}y_1 \ldots \mathrm{d}y_k &= \int_{[-1,1]^{dk}} \prod_{1 \le i < j \le k} 1\Big\{\big\|y_i - y_j\big\| \le 1\Big\}\, \mathrm{d}y_1 \ldots \mathrm{d}y_k \\
&= \int_{[-1,1]^{dk}} \prod_{\ell=1}^{d} \left[\prod_{1 \le i < j \le k} 1\Big\{\big|y_i^\ell - y_j^\ell\big| \le 1\Big\} \right] \mathrm{d}y_1 \ldots \mathrm{d}y_k \\
&= \left(\int_{[-1,1]^k} \prod_{1 \le i < j \le k} 1\Big\{\big|x_i - x_j\big| \le 1\Big\} \Bigg]\, \mathrm{d}x_1 \ldots \mathrm{d}x_k \right)^d,
\end{aligned}$$

where y_i^ℓ, $\ell = 1, \ldots, d$ is the ℓth coordinate of $y_i \in \mathbb{R}^d$. Thus it remains to find the value of the integral of $h_1^k(0, y_1, \ldots, y_k)$ for elements of $\mathbb{R}$. Continuing onward, let us choose x_1 and x_k to be the maximum and minimum, respectively. Then each of the other $k-2$ points must be within $[x_1, x_k]$

$$\begin{aligned}
&\int_{[-1,1]^k} \prod_{1 \le i < j \le k} 1\Big\{\big|x_i - x_j\big| \le 1\Big\}\, \mathrm{d}x_1 \ldots \mathrm{d}x_k \\
&\quad = k(k-1) \left(\int_0^1 \int_{x_1}^1 (x_k - x_1)^{k-2}\, \mathrm{d}x_k\, \mathrm{d}x_1 + \int_{-1}^0 \int_{x_1}^{x_1+1} (x_k - x_1)^{k-2}\, \mathrm{d}x_k\, \mathrm{d}x_1 \right) \\
&\quad = k + 1,
\end{aligned}$$

as there are $k(k-1)$ possible permutations of $x_{(1)}$ and $x_{(k)}$.

For $\mathbb{E}[f(X)^k]$, let us assume that $f(x) = g(x)/C$, where $C = \int_{\mathbb{R}^d} g(x)\, \mathrm{d}x$ and $g(x)$ only contains the x terms (no constants). Then

$$\mathbb{E}[f(X)^k] = C^{-(k+1)} \int_{\mathbb{R}^d} g(x)^{k+1}\, \mathrm{d}x,$$

and the expression in each case follows from standard properties of the uniform and multivariate normal distributions. $\square$

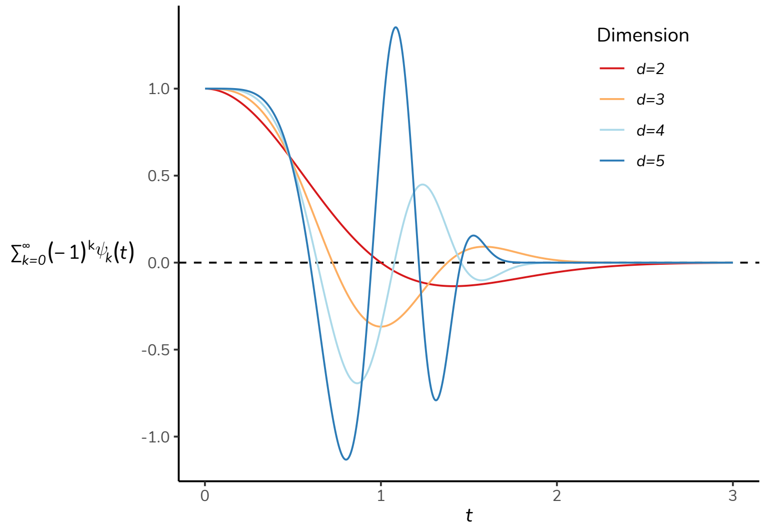


Figure 4.1. Limiting plots of (4.11) for $A = [0, 1]^d$, $d = 2, 3, 4, 5$.

As we discussed in the Introduction, there are preliminary results which connected the mean Euler characteristic process to the appearance of "significant" homology of a topological space M [18]. This was "inspired" by the work [63] connecting traditional percolation thresholds (for bond, site, and continuum percolation for example) and the Euler characteristic. Let us consider some submanifold $M \subset \mathbb{R}^d$ and $\mathcal{X}_n$ a random finite point cloud of size n, sampled from M. Consider $U(\mathcal{X}_n, r_n(t))$, where $ns_n^d = 1$, subject to conditions that ensure $U(\mathcal{X}_n, r_n(t)) \subset M$ for each $n \in \mathbb{N}$, $t \geq 0$. Then, the idea of homological percolation is to discover properties of the ξ_k such that with asymptotic probability 1 for $t > \xi_k$ there is some $\gamma \in H_k(M)$ that also appears in $H_k\big(U(\mathcal{X}_n, r_n(t))\big)$ and for $t < \xi_k$ this occurs with asymptotic probability 0. We might take M to be the d-dimensional torus $\mathbb{T}^d$ equal to $[0, 1]^d$ with periodic boundary conditions. The paper [19] establishes foundational results about this phenomenon. In [18] it was conjectured that the kth zero ζ_k, $k = 1, \ldots, d-1$ of the expected Euler characteristic process $(\mathbb{E}[\chi(U(\mathcal{X}_n, r_n(t)))], t \geq 0)$ is often close to ξ_k, such that $\zeta_k - \xi_k$ tends to some small finite constant as $n \to \infty$. Here, the support of our density f is contractible and has no homology to speak of, so percolation could be interpreted in the sense of a certain homology "dominating" a lower-dimensional one. In [18], this was defined as $\iota_k = \inf\{t \geq 0 : \beta_{k-1}(U(\mathcal{X}_n, r_n(t))) = \beta_k(U(\mathcal{X}_n, r_n(t))\}$. Thus Theorem 4.3.1 has implications for homological percolation given that our submanifold M is locally d-dimensional and the behavior in the critical regime is dominated by local features. Theorem 4.3.1 also supports the conclusion that all we need to assess the evolution of homology is the evolution of purely local features, i.e. the simplices. Understanding the Euler characteristic process, especially in the functional setting, can be used to demonstrate the unimodality of the limiting Betti number process in the critical regime (as conjectured in [49]), if we have some idea of the values of ι_k.

(a) (b)

Figure 4.2. (a) is a scatterplot of $n = 1000$ i.i.d. Gaussian random vectors in $\mathbb{R}^3$ with mean $\mu = (6, 0.8, 2.4)$ and variances $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\sigma_3^2 = 3$ and correlations $\rho_{12} = 0.7$, $\rho_{13} = 0.1$ and $\rho_{23} = -0.3$. In (b) the Euler characteristic process for this point cloud can be seen with $ns_n^3 = 1$ and (4.12) overlaid in orange.

To prove the functional strong law of large numbers, we use Proposition 2.3.1 to extend a pointwise strong law to a functional one. The approach taken here gives an improvement from the viewpoint of assumptions on the density f. Unlike the existing results such as [100], we do not require f to have compact support. The proof technique below takes a similar approach to that of [42].

Proof of Theorem 4.3.1. Since (4.2) is almost surely represented as a sum of finitely many terms, it can be split into two parts,

$$\chi_n(t) = \sum_{k=0}^{\infty} S_{2k,n}(t) - \sum_{k=0}^{\infty} S_{2k+1,n}(t) =: \chi_n^{(0)}(t) - \chi_n^{(1)}(t) \quad \text{a.s.}$$

Denoting by $K(t)$ the limit of (4.4) with $A = \mathbb{R}^d$, we decompose it in a way similar to the above,

$$K(t) = \sum_{k=0}^{\infty} \psi_{2k}(t) - \sum_{k=0}^{\infty} \psi_{2k+1}(t) =: K^{(0)}(t) - K^{(1)}(t).$$

Our final goal is to prove that for every $0 < T < \infty$,

$$\sup_{0 \le t \le T} \left| \frac{\chi_n(t)}{n} - K(t) \right| \to 0, \quad n \to \infty, \quad \text{a.s.},$$

which is clearly implied by

$$\sup_{0 \le t \le T} \left| \frac{\chi_n^{(i)}(t)}{n} - K^{(i)}(t) \right| \to 0, \quad n \to \infty, \quad \text{a.s.}$$

for each $i = 0, 1$. As $\chi_n^{(i)}(t)/n$ and $K^{(i)}(t)$ satisfy the conditions of Proposition 2.3.1, it suffices to show that

$$\frac{\chi_n^{(i)}(t)}{n} \to K^{(i)}(t), \quad n \to \infty, \quad \text{a.s.},$$

for every $t \ge 0$. We will only prove the case $i = 0$, and henceforth omit the superscript i from $\chi_n^{(i)}(t)$ and $K^{(i)}(t)$. It then suffices to show that

$$n^{-1} |\chi_n(t) - \mathbb{E}[\chi_n(t)]| \to 0, \quad n \to \infty, \quad \text{a.s.}, \tag{4.13}$$

and

$$\left| n^{-1} \mathbb{E}[\chi_n(t)] - K(t) \right| \to 0, \quad n \to \infty. \tag{4.14}$$

First we will deal with (4.14). It follows from the customary change of variables as in the proof of Lemma 4.2.1, that

$$\begin{aligned} &\left| n^{-1} \mathbb{E}[\chi_n(t)] - K(t) \right| \\ &\quad = \left| \sum_{k=1}^{\infty} \frac{1}{(2k+1)!} \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{2k}} h_t^{2k}(0, y_1, \dots, y_{2k}) \right. \\ &\qquad\qquad \left. \times f(x) \Big(\prod_{i=1}^{2k} f(x + s_n y_i) - f(x)^{2k} \Big) \, \mathrm{d}\mathbf{y} \, \mathrm{d}x \right| \end{aligned}$$

$$\leq \sum_{k=1}^{\infty} \frac{1}{(2k+1)!} \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{2k}} h_t^{2k}(0, y_1, \dots, y_{2k}) f(x) \Big| \prod_{i=1}^{2k} f(x + s_n y_i) - f(x)^{2k} \Big| \,\mathrm{d}\mathbf{y}\,\mathrm{d}x.$$

Similarly to the proof of Lemma 4.2.1(ii) or (iv), one can show that the last term above is bounded by $2\sum_{k=1}^{\infty}(a_t)^{2k}/(2k+1)! < \infty$, where a_t is defined at (4.6). Thus, the dominated convergence theorem concludes (4.14).

Now, let us turn our attention to (4.13). From the Borel-Cantelli lemma it suffices to show that, for every $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\big| \chi_n(t) - \mathbb{E}[\chi_n(t)] \big| > \epsilon n \Big) < \infty.$$

By Markov's inequality, the left hand side above is bounded by

$$\frac{1}{\epsilon^4} \sum_{n=1}^{\infty} \frac{1}{n^4} \mathbb{E}\Big[\Big(\chi_n(t) - \mathbb{E}[\chi_n(t)] \Big)^4 \Big].$$

Since $\sum_n n^{-2} < \infty$, we only need to show that

$$\limsup_{n\to\infty} \frac{1}{n^2} \mathbb{E}\Big[\Big(\chi_n(t) - \mathbb{E}[\chi_n(t)] \Big)^4 \Big] < \infty. \tag{4.15}$$

Applying Fubini's theorem as in the proof of Proposition 4.2.1, along with Hölder's inequality, we get that

$$\begin{aligned}
&\frac{1}{n^2} \mathbb{E}\Big[\Big(\chi_n(t) - \mathbb{E}[\chi_n(t)] \Big)^4 \Big] \\
&\qquad = \frac{1}{n^2} \sum_{(k_1,\dots,k_4)\in\mathbb{N}^4} \mathbb{E}\Big[\prod_{i=1}^{4} \Big(S_{2k_i,n}(t) - E\big[S_{2k_i,n}(t) \big] \Big) \Big] \\
&\qquad \leq \Bigg[\sum_{k=1}^{\infty} \Big\{ \frac{1}{n^2} \mathbb{E}\Big[\Big(S_{2k,n}(t) - \mathbb{E}\big[S_{2k,n}(t) \big] \Big)^4 \Big] \Big\}^{1/4} \Bigg]^4.
\end{aligned}$$

Now, (4.15) can be obtained if we show that

$$\sum_{k=1}^{\infty} \Big\{ \limsup_{n\to\infty} \frac{1}{n^2} \mathbb{E}\Big[\Big(S_{2k,n}(t) - \mathbb{E}\big[S_{2k,n}(t) \big] \Big)^4 \Big] \Big\}^{1/4} < \infty. \tag{4.16}$$

From this point on, let us introduce a shorthand notation, $S_{2k} := S_{2k,n}(t)$. In order to find an appropriate upper bound for (4.16), by the binomial expansion we write

$$\mathbb{E}\Big[\Big(S_{2k} - \mathbb{E}[S_{2k}]\Big)^4\Big] = \sum_{\ell=0}^{4} \binom{4}{\ell} (-1)^\ell \mathbb{E}[S_{2k}^\ell]\Big(\mathbb{E}[S_{2k}]\Big)^{4-\ell}. \tag{4.17}$$

For every $\ell \in \{0, \dots, 4\}$, one can denote $\mathbb{E}[S_{2k}^\ell]\Big(\mathbb{E}[S_{2k}]\Big)^{4-\ell}$ as

$$\mathbb{E}\Bigg[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n^{(1)}} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n^{(2)}} \sum_{\mathcal{Y}_3 \subset \mathcal{P}_n^{(3)}} \sum_{\mathcal{Y}_4 \subset \mathcal{P}_n^{(4)}} \prod_{i=1}^{4} h_{r_n(t)}^{2k}(\mathcal{Y}_i)\Bigg], \tag{4.18}$$

where for every $i, j \in \{1, \dots, 4\}$, we have either $\mathcal{P}_n^{(i)} = \mathcal{P}_n^{(j)}$ or $\mathcal{P}_n^{(i)}$ is an independent copy of $\mathcal{P}_n^{(j)}$. If $|\mathcal{Y}_1 \cup \cdots \cup \mathcal{Y}_4| = 8k + 4$, i.e., $\mathcal{Y}_1, \dots, \mathcal{Y}_4$ do not have any common elements, Palm theory (Lemma 2.2.4) shows that (4.18) is equal to $\Big(\mathbb{E}[S_{2k}]\Big)^4$, which grows at the rate of $O(n^4)$ (see Lemma 4.2.1(i)). In this case, the total contribution to (4.17) disappears, because

$$\sum_{\ell=0}^{4} \binom{4}{\ell} (-1)^\ell \Big(\mathbb{E}[S_{2k}]\Big)^4 = 0.$$

Suppose next that $|\mathcal{Y}_1 \cup \cdots \cup \mathcal{Y}_4| = 8k + 3$, that is, there is exactly one common element between $\mathcal{Y}_i$ and $\mathcal{Y}_j$ for some $i \neq j$ with no other overlappings. Then (4.18) is equal to

$$\mathbb{E}\Bigg[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} h_{r_n(t)}^{2k}(\mathcal{Y}_1) h_{r_n(t)}^{2k}(\mathcal{Y}_2)\, \mathbb{1}\{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = 1\}\Bigg] \Big(\mathbb{E}[S_{2k}]\Big)^2.$$

Although the growth rate of the above term is $O(n^3)$ (see Lemma 4.2.1(i) and (iii)), an overall contribution to (4.17) is again canceled. This is because

$$\begin{aligned}
&\left\{\binom{4}{2}(-1)^2 + \binom{4}{3}(-1)^3\binom{3}{2} + \binom{4}{4}(-1)^4\binom{4}{2}\right\} \\
&\quad \times \mathbb{E}\Bigg[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} h_{r_n(t)}^{2k}(\mathcal{Y}_1) h_{r_n(t)}^{2k}(\mathcal{Y}_2)\, \mathbb{1}\{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = 1\}\Bigg] \Big(\mathbb{E}[S_{2k}]\Big)^2 = 0.
\end{aligned}$$

By the above discussion, we only need to consider the case where there are at least two common elements within $\mathcal{Y}_1, \dots, \mathcal{Y}_4$. Among many such cases, let us deal with a specific term,

$$n^{-2}\mathbb{E}\Big[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n}\sum_{\mathcal{Y}_2\subset\mathcal{P}_n}\sum_{\mathcal{Y}_3\subset\mathcal{P}_n}\sum_{\mathcal{Y}_4\subset\mathcal{P}_n}\prod_{i=1}^{4} h_{r_n(t)}^{2k}(\mathcal{Y}_i) \tag{4.19}$$
$$\times\, 1\Big\{|\mathcal{Y}_1\cap\mathcal{Y}_2| = \ell_1,\ |\mathcal{Y}_3\cap\mathcal{Y}_4| = \ell_2,\ \Big|(\mathcal{Y}_1\cup\mathcal{Y}_2)\cap(\mathcal{Y}_3\cup\mathcal{Y}_4)\Big| = 0\Big\}\Big],$$

where $\ell_1, \ell_2 \in \{1, \dots, 2k+1\}$. Palm theory allows us to write (4.19) as

$$\prod_{i=1}^{2} \frac{n^{4k+1-\ell_i}}{\ell_i!\big((2k+1-\ell_i)!\big)^2}\, \mathbb{E}\Big[h_{r_n(t)}^{2k}(\mathcal{X}_1)h_{r_n(t)}^{2k}(\mathcal{X}_2)\, 1\{|\mathcal{X}_1\cap\mathcal{X}_2| = \ell_i\}\Big]. \tag{4.20}$$

By Lemma 4.2.1(iv) and $\ell!(2k+1-\ell)! \geq k!$ for any $\ell \in \{1, \dots, 2k+1\}$, one can bound (4.20) by

$$\prod_{i=1}^{2} \frac{(a_t)^{4k+1-\ell_i}}{\ell_i!\big((2k+1-\ell_i)!\big)^2} \leq \frac{(a_t)^{8k+2-\ell_1-\ell_2}}{k!}.$$

Now, the ratio test shows that

$$\sum_{k=1}^{\infty} \left\{ \frac{(a_t)^{8k+2-\ell_1-\ell_2}}{k!} \right\}^{1/4} < \infty$$

as desired. Notice that all the cases except (4.19) can be handled in a very similar way, and so, (4.16) follows. $\square$

4.4 Functional central limit theorem

The main result of this section is the functional central limit theorem for $(\bar{\chi}_n(t),\, t \geq 0)$ in (D, J_1), where

$$\bar{\chi}_n(t) := n^{-1/2}\Big(\chi_n(t) - \mathbb{E}[\chi_n(t)]\Big), \quad t \geq 0.$$

Before stating the FCLT for this section let us define the limiting process for $(\bar{\chi}_n(t),\, t \geq 0)$. First define $(\mathcal{L}_k,\ k \in \mathbb{N}_0)$ as a family of zero-mean Gaussian processes on a generic probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with intra-process covariance

$$\mathbb{E}[\mathcal{L}_k(t)\mathcal{L}_k(s)] = \Psi_{k,k}(t,s), \tag{4.21}$$

and inter-process convariance

$$\mathbb{E}[\mathcal{L}_{k_1}(t)\mathcal{L}_{k_2}(s)] = \Psi_{k_1,k_2}(t,s), \tag{4.22}$$

for all $k, k_1, k_2 \in \mathbb{N}_0$ with $k_1 \neq k_2$ and $t, s \geq 0$. In the proof of Proposition 4.2.1, the functions $\Psi_{k_1,k_2}(t,s)$ naturally appear in the covariance calculation of χ_n, which in turn implies that the covariance functions in (4.21) and (4.22) are well-defined. With these notations in mind, we now define the limiting Gaussian process for $\bar{\chi}_n$ as

$$\mathcal{L}(t) := \sum_{k=0}^{\infty} (-1)^k \mathcal{L}_k(t), \quad t \geq 0, \tag{4.23}$$

such that

$$\mathbb{E}[\mathcal{L}(t)\mathcal{L}(s)] = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} (-1)^{k_1+k_2} \Psi_{k_1,k_2}(t,s), \quad t, s \geq 0. \tag{4.24}$$

Once again, Proposition 4.2.1 implies that the right hand side of (4.24) can define the covariance functions of a limiting Gaussian process, since it is obtained as a (scaled) limit of the covariance functions of χ_n. In particular, since (4.24) is convergent, for every $t \geq 0$, $(\mathcal{L}(t),\, t \geq 0)$ is definable in the $L^2(\Omega)$-sense. Note that the Euler characteristic in (4.2) and the process (4.23) exhibit similar structure, in the sense that $S_{k,n}(t)$ in (4.2) and $\mathcal{L}_k(t)$ both correspond to the spatial distribution of k-simplices.

Theorem 4.4.1. *Let $\mathcal{K} = \mathcal{R}$, take Φ_n to be Poisson, and suppose $ns_n^d \to 1$. Assume that f is an essentially bounded probability density. Then, as $n \to \infty$,*

$$(\bar{\chi}_n(t),\, t \geq 0) \Rightarrow (\mathcal{L}(t),\, t \geq 0), \quad \text{in } (D, J_1), \tag{4.25}$$

Furthermore, for every $0 < T < \infty$*, we have that* $\Big(\mathcal{L}(t),\, 0 \le t \le T\Big)$ *has a continuous version with Hölder continuous sample paths of any exponent* $\gamma \in [0, 1/2)$*.*

Remark 4.4.2. This result was expanded upon in [57] to include $\mathcal{K} = \check{C}$ and Φ_n to be binomial as well. The proof of the functional central limit theorem for $(\bar{\chi}_n(t),\, t \ge 0)$ in [57] relies on a slightly different method than the one seen below. The authors used approximation arguments from [9], where one needs to only demonstrate that (2.10) holds for $t_1 \le s \le t_2$ elements of a well-behaved finite set $\Gamma_n \subset [0, T]$, which allows them to use a version of Lemma 4.4.3(i) for Čech complexes to prove (2.10) for general Φ_n, in conjunction with a martingale difference argument. The finite-dimensional convergence for the binomial process case was established via de-Poissonization—see Theorem 2.12 in [76]. All of the results in [57] hold for a density with bounded support. It is highly likely the results hold for the Čech and binomial cases for unbounded support as well.

The proof of Theorem 4.4.1 is broken into three parts. The first part, seen directly below, is devoted to the proof of finite-dimensional weak convergence of $(\bar{\chi}_n(t),\, t \ge 0)$, and the second part to tightness. The third and final part concerns the Hölder continuity of the limiting Gaussian process.

Proof of finite-dimensional convergence in Theorem 4.4.1. Recall (4.3) and define $\bar{\chi}_{n,A}(t)$ analogously to $\bar{\chi}_n(t)$ by mean-centering and scaling by $n^{-1/2}$. We first consider the case where A is an open and bounded subset of $\mathbb{R}^d$ with $m(\partial A) = 0$.

From the viewpoint of the Cramér-Wold device, one needs to establish weak convergence of $\sum_{i=1}^m a_i \bar{\chi}_n(t_i)$ for every $0 < t_1 < \cdots < t_m$, $m \in \mathbb{N}$, and $a_i \in \mathbb{R}$, $i = 1, \dots, m$. Our proof exploits Stein's normal approximation method in Theorem 2.2.3 and as in the proofs of Theorems 3.4.1 and 3.5.1 in the previous chapter. Let $(Q_{j,n},\, j \ge 1)$ be an enumeration of disjoint subsets of $\mathbb{R}^d$ congruent to $(0, r_n(t_m)/2]^d$, such that $\mathbb{R}^d = \bigcup_{j=1}^\infty Q_{j,n}$. Let us define $H_n := \{j \in \mathbb{N} : Q_{j,n} \cap A \neq \emptyset\}$ and

$$\xi_{j,n} := \sum_{k=0}^{\infty} (-1)^k \sum_{\mathcal{Y} \subset \mathcal{P}_n} \sum_{i=1}^{m} a_i h^k_{r_n(t_i)}(\mathcal{Y}) 1\Big\{\mathrm{LMP}(\mathcal{Y}) \in A \cap Q_{j,n}\Big\},$$

and also,

$$\bar{\xi}_{j,n} := \frac{\xi_{j,n} - \mathbb{E}[\xi_{j,n}]}{\sqrt{\mathrm{Var}\Big(\sum_{i=1}^{m} a_i \chi_{n,A}(t_i)\Big)}}.$$

Then, we have $\sum_{i=1}^{m} a_i \chi_{n,A}(t_i) = \sum_{j \in H_n} \xi_{j,n}$.

Now, let us define H_n to be the vertex set of a dependency graph (see discussion proceeding Theorem 2.2.3) for the random variables $(\bar{\xi}_{j,n},\ j \in H_n)$ by setting $j \sim j'$ if and only if the condition

$$\inf\Big\{\|x-y\| : x \in Q_{j,n},\ y \in Q_{j',n}\Big\} \le 2r_n(t_m),$$

is satisfied. This is because $\xi_{j,n}$ and $\xi_{j',n}$ become independent whenever $j \sim j'$ fails to hold. Now we must ensure that the other conditions to use Stein's method are satisfied with respect to the dependency graph $(H_n, \sim)$. First, $\sum_{j \in H_n} \bar{\xi}_{j,n}$ is a zero-mean random variable with unit variance. We know that $|H_n| = O(s_n^{-d})$ as A is bounded. Furthermore, the maximum degree of any vertex of H_n is uniformly bounded by a positive and finite constant. Let Z denote a standard normal random variable. Then Theorem 2.2.3 implies that

$$\begin{aligned}
\Big|\mathbb{P}\Big(\sum_{j \in H_n} \bar{\xi}_{j,n} \le x\Big) - \mathbb{P}(Z \le x)\Big| &\le C^* \left(\sqrt{s_n^{-d} \max_j \mathbb{E}\Big[|\bar{\xi}_{j,n}|^3\Big]} + \sqrt{s_n^{-d} \max_j \mathbb{E}\Big[|\bar{\xi}_{j,n}|^4\Big]}\right) \\
&\le C^* \left(\sqrt{s_n^{-d} n^{-3/2} \max_j \mathbb{E}\Big[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^3\Big]} + \sqrt{s_n^{-d} n^{-2} \max_j \mathbb{E}\Big[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^4\Big]}\right),
\end{aligned} \tag{4.26}$$

where the second inequality follows from Proposition 4.2.1 that claims that

$$\mathrm{Var}\Big(\sum_{i=1}^{m} a_i \chi_{n,A}(t_i))\Big)$$

is asymptotically equal to n up to multiplicative constants. Minkowski's inequality implies that

$$\left(\mathbb{E}\Big[|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]|^p\Big]\right)^{1/p} \le \left(\mathbb{E}\Big[|\xi_{j,n}|^p\Big]\right)^{1/p} + \mathbb{E}\Big[|\xi_{j,n}|\Big].$$

Recall that for fixed $\mathcal{Y} \subset \mathbb{R}^d$, $h_t^k(\mathcal{Y})$ is nondecreasing in t. Then, we have that

$$
\begin{aligned}
|\xi_{j,n}| &\le \sum_{k=0}^{\infty} \sum_{\mathcal{Y}\subset\mathcal{P}_n} \sum_{i=1}^{m} |a_i| h^k_{r_n(t_i)}(\mathcal{Y}) \mathbb{1}\Big\{\mathrm{LMP}(\mathcal{Y}) \in A \cap Q_{j,n}\Big\} \\
&\le C^* \sum_{k=0}^{\infty} \sum_{\mathcal{Y}\subset\mathcal{P}_n} h^k_{r_n(t_m)}(\mathcal{Y}) \mathbb{1}\Big\{\mathrm{LMP}(\mathcal{Y}) \in A \cap Q_{j,n}\Big\} \\
&\le C^* \sum_{k=0}^{\infty} \binom{\mathcal{P}_n\Big(\mathrm{Tube}(Q_{j,n}, r_n(t_m))\Big)}{k+1} \\
&\le C^* \cdot 2^{\mathcal{P}_n(\mathrm{Tube}(Q_{j,n}, r_n(t_m)))},
\end{aligned}
$$

where

$$
\mathrm{Tube}\Big(Q_{j,n}, r_n(t_m)\Big) = \Big\{x \in \mathbb{R}^d : \inf_{y \in Q_{j,n}} \|x - y\| \le r_n(t_m)\Big\}.
$$

By the assumption $ns_n^d = 1$, one can easily show that $\mathcal{P}_n\Big(\mathrm{Tube}(Q_{j,n},\, r_n(t_m))\Big)$ is stochastically dominated by a Poisson random variable with positive and finite parameter, which does not depend on j and n. Denote such a Poisson random variable by Y. Then, for $p = 3, 4$,

$$
\max_j \mathbb{E}\Big[\Big|\xi_{j,n} - \mathbb{E}[\xi_{j,n}]\Big|^p\Big] \le C^* \Big[\Big(\mathbb{E}[2^{pY}]\Big)^{1/p} + \mathbb{E}(2^Y)\Big] < \infty.
$$

Referring back to (4.26) and noting $ns_n^d = 1$, we can see that

$$
\Big|\mathbb{P}\Big(\sum_{j\in H_n} \bar{\xi}_{j,n} \le x\Big) - \mathbb{P}(Z \le x)\Big| \le C^* \Big(\sqrt{s_n^{-d} n^{-3/2}} + \sqrt{s_n^{-d} n^{-2}}\Big) = O(n^{-1/4}) \to 0, \quad n \to \infty,
$$

which implies that $\sum_{j\in H_n} \bar{\xi}_{j,n} \Rightarrow \mathcal{N}(0,1)$ as $n \to \infty$; equivalently,

$$
\sum_{i=1}^{m} a_i \bar{\chi}_{n,A}(t_i) \Rightarrow \mathcal{N}(0, \Sigma_A), \quad n \to \infty,
$$

where

$$
\Sigma_A := \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \sum_{k_i=0}^{\infty} \sum_{k_j=0}^{\infty} (-1)^{k_i+k_j} \Psi_{k_i,k_j,A}(t_i, t_j).
$$

Subsequently we claim that

$$
\sum_{i=1}^{m} a_i \bar{\chi}_n(t_i) \Rightarrow \mathcal{N}(0, \Sigma_{\mathbb{R}^d}), \quad n \to \infty,
$$

which completes the proof. To show this, take $A_K = (-K, K)^d$ for $K > 0$. It then suffices to verify that

$$\mathcal{N}(0, \Sigma_{A_K}) \Rightarrow \mathcal{N}(0, \Sigma_{\mathbb{R}^d}), \quad K \to \infty,$$

and for each $t \geq 0$ and $\epsilon > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(\Big|\bar{\chi}_n(t) - \bar{\chi}_{n,A_K}(t)\Big| > \epsilon\Big) = 0.$$

The former condition is obvious from $\Sigma_{A_K} \to \Sigma_{\mathbb{R}^d}$ as $K \to \infty$, and the induced convergence of the corresponding characteristic functions. The latter is also a direct consequence of Proposition 4.2.1, together with Chebyshev's inequality and the fact that $\chi_n(t) - \chi_{n,A_K}(t) = \chi_{n,\mathbb{R}^d\setminus A_K}(t)$. □

Before we begin the proof of tightness for Theorem 4.4.1, a few more useful properties of h_t^k are needed—these are supplied in Lemma 4.4.3. For $0 \leq s < t < \infty$, we denote

$$h_{t,s}^k(\mathcal{Y}) = h_t^k(\mathcal{Y}) - h_s^k(\mathcal{Y}), \quad \mathcal{Y} = (y_0, \dots, y_k) \in (\mathbb{R}^d)^{k+1}.$$

Lemma 4.4.3.

(i) For any $0 \leq s \leq t \leq T < \infty$,

$$\int_{(\mathbb{R}^d)^k} h_{t,s}^k(0, y_1, \dots, y_k)\, \mathrm{d}\mathbf{y} \leq C_{d,k,T}(t^d - s^d),$$

where $C_{d,k,T} = k^2 \omega_d^k T^{d(k-1)}$.

(ii) Let $j \in \{1, \dots, (k_1 \wedge k_2) + 1\}$ *and suppose that* $\mathbf{y}_0 \in (\mathbb{R}^d)^{j-1}$, $\mathbf{y}_1 \in (\mathbb{R}^d)^{k_1+1-j}$ *and* $\mathbf{y}_2 \in (\mathbb{R}^d)^{k_2+1-j}$. *Then, for* $0 \leq t_1 \leq s \leq t_2 \leq T < \infty$,

$$\int_{(\mathbb{R}^d)^{k_1+k_2+1-j}} h_{s,t_1}^{k_1}(0, \mathbf{y}_0, \mathbf{y}_1) h_{t_2,s}^{k_2}(0, \mathbf{y}_0, \mathbf{y}_2)\, \mathrm{d}\mathbf{y}_0\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}\mathbf{y}_2 \\ \leq 36(k_1k_2)^6 (T^d\omega_d)^{2(k_1+k_2)}(t_2^d - t_1^d)^2.$$

Proof. We note that for any $0 \le s < t$ with $y_0 \equiv 0$,

$$
\begin{aligned}
h_{t,s}^k(0, y_1, \dots, y_k) &= 1\Big\{s < \max_{0\le i<j\le k} \big\|y_i - y_j\big\| \le t\Big\} \\
&\le \prod_{i=1}^k 1\Big\{y_i \in B(0,T)\Big\}\Bigg(\sum_{i=1}^k 1\Big\{s < \|y_i\| \le t\Big\} + \sum_{1\le i<j\le k} 1\Big\{s < \big\|y_i - y_j\big\| \le t\Big\}\Bigg).
\end{aligned}
$$

For each $i = 1, \dots, k$, let $\mathbf{y}^{(i)}$ be the tuple $(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_k) \in (\mathbb{R}^d)^{k-1}$ with the i^{th} coordinate omitted. Then,

$$
\begin{aligned}
\int_{(\mathbb{R}^d)^k} h_{t,s}^k(0, y_1, \dots, y_k)\,\mathrm{d}\mathbf{y} &\le \sum_{i=1}^k \int_{B(0,T)^{k-1}} \int_{\mathbb{R}^d} 1\Big\{s < \|y_i\| \le t\Big\}\,\mathrm{d}y_i\,\mathrm{d}\mathbf{y}^{(i)} \\
&\quad + \sum_{1\le i<j\le k} \int_{B(0,T)^{k-1}} \int_{\mathbb{R}^d} 1\Big\{s < \big\|y_i - y_j\big\| \le t\Big\}\,\mathrm{d}y_i\,\mathrm{d}\mathbf{y}^{(i)}. \\
&= \left(k + \binom{k}{2}\right) m\Big(B(0,T)\Big)^{k-1}\Big[m\Big(B(0,t)\Big) - m\Big(B(0,s)\Big)\Big] \\
&\le C_{d,k,T}(t^d - s^d)
\end{aligned}
$$

as required.

Part (ii) is essentially the same as Lemma 7.1 in [67], so the proof is skipped. □

Proof of tightness in Theorem 4.4.1. We now demonstrate tightness using Theorem 2.3.2, which was adapted from [10]. Hence, if we can show that for every $0 < T < \infty$, there exists a $C > 0$ such that

$$
\mathbb{E}\Big[|\bar{\chi}_n(t_2) - \bar{\chi}_n(s)|^2 |\bar{\chi}_n(s) - \bar{\chi}_n(t_1)|^2\Big] \le C(t_2^d - t_1^d)^2, \tag{4.27}
$$

for all $0 \le t_1 \le s \le t_2 \le T$ and $n \in \mathbb{N}$—we are finished. To demonstrate (4.27), we will give an abridged proof – tightness will be similarly established for analogous processes seen in [67, 75]. Let us begin with some helpful notation, namely,

$$
\begin{aligned}
h_{n,t,s}^k(\mathcal{Y}) &:= h_{r_n(t), r_n(s)}^k(\mathcal{Y}) = h_{r_n(t)}^k(\mathcal{Y}) - h_{r_n(s)}^k(\mathcal{Y}), \\
\zeta_{n,t,s}^k &:= S_{k,n}(t) - S_{k,n}(s) = \sum_{\mathcal{Y}\subset\mathcal{P}_n} h_{n,t,s}^k(\mathcal{Y}).
\end{aligned}
$$

By the same argument as in the proof of Proposition 4.2.1, one can apply Fubini's theorem to obtain

$$\begin{aligned}
&\mathbb{E}\Big[|\bar{\chi}_n(t_2)-\bar{\chi}_n(s)|^2|\bar{\chi}_n(s)-\bar{\chi}_n(t_1)|^2\Big] \qquad (4.28)\\
&\quad=\frac{1}{n^2}\sum_{(k_1,k_2,k_3,k_4)\in\mathbb{N}_0^4}(-1)^{k_1+k_2+k_3+k_4}\mathbb{E}\Big[\Big(\zeta^{k_1}_{n,t_2,s}-\mathbb{E}[\zeta^{k_1}_{n,t_2,s}]\Big)\Big(\zeta^{k_2}_{n,t_2,s}-\mathbb{E}[\zeta^{k_2}_{n,t_2,s}]\Big)\\
&\qquad\qquad\times\Big(\zeta^{k_3}_{n,s,t_1}-\mathbb{E}[\zeta^{k_3}_{n,s,t_1}]\Big)\Big(\zeta^{k_4}_{n,s,t_1}-\mathbb{E}[\zeta^{k_4}_{n,s,t_1}]\Big)\Big].
\end{aligned}$$

Our objective now is to find a suitable bound for

$$\mathbb{E}\Big[\Big(\zeta^{k_1}_{n,t_2,s}-\mathbb{E}[\zeta^{k_1}_{n,t_2,s}]\Big)\Big(\zeta^{k_2}_{n,t_2,s}-\mathbb{E}[\zeta^{k_2}_{n,t_2,s}]\Big)\Big(\zeta^{k_3}_{n,s,t_1}-\mathbb{E}[\zeta^{k_3}_{n,s,t_1}]\Big)\Big(\zeta^{k_4}_{n,s,t_1}-\mathbb{E}[\zeta^{k_4}_{n,s,t_1}]\Big)\Big]. \qquad (4.29)$$

To this end, let us refine the notation once more by denoting $\xi_1:=\zeta^{k_1}_{n,t_2,s}$, $\xi_2:=\zeta^{k_2}_{n,t_2,s}$, $\xi_3:=\zeta^{k_3}_{n,s,t_1}$ and $\xi_4:=\zeta^{k_4}_{n,s,t_1}$. Furthermore, let $h_1:=h^{k_1}_{n,t_2,s}$, $h_2:=h^{k_2}_{n,t_2,s}$, $h_3:=h^{k_3}_{n,s,t_1}$ and $h_4:=h^{k_4}_{n,s,t_1}$. Recall $[n]=\{1,2,\dots,n\}$ and for any $\sigma\subset[4]$ let $\xi_\sigma=\prod_{i\in\sigma}\xi_i$ where we set $\xi_\emptyset=1$ by convention. Then we can express (4.29) quite simply as

$$\sum_{\sigma\subset[4]}(-1)^{|\sigma|}\,\mathbb{E}[\xi_\sigma]\prod_{i\in[4]\setminus\sigma}\mathbb{E}[\xi_i]. \qquad (4.30)$$

For $\sigma\subset[4]$ with $\sigma\neq\emptyset$, and finite subsets $\mathcal{Y}_j\subset\mathbb{R}^d$, $j\in\sigma$, we define $\mathcal{Y}_\sigma:=\bigcup_{j\in\sigma}\mathcal{Y}_j$. Given a subset $\tau\subset\sigma\subset[4]$, we also define

$$\begin{aligned}
\mathcal{I}_{\tau,\sigma}(\mathcal{Y}_\sigma):=&\prod_{j\in\tau}1\Big\{\text{there exists } p\in\tau\setminus\{j\}\text{ such that }\mathcal{Y}_j\cap\mathcal{Y}_p\neq\emptyset\Big\}\\
&\times\prod_{j\in\sigma\setminus\tau}1\Big\{\mathcal{Y}_j\cap\mathcal{Y}_q=\emptyset\text{ for all }q\in\sigma\setminus\{j\}\Big\}.
\end{aligned}$$

Note that $\mathcal{I}_{\tau,\sigma}(\mathcal{Y}_\sigma)=0$ whenever $|\tau|=1$, and

$$\sum_{\tau\subset\sigma}\mathcal{I}_{\tau,\sigma}(\mathcal{Y}_\sigma)=1. \qquad (4.31)$$

Furthermore, if $\tau = \sigma$, we write $\mathcal{I}_\sigma(\cdot) := \mathcal{I}_{\sigma,\sigma}(\cdot)$. It follows from (4.31) and Palm theory in Lemma 2.2.4 that, for each non-empty $\sigma \subset [4]$,

$$\begin{aligned}
\mathbb{E}[\xi_\sigma] &= \mathbb{E}\Big[\sum_{\mathcal{Y}_j \subset \mathcal{P}_n,\, j \in \sigma} \prod_{i \in \sigma} h_i(\mathcal{Y}_i)\Big] \\
&= \sum_{\tau \subset \sigma} \mathbb{E}\Big[\sum_{\mathcal{Y}_j \subset \mathcal{P}_n,\, j \in \sigma} \mathcal{I}_{\tau,\sigma}(\mathcal{Y}_\sigma) \prod_{i \in \sigma} h_i(\mathcal{Y}_i)\Big] \\
&= \sum_{\tau \subset \sigma} \mathbb{E}\Big[\sum_{\mathcal{Y}_j \subset \mathcal{P}_n,\, j \in \tau} \mathcal{I}_{\tau}(\mathcal{Y}_\tau) \prod_{i \in \tau} h_i(\mathcal{Y}_i)\Big] \prod_{i \in \sigma \setminus \tau} \mathbb{E}[\xi_i].
\end{aligned}$$

Hence, (4.30) is equal to

$$\begin{aligned}
&\sum_{\sigma \subset [4]} \sum_{\tau \subset \sigma} (-1)^{|\sigma|}\, \mathbb{E}\Big[\sum_{\mathcal{Y}_j \subset \mathcal{P}_n,\, j \in \tau} \mathcal{I}_{\tau}(\mathcal{Y}_\tau) \prod_{i \in \tau} h_i(\mathcal{Y}_i)\Big] \prod_{i \in \sigma \setminus \tau} \mathbb{E}[\xi_i] \prod_{i \in [4] \setminus \sigma} \mathbb{E}[\xi_i] \\
&= \sum_{\tau \subset [4]} \mathbb{E}\Big[\sum_{\mathcal{Y}_j \subset \mathcal{P}_n,\, j \in \tau} \mathcal{I}_{\tau}(\mathcal{Y}_\tau) \prod_{i \in \tau} h_i(\mathcal{Y}_i)\Big] \prod_{i \in [4] \setminus \tau} \mathbb{E}[\xi_i] \sum_{\tau \subset \sigma \subset [4]} (-1)^{|\sigma|} \\
&= \mathbb{E}\Big[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_3 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_4 \subset \mathcal{P}_n} \mathcal{I}_{[4]}(\mathcal{Y}_{[4]}) \prod_{i=1}^{4} h_i(\mathcal{Y}_i)\Big],
\end{aligned}$$

where the last line follows from $\sum_{\tau \subset \sigma \subset [4]} (-1)^{|\sigma|} = \binom{4-|\tau|}{0}(-1)^{|\tau|} + \cdots + \binom{4-|\tau|}{4-|\tau|}(-1)^4 = 0$, unless $\tau = [4]$. Substituting this back into (4.28) and taking the absolute value of $(-1)^{k_1+k_2+k_3+k_4}$, we get

$$\begin{aligned}
&\mathbb{E}\Big[|\bar{\chi}_n(t_2) - \bar{\chi}_n(s)|^2 |\bar{\chi}_n(s) - \bar{\chi}_n(t_1)|^2\Big] \\
&\quad \le \sum_{(k_1,k_2,k_3,k_4) \in \mathbb{N}_0^4} \frac{1}{n^2} \mathbb{E}\Big[\sum_{\mathcal{Y}_1 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_2 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_3 \subset \mathcal{P}_n} \sum_{\mathcal{Y}_4 \subset \mathcal{P}_n} \mathcal{I}_{[4]}(\mathcal{Y}_{[4]}) \prod_{i=1}^{4} h_i(\mathcal{Y}_i)\Big].
\end{aligned}$$

Now, it suffices to show that the right-hand side above is less than $C(t_2^d - t_1^d)^2$ for some $C > 0$. We can break the above summand into four distinct cases:

(I) $b_{12} = |\mathcal{Y}_1 \cap \mathcal{Y}_2| > 0$, $b_{34} = |\mathcal{Y}_3 \cap \mathcal{Y}_4| > 0$ with all other pairwise intersections empty.

(II) $b_{13} = |\mathcal{Y}_1 \cap \mathcal{Y}_3| > 0$, $b_{24} = |\mathcal{Y}_2 \cap \mathcal{Y}_4| > 0$ with all other pairwise intersections empty.

(III) $b_{14} = |\mathcal{Y}_1 \cap \mathcal{Y}_4| > 0$, $b_{23} = |\mathcal{Y}_2 \cap \mathcal{Y}_3| > 0$ with all other pairwise intersections empty.

(IV) For each i, there exists a $j \neq i$ such that $\mathcal{Y}_i \cap \mathcal{Y}_j \neq \emptyset$, but **(I)**-**(III)** do not hold.

We prove appropriate upper bounds for cases **(I)** and **(IV)**, and the other two cases follow from the proof for **(I)**. Palm theory in Lemma 2.2.4(iv) implies that

$$
\begin{aligned}
&\frac{1}{n^2}\mathbb{E}\Big[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n}\sum_{\mathcal{Y}_2\subset\mathcal{P}_n}\sum_{\mathcal{Y}_3\subset\mathcal{P}_n}\sum_{\mathcal{Y}_4\subset\mathcal{P}_n}\prod_{i=1}^4 h_i(\mathcal{Y}_i) \qquad (4.32)\\
&\qquad\qquad \times 1\Big\{|\mathcal{Y}_1\cap\mathcal{Y}_2|=b_{12}, |\mathcal{Y}_3\cap\mathcal{Y}_4|=b_{34}, |\mathcal{Y}_i\cap\mathcal{Y}_j|=0 \text{ for other } i,j\text{'s}\Big\}\Big]\\
&=\frac{1}{n^2}\mathbb{E}\Big[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n}\sum_{\mathcal{Y}_2\subset\mathcal{P}_n} h_1(\mathcal{Y}_1)h_2(\mathcal{Y}_2)1\Big\{|\mathcal{Y}_1\cap\mathcal{Y}_2|=b_{12}\Big\}\Big]\\
&\qquad\qquad \times \mathbb{E}\Big[\sum_{\mathcal{Y}_3\subset\mathcal{P}_n}\sum_{\mathcal{Y}_4\subset\mathcal{P}_n} h_3(\mathcal{Y}_3)h_4(\mathcal{Y}_4)1\Big\{|\mathcal{Y}_3\cap\mathcal{Y}_4|=b_{34}\Big\}\Big]\\
&=\frac{n^{k_1+k_2+1-b_{12}}}{b_{12}!(k_1+1-b_{12})!(k_2+1-b_{12})!}\\
&\qquad\qquad \times\mathbb{E}\Big[h_1(X_1,\dots,X_{k_1+1})h_2(X_1,\dots,X_{b_{12}},X_{k_1+2},\dots,X_{k_1+k_2+2-b_{12}})\Big]\\
&\times\frac{n^{k_3+k_4+1-b_{34}}}{b_{34}!(k_3+1-b_{34})!(k_4+1-b_{34})!}\\
&\qquad\qquad \times\mathbb{E}\Big[h_3(X_1,\dots,X_{k_3+1})h_4(X_1,\dots,X_{b_{34}},X_{k_3+2},\dots,X_{k_3+k_4+2-b_{34}})\Big].
\end{aligned}
$$

For the remainder of the proof, assume that $T^d\omega_d > 1$, $\|f\|_\infty > 1$ and $T > 1$ for ease of description. Moreover, assume, without loss of generality that $k_1 \ge k_2$ and $k_3 \ge k_4$. Using trivial bounds and the customary changes of variable, i.e., $x_1 = x$, $x_i = x + s_n y_{i-1}$, $i = 2,\dots,k_1+k_2+2-b_{12}$, applying Lemma 4.4.3($i$), and recalling $a_T = T^d\omega_d\|f\|_\infty$, we see that

$$
\begin{aligned}
&n^{k_1+k_2+1-b_{12}}\mathbb{E}[h_1(X_1,\dots,X_{k_1+1})h_2(X_1,\dots,X_{b_{12}},X_{k_1+2},\dots,X_{k_1+k_2+2-b_{12}})]\\
&\le(\|f\|_\infty)^{k_1+k_2+1-b_{12}}\int_{(\mathbb{R}^d)^{k_2+1-b_{12}}}\int_{(\mathbb{R}^d)^{k_1+1-b_{12}}}\int_{(\mathbb{R}^d)^{b_{12}-1}} h_{t_2,s}^{k_1}(0,\mathbf{y}_0,\mathbf{y}_1)\\
&\qquad\qquad\qquad\qquad\qquad\qquad \times h_{t_2,s}^{k_2}(0,\mathbf{y}_0,\mathbf{y}_2)\,\mathrm{d}\mathbf{y}_0\,\mathrm{d}\mathbf{y}_1\,\mathrm{d}\mathbf{y}_2\\
&\le(\|f\|_\infty)^{k_1+k_2}\Big(T^d\omega_d\Big)^{k_2+1-b_{12}}\int_{(\mathbb{R}^d)^{k_1+1-b_{12}}}\int_{(\mathbb{R}^d)^{b_{12}-1}} h_{t_2,s}^{k_1}(0,\mathbf{y}_0,\mathbf{y}_1)\,\mathrm{d}\mathbf{y}_0\,\mathrm{d}\mathbf{y}_1\\
&\le(\|f\|_\infty)^{k_1+k_2}\Big(T^d\omega_d\Big)^{k_2+1-b_{12}}C_{d,k_1,T}(t_2^d-s^d)\\
&\le k_1^2(a_T)^{k_1+k_2}(t_2^d-s^d).
\end{aligned}
$$

Hence, (4.32) is bounded by

$$\begin{aligned}&\frac{(a_T)^{k_1+k_2}k_1^2}{b_{12}!(k_1+1-b_{12})!(k_2+1-b_{12})!}(t_2^d-s^d)\frac{(a_T)^{k_3+k_4}k_3^2}{b_{34}!(k_3+1-b_{34})!(k_4+1-b_{34})!}(s^d-t_1^d)\\&\le\frac{(a_T)^{k_1+k_2+k_3+k_4}k_1^2k_3^2}{b_{12}!(k_1+1-b_{12})!(k_2+1-b_{12})!b_{34}!(k_3+1-b_{34})!(k_4+1-b_{34})!}(t_2^d-t_1^d)^2.\end{aligned}$$

Finally we see that

$$\sum_{\substack{k_1\ge k_2,\,k_3\ge k_4,\\1\le b_{12}\le k_2+1,\\1\le b_{34}\le k_4+1}}\frac{(a_T)^{k_1+k_2+k_3+k_4}k_1^2k_3^2}{b_{12}!(k_1+1-b_{12})!(k_2+1-b_{12})!b_{34}!(k_3+1-b_{34})!(k_4+1-b_{34})!}<\infty,$$

since

$$\begin{aligned}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{k_1}\sum_{\ell=1}^{k_2+1}\frac{(a_T)^{k_1+k_2}k_1^2}{\ell!(k_1+1-\ell)!(k_2+1-\ell)!}&=\sum_{\ell=1}^{\infty}\sum_{k_1=\ell-1}^{\infty}\frac{(a_T)^{k_1}k_1^2}{\ell!(k_1+1-\ell)!}\sum_{k_2=\ell-1}^{k_1}\frac{(a_T)^{k_2}}{(k_2+1-\ell)!}\\&\le e^{a_T}\sum_{\ell=1}^{\infty}\frac{(a_T)^{\ell-1}}{\ell!}\sum_{k_1=\ell-1}^{\infty}\frac{(a_T)^{k_1}k_1^2}{(k_1+1-\ell)!}<\infty.\end{aligned}\tag{4.33}$$

Now, for cases **(I)** - **(III)**, we have an upper bound of the form $C(t_2^d-t_1^d)^2$ as desired.

Thus we need only demonstrate the same for case **(IV)**. In addition to the notation b_{ij}, $1\le i<j\le 4$ as above, define for $\mathcal{Y}_i\in(\mathbb{R}^d)^{k_i+1}$, $k_i\in\mathbb{N}_0$, $i=1,\dots,4$,

$$b_{ijk}:=|\mathcal{Y}_i\cap\mathcal{Y}_j\cap\mathcal{Y}_k|,\ 1\le i<j<k\le 4,\quad\text{and}\quad b_{1234}:=|\mathcal{Y}_1\cap\mathcal{Y}_2\cap\mathcal{Y}_3\cap\mathcal{Y}_4|,$$

and also recall the definition of b on p. 50,

$$b=b_{12}+b_{13}+b_{14}+b_{23}+b_{24}+b_{34}-b_{123}-b_{124}-b_{134}-b_{234}+b_{1234},\tag{4.34}$$

so that $|\mathcal{Y}_1\cup\mathcal{Y}_2\cup\mathcal{Y}_3\cup\mathcal{Y}_4|=k_1+k_2+k_3+k_4+4-b$ with $b\ge 3$. Let $\mathcal{B}$ be the collection of $\mathbf{b}=(b_{12},\dots,b_{1234})\in\mathbb{N}_0^{11}$ satisfying the conditions in Case **(IV)**. For a non-empty $\sigma\subset[4]$ and $\mathcal{X}_i\in(\mathbb{R}^d)^{k_i+1}$, $i=1,\dots,4$, recall (also from p. 50) that

$$j_\sigma:=\left|\bigcap_{i\in\sigma}\Big(\mathcal{X}_i\setminus\bigcup_{j\in[4]\setminus\sigma}\mathcal{X}_j\Big)\right|\tag{4.35}$$

In particular, j_σ's are functions of $\mathbf{b}$ such that $\sum_{\sigma\subset[4],\,\sigma\neq\emptyset} j_\sigma = |\mathcal{X}_1\cup\mathcal{X}_2\cup\mathcal{X}_3\cup\mathcal{X}_4|$. Then, Lemma 2.2.4($iv$) yields

$$\begin{aligned}
&\frac{1}{n^2}\,\mathbb{E}\Big[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n}\sum_{\mathcal{Y}_2\subset\mathcal{P}_n}\sum_{\mathcal{Y}_3\subset\mathcal{P}_n}\sum_{\mathcal{Y}_4\subset\mathcal{P}_n}\prod_{i=1}^4 h_i(\mathcal{Y}_i)\,1\Big\{\text{case }\textbf{(IV)}\text{ holds}\Big\}\Big]\\
&=\sum_{\mathbf{b}\in\mathcal{B}}\frac{1}{n^2}\,\mathbb{E}\Big[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n}\sum_{\mathcal{Y}_2\subset\mathcal{P}_n}\sum_{\mathcal{Y}_3\subset\mathcal{P}_n}\sum_{\mathcal{Y}_4\subset\mathcal{P}_n}\prod_{i=1}^4 h_i(\mathcal{Y}_i)\,1\Big\{|\mathcal{Y}_1\cap\mathcal{Y}_2|=b_{12},|\mathcal{Y}_1\cap\mathcal{Y}_3|=b_{13},\\
&\qquad\qquad\dots,|\mathcal{Y}_1\cap\mathcal{Y}_2\cap\mathcal{Y}_3\cap\mathcal{Y}_4|=b_{1234}\Big\}\Big]\\
&=\sum_{\mathbf{b}\in\mathcal{B}}\frac{n^{k_1+k_2+k_3+k_4+2-b}}{\prod_{\sigma\subset[4],\,\sigma\neq\emptyset} j_\sigma!}\,\mathbb{E}\Big[\prod_{i=1}^4 h_i(\mathcal{X}_i)\,1\Big\{|\mathcal{X}_1\cap\mathcal{X}_2|=b_{12},|\mathcal{X}_1\cap\mathcal{X}_3|=b_{13},\\
&\qquad\qquad\dots,|\mathcal{X}_1\cap\mathcal{X}_2\cap\mathcal{X}_3\cap\mathcal{X}_4|=b_{1234}\Big\}\Big].
\end{aligned}$$

Under the conditions in Case **(IV)**, at least one of the b_{ij}'s is non-zero, so we may assume without loss of generality that $b_{13}>0$. Then we have

$$\begin{aligned}
&n^{k_1+k_2+k_3+k_4+2-b}\mathbb{E}\Big[\prod_{i=1}^4 h_i(\mathcal{X}_i)\,1\Big\{|\mathcal{X}_1\cap\mathcal{X}_2|=b_{12},\dots,|\mathcal{X}_1\cap\mathcal{X}_2\cap\mathcal{X}_3\cap\mathcal{X}_4|=b_{1234}\Big\}\Big]\\
&=n^{k_1+k_2+k_3+k_4+2-b}\int_{(\mathbb{R}^d)^{k_1+k_2+k_3+k_4+4-b}} h_1(\mathbf{x}_0,\mathbf{x}_1)h_3(\mathbf{x}_0,\mathbf{x}_3)h_2(\mathbf{x}_2)h_4(\mathbf{x}_4)\\
&\qquad\times\prod_{x\in\bigcup_{i=0}^4\mathbf{x}_i} f(x)\,\mathrm{d}(\mathbf{x}_0\cup\mathbf{x}_1\cup\cdots\cup\mathbf{x}_4),
\end{aligned}$$

where $\mathbf{x}_0$ is a collection of elements in $\mathbb{R}^d$ with $|\mathbf{x}_0|=b_{13}>0$. In other words, $\mathbf{x}_0\in(\mathbb{R}^d)^{b_{13}}$, so that $\mathbf{x}_1\in(\mathbb{R}^d)^{k_1+1-b_{13}}$ and $\mathbf{x}_3\in(\mathbb{R}^d)^{k_3+1-b_{13}}$ with $\mathbf{x}_1\cap\mathbf{x}_3=\emptyset$. Moreover, $\mathbf{x}_2\in(\mathbb{R}^d)^{k_2+1}$ and $\mathbf{x}_4\in(\mathbb{R}^d)^{k_4+1}$, such that if $\mathbf{x}_2\cap\mathbf{x}_4=\emptyset$, then $\mathbf{x}_i\cap(\mathbf{x}_0\cup\mathbf{x}_1\cup\mathbf{x}_3)\neq\emptyset$ for $i=2,4$, and if $\mathbf{x}_2\cap\mathbf{x}_4\neq\emptyset$, then $(\mathbf{x}_2\cup\mathbf{x}_4)\cap(\mathbf{x}_0\cup\mathbf{x}_1\cup\mathbf{x}_3)\neq\emptyset$.

Now, let us perform a change of variables by $\mathbf{x}_i=x1+s_n\mathbf{y}_i$ for $i=0,\dots,4$, where 1 is a vector with all entries 1, and the first element of $\mathbf{y}_0$ is taken to be 0. In addition to this, we apply the translation and scale invariance of h_i's to get

$$\begin{aligned}
&n^{k_1+k_2+k_3+k_4+2-b}\mathbb{E}\Big[\prod_{i=1}^4 h_i(\mathcal{X}_i)\,1\Big\{|\mathcal{X}_1\cap\mathcal{X}_2|=b_{12},\dots,|\mathcal{X}_1\cap\mathcal{X}_2\cap\mathcal{X}_3\cap\mathcal{X}_4|=b_{1234}\Big\}\Big]\\
&=n^{k_1+k_2+k_3+k_4+2-b}s_n^{d(k_1+k_2+k_3+k_4+3-b)}\int_{\mathbb{R}^d}\int_{(\mathbb{R}^d)^{k_1+k_2+k_3+k_4+3-b}} h_{t_2,s}^{k_1}(\mathbf{y}_0,\mathbf{y}_1)h_{s,t_1}^{k_3}(\mathbf{y}_0,\mathbf{y}_3)
\end{aligned}$$

$$\times\, h^{k_2}_{t_2,s}(\mathbf{y}_2)h^{k_4}_{s,t_1}(\mathbf{y}_4) \prod_{y\in\bigcup_{i=0}^4 \mathbf{y}_i} f(x+s_n y)\,\mathrm{d}\Big((\mathbf{y}_0\cup\cdots\cup\mathbf{y}_4)\setminus\{0\}\Big)\,\mathrm{d}x.$$

Using $ns_n^d = 1$, together with the trivial bounds $h^{k_2}_{t_2,s}(\mathbf{y}_2) \le h^{k_2}_T(\mathbf{y}_2)$, $h^{k_4}_{s,t_1}(\mathbf{y}_4) \le h^{k_4}_T(\mathbf{y}_4)$, and $f(x+s_n y) \le \|f\|_\infty$, one can bound the last expression by

$$\begin{aligned}
&\|f\|_\infty^{k_1+k_2+k_3+k_4+3-b} \int_{(\mathbb{R}^d)^{k_1+k_2+k_3+k_4+3-b}} h^{k_1}_{t_2,s}(\mathbf{y}_0,\mathbf{y}_1)h^{k_3}_{s,t_1}(\mathbf{y}_0,\mathbf{y}_3)\\
&\qquad\qquad\qquad\qquad\qquad \times h^{k_2}_T(\mathbf{y}_2)h^{k_4}_T(\mathbf{y}_4)\,\mathrm{d}\Big((\mathbf{y}_0\cup\cdots\cup\mathbf{y}_4)\setminus\{0\}\Big)\\
&= \|f\|_\infty^{k_1+k_2+k_3+k_4+3-b} \int_{(\mathbb{R}^d)^{k_1+k_3+1-b_{13}}} h^{k_1}_{t_2,s}(\mathbf{y}_0,\mathbf{y}_1)h^{k_3}_{s,t_1}(\mathbf{y}_0,\mathbf{y}_3)\\
&\quad \times \bigg\{ \int_{(\mathbb{R}^d)^{k_2+k_4+2-b+b_{13}}} h^{k_2}_T(\mathbf{y}_2)h^{k_4}_T(\mathbf{y}_4)\,\mathrm{d}\Big((\mathbf{y}_2\cup\mathbf{y}_4)\setminus(\mathbf{y}_0\cup\mathbf{y}_1\cup\mathbf{y}_3)\Big)\bigg\}\,\mathrm{d}\Big(\mathbf{y}_0\setminus\{0\}\Big)\,\mathrm{d}\mathbf{y}_1\,\mathrm{d}\mathbf{y}_3.
\end{aligned} \tag{4.36}$$

Suppose $h^{k_2}_T(\mathbf{y}_2)h^{k_4}_T(\mathbf{y}_4) = 1$, such that

$$\mathbf{y}_2\cap\mathbf{y}_4 \ne \emptyset,\quad \mathbf{y}_2\cap(\mathbf{y}_0\cup\mathbf{y}_1\cup\mathbf{y}_3) = \emptyset,\quad \mathbf{y}_4\cap(\mathbf{y}_0\cup\mathbf{y}_1\cup\mathbf{y}_3) \ne \emptyset. \tag{4.37}$$

Then, there exists $y' \in \mathbf{y}_4\cap(\mathbf{y}_0\cup\mathbf{y}_1\cup\mathbf{y}_3)$ such that all points in $\mathbf{y}_2$ are at distance at most $2T$ from y'. Since y' itself lies within distance T from the origin (recall that the first element of $\mathbf{y}_0$ is 0), we conclude that all points in $\mathbf{y}_2\cap\mathbf{y}_4$ are at distance at most $3T$ from the origin. As $b_{13} \le k_1+k_3+1$ and $b \ge 3$, we have

$$\begin{aligned}
&\int_{(\mathbb{R}^d)^{k_2+k_4+2-b+b_{13}}} h^{k_2}_T(\mathbf{y}_2)h^{k_4}_T(\mathbf{y}_4)\,\mathrm{d}\Big((\mathbf{y}_2\cup\mathbf{y}_4)\setminus(\mathbf{y}_0\cup\mathbf{y}_1\cup\mathbf{y}_3)\Big)\\
&\le m\Big(B(0,3T)\Big)^{k_2+k_4+2-b+b_{13}} = \Big((3T)^d\omega_d\Big)^{k_2+k_4+2-b+b_{13}} \le \Big((3T)^d\omega_d\Big)^{k_1+k_2+k_3+k_4}.
\end{aligned} \tag{4.38}$$

If $\mathbf{y}_2$ and $\mathbf{y}_4$ do not satisfy (4.37), it is still easy to check (4.38).

Applying (4.38), along with Lemma 4.4.3(ii), one can bound (4.36) by

$$\begin{aligned}
&\|f\|_\infty^{k_1+k_2+k_3+k_4+3-b}\Big((3T)^d\omega_d\Big)^{k_1+k_2+k_3+k_4} \times 36(k_1k_3)^6\Big(T^d\omega_d\Big)^{2(k_1+k_3)}(t_2^d-t_1^d)^2\\
&\le 36(k_1k_2k_3k_4)^6\Big((3T)^d\omega_d\|f\|_\infty\Big)^{3(k_1+k_2+k_3+k_4)}(t_2^d-t_1^d)^2.
\end{aligned}$$

Thus, we conclude that

$$\frac{1}{n^2}\,\mathbb{E}\Big[\sum_{\mathcal{Y}_1\subset\mathcal{P}_n}\sum_{\mathcal{Y}_2\subset\mathcal{P}_n}\sum_{\mathcal{Y}_3\subset\mathcal{P}_n}\sum_{\mathcal{Y}_4\subset\mathcal{P}_n}\prod_{i=1}^{4}h_i(\mathcal{Y}_i)\,1\Big\{\text{case } \textbf{(IV)} \text{ holds}\Big\}\Big]$$
$$\le 36(k_1k_2k_3k_4)^6\Big((3T)^d\omega_d\|f\|_\infty\Big)^{3(k_1+k_2+k_3+k_4)}\sum_{\mathbf{b}\in\mathcal{B}}\frac{1}{\prod_{\sigma\subset[4],\,\sigma\neq\emptyset}j_\sigma!}\,(t_2^d-t_1^d)^2.$$

To complete the proof, we need to show that

$$\sum_{k_1\le k_2\le k_3\le k_4}(k_1k_2k_3k_4)^6\Big((3T)^d\omega_d\|f\|_\infty\Big)^{3(k_1+k_2+k_3+k_4)}\sum_{\mathbf{b}\in\mathcal{B}}\frac{1}{\prod_{\sigma\subset[4],\,\sigma\neq\emptyset}j_\sigma!}<\infty.$$

As seen in the calculation at (4.33), the term $(k_1k_2k_3k_4)^6\Big((3T)^d\omega_d\|f\|_\infty\Big)^{3(k_1+k_2+k_3+k_4)}$ is negligible, while proving

$$\sum_{k_1\le k_2\le k_3\le k_4}\sum_{\mathbf{b}\in\mathcal{B}}\frac{1}{\prod_{\sigma\subset[4],\,\sigma\neq\emptyset}j_\sigma!}<\infty$$

is straightforward. $\square$

Proof of Hölder continuity in Theorem 4.4.1. Since $\mathcal{L}(t)-\mathcal{L}(s)$ has a normal distribution for $0\le s<t<\infty$, we have for every $m\in\mathbb{N}$,

$$\mathbb{E}\Big[\Big(\mathcal{L}(t)-\mathcal{L}(s)\Big)^{2m}\Big]=\prod_{i=1}^{m}(2i-1)\Big(\mathbb{E}\Big[\Big(\mathcal{L}(t)-\mathcal{L}(s)\Big)^2\Big]\Big)^m,$$

Proposition 4.2.1 ensures that $\Big(\sum_{k=0}^{M}(-1)^k\mathcal{L}_k(t),\ M\in\mathbb{N}_0\Big)$ constitutes a Cauchy sequence in $L^2(\Omega)$. Therefore we have

$$\mathbb{E}\Big[\Big(\mathcal{L}(t)-\mathcal{L}(s)\Big)^2\Big]=\lim_{M\to\infty}\mathbb{E}\Big[\Big(\sum_{k=0}^{M}(-1)^k\Big(\mathcal{L}_k(t)-\mathcal{L}_k(s)\Big)\Big)^2\Big]$$
$$\le\Big[\sum_{k=0}^{\infty}\Big\{\mathbb{E}\Big[\Big(\mathcal{L}_k(t)-\mathcal{L}_k(s)\Big)^2\Big]\Big\}^{1/2}\Big]^2,$$

where the second line is due to the Cauchy-Schwarz inequality. We see at once that

$$\mathbb{E}\Big[\Big(\mathcal{L}_k(t)-\mathcal{L}_k(s)\Big)^2\Big]=\Psi_{k,k}(t,t)-2\Psi_{k,k}(t,s)+\Psi_{k,k}(s,s)$$

$$\leq \Psi_{k,k}(t,t) - \Psi_{k,k}(t,s) = \sum_{j=1}^{k+1} \Big(\psi_{j,k,k}(t,t) - \psi_{j,k,k}(t,s)\Big) \tag{4.39}$$

by monotonicity due to the the monotonicity of simplicial construction functions and the symmetry of $\Psi_{k,k}(\cdot,\cdot)$ in its arguments. Now, we note that

$$\begin{aligned}\psi_{j,k,k}(t,t) - \psi_{j,k,k}(t,s) &= \frac{\int_{\mathbb{R}^d} f(x)^{2k+2-j}\,\mathrm{d}x}{j!((k+1-j)!)^2} \\ &\quad \times \int_{(\mathbb{R}^d)^{k+1-j}} \int_{(\mathbb{R}^d)^{k+1-j}} \int_{(\mathbb{R}^d)^{j-1}} h_t^k(0,\mathbf{y}_0,\mathbf{y}_1) h_{t,s}^k(0,\mathbf{y}_0,\mathbf{y}_2)\,\mathrm{d}\mathbf{y}_0\,\mathrm{d}\mathbf{y}_1\,\mathrm{d}\mathbf{y}_2.\end{aligned}$$

Applying a bound $h_t^k(0,\mathbf{y}_0,\mathbf{y}_1) \leq \prod_{y\in\mathbf{y}_1} 1\{\|y\| \leq T\}$ followed by integrating out $\mathbf{y}_1$, as well as using Lemma 4.4.3(i), we get

$$\begin{aligned}\psi_{j,k,k}(t,t) - \psi_{j,k,k}(t,s) &\leq \frac{k^2}{T^d j!\Big((k+1-j)!\Big)^2} (a_T)^{2k+1-j}(t^d - s^d) \\ &\leq \frac{dk^2}{Tj!\Big((k+1-j)!\Big)^2} (a_T)^{2k+1-j}(t-s),\end{aligned}$$

where a_T is given in (4.6). Substituting this back into (4.39), we obtain

$$\begin{aligned}\mathbb{E}\Big[\Big(\mathcal{L}_k(t) - \mathcal{L}_k(s)\Big)^2\Big] &\leq \frac{dk^2}{T} \sum_{j=1}^{k+1} \frac{(a_T)^{2k+1-j}}{j!\Big((k+1-j)!\Big)^2} (t-s) \\ &\leq \frac{dk^2}{T(k+1)!a_T} \Big(a_T(1+a_T)\Big)^{k+1}(t-s).\end{aligned}$$

Therefore, we conclude that

$$\mathbb{E}\Big[\Big(\mathcal{L}(t) - \mathcal{L}(s)\Big)^{2m}\Big] \leq \prod_{i=1}^{m}(2i-1)\Big(\frac{d}{Ta_T}\Big)^m \Bigg(\sum_{k=0}^{\infty} \frac{k\Big(a_T(1+a_T)\Big)^{(k+1)/2}}{\sqrt{(k+1)!}}\Bigg)^{2m} (t-s)^m.$$

One can easily check that the infinite sum on the right hand side converges via the ratio test. As a result, we can apply the Kolmogorov continuity theorem [53]. This implies that there exists a continuous version of $(\mathcal{L}_k(t),\, 0 \leq t \leq T)$ with Hölder continuous sample paths on $[0,T]$ with any exponent $\gamma \in [0,(m-1)/2m)$. As m is arbitrary, we are done by letting $m \to \infty$. □

5. LIMIT THEORY FOR EULER CHARACTERISTIC PROCESSES OF EXTREME SAMPLE CLOUDS

NB: A version of this chapter has been published in *Extremes*, May 2020, DOI: https://doi.org/10.1007/s10687-021-00419-1. Reproduced with permission from Springer Nature.

Extreme value theory often deals with the empirical measures of extreme point clouds [4, 6, 82]. In this section, we generalize the counting of elements in extreme point clouds, as well as advance the study of topological crackle [4, 67, 68, 69, 70, 71], by investigating FSLLNs for the process

$$(\tilde{\chi}_n(t),\, t \geq 0) := \Big(\chi(\mathcal{K}(\Phi_n \cap B(0, R_n)^c, t)),\, t \geq 0\Big), \tag{5.1}$$

where as usual Φ_n has intensity nf. However, in contrast to the previous chapter, wherein we dealt with subconnective random geometric complexes, in this setting there will be a large contractible core of points. Owing to the fact that the behavior of the radius that governs the formation of simplices is well within the dense regime (perhaps one might say *hyperdense*), even though the points appear in regions of relatively low density, we must pay much closer attention to the density f, which we assume spherically symmetric throughout. Hence, we impose certain constraints on the tail behavior of f. We handle the case when R_n coincides with a weak core $R_n^{(w)}$ (see p. 28), so that $nf(R_n e_1) \to \xi \in (0, \infty)$ and simplex counts of each dimension k, $k \in \mathbb{N}_0$, grow at the same rate.

In the next section, we will detail the setup and the background on extreme value theory necessary for this chapter. We begin with a treatment of regular variation needed for this chapter and define our heavy-tailed densities. After this, we prove two FSLLNs for distributions that undergo topological crackle. First, we prove the FSLLN for (5.1) for the case when f is heavy-tailed, and conclude with the FSLLN for the case when f has an exponentially-decaying tail. We will define both of these notions in the upcoming sections. It is important to note that unlike [92] and [42], the proofs here cannot be realized through a direct application of the Borel-Cantelli lemma—the growth rate of $(\tilde{\chi}_n(t),\, t \geq 0)$ is too slow (logarithmic in the case of an exponentially-decaying tail). In the same vein, this chapter drew inspiration from a very early draft of [73], which deals with FSLLNs for Betti number processes in same setup as this chapter—the EVT setup. As mentioned above, the material in this section has been published in the journal *Extremes* [93].

5.1 Setup

Much of the material required for this chapter has been covered in the previous chapters, however there are a few important notions to discuss. We begin by characterizing $(\tilde{\chi}_n(t),\, t \geq 0)$ in a similar manner to (4.2). Namely, $\tilde{\chi}_n(t) := \sum_{k=0}^{\infty}(-1)^k \tilde{S}_{k,n}(t)$, where for $t \geq 0$

$$\tilde{S}_{k,n}(t) := S_k(\mathcal{K}(\Phi_n \cap B(0, R_n)^c, t)). \tag{5.2}$$

In particular, it still holds that $\tilde{S}_{k,n}(t) \equiv 0$ for all $k \geq |\Phi_n|$, so that (5.1) is a.s. a finite sum.

A useful concept in the characterization of heavy tails is that of *regular variation.* A function $g : [0, \infty) \to (0, \infty)$ is said to be *regularly varying* (at infinity) if

$$\lim_{r\to\infty} \frac{g(rt)}{g(r)} = t^{\beta}, \quad t > 0, \tag{5.3}$$

for some $\beta \in \mathbb{R}$. If (5.3) holds we denote this as $g \in RV_\beta$, and we say that β is the *regular variation exponent.* If $\beta = 0$ we say that g is *slowly varying.* If $g \in RV_\beta$ then we may represent $g(t) = t^\beta L(t)$, where L is slowly varying.

We continue with a highly useful inequality for the quotient $g(rt)/g(r)$, that plays a prominent role in establishing the dominated convergence assumption for Lemmas 5.2.3 and 5.3.4.

Lemma 5.1.1 (**Potter's bound**—Proposition 2.6 (ii) in [82]). *Suppose that g is a positive function on $[0, \infty)$ such that $g \in RV_\beta$ for $\beta \in \mathbb{R}$. Then for any $\epsilon > 0$ there exists an r_0 such that for all $r \geq r_0$ and $t \geq 1$ we have*

$$(1 - \epsilon)t^{\beta-\epsilon} < \frac{g(rt)}{g(r)} < (1 + \epsilon)t^{\beta+\epsilon}.$$

Now, we make note of a property of regular variation (and regular variation exponents) of inverse functions. For a nondecreasing function $g : [0, \infty) \to \mathbb{R}$, its inverse is defined

$$g^{\leftarrow}(y) := \inf\{x \in [0, \infty) : g(x) \geq y\}.$$

If additionally the function $g \in RV_\beta$, $\beta \geq 0$, with $\lim_{r\to\infty} g(r) = \infty$, then one has that $g^{\leftarrow} \in RV_{1/\beta}$ [81, Proposition 0.8].

5.2 Heavy tail case

In this section we detail the limiting behavior of (4.2) where the underlying extreme sample cloud consists of points with a heavy-tailed density f. When we speak of a density f with a heavy tail, this means that there exists $\alpha > d$, such that

$$\lim_{r\to\infty} \frac{f(rt\theta)}{f(r\theta)} = t^{-\alpha}, \quad t > 0, \tag{5.4}$$

for every (equivalently, some) $\theta \in S^{d-1}$. By the spherical symmetry assumption, we have $f(r\theta) = f(re_1)$ for all $\theta \in S^{d-1}$, so we may abuse notation a little bit to denote $f(r) := f(re_1)$. Our density $f : \mathbb{R}^d \to [0, \infty)$ may thus be considered regularly varying at infinity with exponent $-\alpha$, if we consider it a function of the value $r \geq 0$. To give further insight into our heavy-tailed density, let us define $\bar{F}(t) = P(\|X\| > t)$ to be the tail function of the distance of each $X_i \in \mathcal{X}_n$ from the origin. Then we have

$$\bar{F}(t) = s_{d-1} \int_t^\infty r^{d-1} f(r)\, \mathrm{d}r,$$

by the standard polar coordinate transform. It is clear that $r \mapsto s_{d-1} r^{d-1} f(r)$ is regularly varying with exponent $d - 1 - \alpha < -1$, thus Karamata's theorem—Theorem 2.1 in [82]—implies that $\bar{F} \in RV_{d-\alpha}$. This means that $\bar{F}$ lies in the max-domain of attraction of the Fréchet distribution $G(x) = \exp\{-x^{-(\alpha-d)}\}$, $x > 0$. A good exposition on conditions for the domains of attraction of the extreme-value distributions can be found in [81].

Finally, since $nf(R_n) \to \xi \in (0,\infty)$, if f assumed eventually nonincreasing this implies that one can take

$$R_n = \xi^{-1/\alpha}\Big(1/f\Big)^{\leftarrow}(n), \tag{5.5}$$

so that $(R_n)_{n\geq 1}$ is a regularly varying sequence of exponent $1/\alpha$. However, we do not necessarily assume the regular variation of R_n. We are now ready to state the FSLLN for a heavy-tailed density.

Theorem 5.2.1. *Let $\mathcal{K}$ be any right-continuous simplicial construction function and take Φ_n to be either Poisson or binomial. Suppose that f is a spherically symmetric probability density with $f \in RV_{-\alpha}$. Then if $nf(R_n) \to \xi \in (0,\infty)$ we have the following functional strong law of large numbers, i.e.,*

$$\left(\frac{\tilde{\chi}_n(t)}{R_n^d},\, t \geq 0\right) \to \left(\sum_{k=0}^{\infty}(-1)^k s_k(t),\, t \geq 0\right), \quad \text{a.s. in } (D,U), \tag{5.6}$$

where

$$s_k(t) := \frac{s_{d-1}\xi^{k+1}}{(k+1)!\Big(\alpha(k+1)-d\Big)} \int_{(\mathbb{R}^d)^k} h_t^k(0,y_1,\dots,y_k)\,\mathrm{d}\mathbf{y},\ \ t \geq 0, \tag{5.7}$$

and $s_0(t) \equiv s_{d-1}\xi/(\alpha-d)$. In particular, the limit in (5.6) is convergent for all $t \geq 0$.

The following example illustrates the uniform convergence that takes place in the above theorem.

Example 5.2.2. Consider the power-law density defined by

$$f(x) = \frac{2}{\pi\omega_d(1+\|x\|^{2d})}, \quad x \in \mathbb{R}^d.$$

Define $R_n := (2n/\pi\omega_d)^{1/(2d)}$, so that $nf(R_n) \to 1$. We consider the Vietoris-Rips complex induced by the filtration function $\kappa(\sigma) = \sqrt{d}(\mathrm{diam}(\sigma))$, where diam is calculated here with respect to the L^∞ norm. Then, it follows from Theorem 5.2.1 that, as $n \to \infty$,

$$\left(\sqrt{\frac{\pi\omega_d}{2n}}\tilde{\chi}_n(t),\, t \geq 0\right) \to \left(\sum_{k=0}^{\infty}(-1)^k s_k(t),\, t \geq 0\right), \ \text{ a.s. in } D[0,\infty).$$

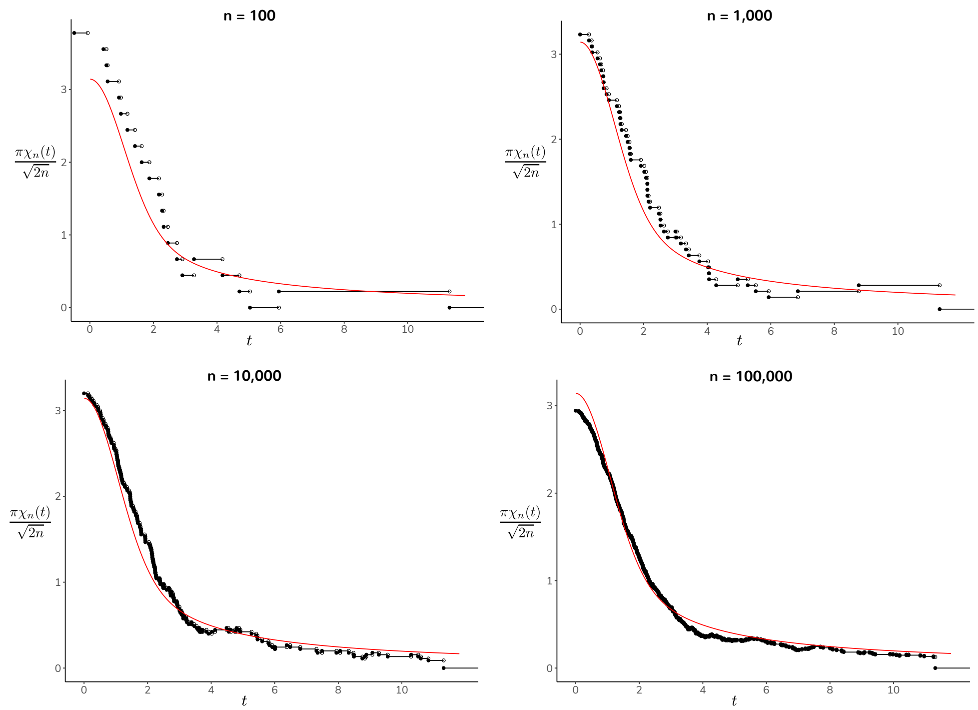


Figure 5.1. Plots of random realizations of $\sqrt{\pi\omega_d/(2n)}\,\tilde{\chi}_n(t)$ for $d = 2$ (in black) in the setup of Example 5.2.2. In the plots above, as n increases from left to right, the random function converges uniformly to $\sum_{k=0}^{\infty}(-1)^k s_k(t)$ (in red).

The limiting function above can be simplified as follows:

$$\sum_{k=0}^{\infty}(-1)^k s_k(t) = s_{d-1}\sum_{k=0}^{\infty}\frac{(-1)^k(t/\sqrt{d})^{dk}}{(k+1)!(2d(k+1)-d)}\int_{(\mathbb{R}^d)^k} h_{\sqrt{d}}(0, y_1, \dots, y_k)\,\mathrm{d}\mathbf{y}$$
$$= \omega_d \sum_{k=0}^{\infty}\frac{(-1)^k(t/\sqrt{d})^{dk}(k+1)^d}{(k+1)!(2k+1)}. \tag{5.8}$$

See Figures 5.1 and 5.2 for actual plots of the limiting function in (5.8) for $d = 2, 3, 4, 5$.

One of the implications of Theorem 5.2.1 is that one can immediately obtain various limit theorems of functions of the Euler characteristic process. For every continuous function T on $D[0, \infty)$, it indeed holds that, as $n \to \infty$,

$$T\left(\sqrt{\frac{\pi\omega_d}{2n}}\tilde{\chi}_n\right) \to T\left(\sum_{k=0}^{\infty}(-1)^k s_k\right), \quad \text{a.s.}$$

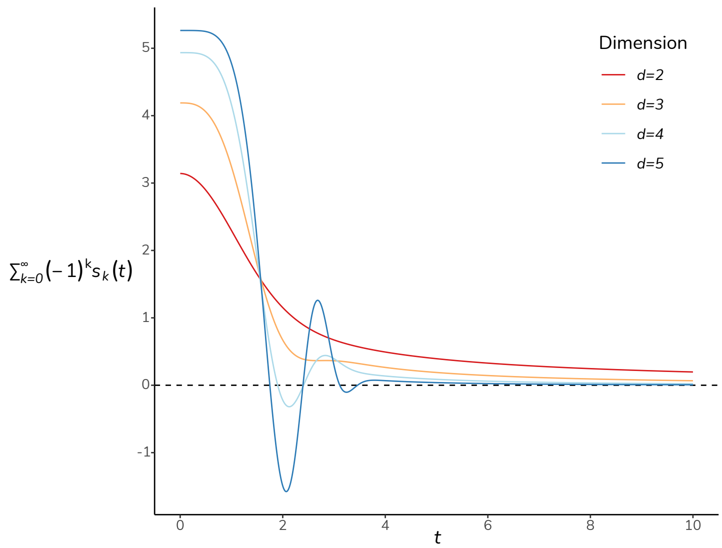


Figure 5.2. Plots of $\sum_{k=0}^{\infty}(-1)^k s_k(t)$ at (5.8) for $d = 2, 3, 4, 5$.

For example, if $U_{a,b} : D[0,\infty) \to [0,\infty)$ is defined by $U_{a,b}(x) := \sup_{a \le x \le b} \left| x(t) \right|$, $0 \le a < b < \infty$, we have

$$\sqrt{\frac{\pi \omega_d}{2n}} \sup_{a \le t \le b} \left| \tilde{\chi}_n(t) \right| \to \sup_{a \le t \le b} \left| \sum_{k=0}^{\infty} (-1)^k s_k(t) \right| \quad \text{a.s.}$$

Furthermore, let $I : D[0,\infty) \to D[0,\infty)$ be defined by $I(x)(t) := \int_0^t x(r)\,\mathrm{d}r$; then,

$$\left(\sqrt{\frac{\pi \omega_d}{2n}} \int_0^t \tilde{\chi}_n(r)\,\mathrm{d}r, \ t \ge 0 \right) \to \left(\sum_{k=0}^{\infty} (-1)^k \int_0^t s_k(r)\,\mathrm{d}r, \ t \ge 0 \right), \quad \text{a.s. in } D[0,\infty).$$

This result is especially important for applications in TDA. Indeed, if we use reduced homology (see [33, 44]) then $\int_0^t \tilde{\chi}_n(r)\,\mathrm{d}r$ represents an alternating sum of the total length of persistence barcodes of all dimensions, up to time t. Recall that a persistence barcode is a graphical descriptor of persistent homology, which allows us to visualize the birth time and death time of cycles [24, 39]. In light of the TDA literature (e.g., Section 6 of [15]), the limit $\int_0^\infty \chi_n(r)\,\mathrm{d}r$ is defined as the Euler characteristic of persistence barcodes of the filtration $\mathcal{K}(\Phi_n \cap B(0, R_n)^c)$. This gives us an estimate of how long the cycles of any dimension live in our extreme sample cloud.

For the proof of Theorem 5.2.1 we use the following collections of strong laws of large numbers for simplex counts.

Proposition 5.2.1. *Suppose that f is a spherically symmetric probability density with $f \in RV_{-\alpha}$ and $nf(R_n) \to \xi \in (0,\infty)$. Then for every $t \geq 0$,*

$$\frac{\tilde{S}_{k,n}(t)}{R_n^d} \to s_k(t), \quad \text{a.s.}, \quad n \to \infty,$$

where $s_k(t)$ is defined at (5.7). Additionally, we have

$$\frac{\chi_n^{(i)}(t)}{R_n^d} \to \sum_{k=0}^{\infty} s_{2k+i}(t), \quad \text{a.s.}, \quad n \to \infty, \tag{5.9}$$

for $i = 0, 1$—where $\chi_n^{(i)}(t) := \sum_{k=0}^{\infty} \tilde{S}_{2k+i,n}(t)$.

To prove Proposition 5.2.1, we need a lemma establishing the asymptotics of the first moment of the sums of simplex counts of various dimensions, as well as the finiteness of the variance of the same quantity.

Lemma 5.2.3. *Let $\mathcal{K}$ be any simplicial construction function, let Φ_n be Poisson or binomial, and suppose that f is spherically symmetric and has a regularly varying tail. Let us assume that $(n_m)_{m\geq 1}, (R_m)_{m \geq 1}$ are sequences tending to infinity as $m \to \infty$, satisfying*

$$0 < \gamma_L \leq \liminf_{m\to\infty} n_m f(R_m) \leq \limsup_{m \to \infty} n_m f(R_m) \leq \gamma_U < \infty.$$

Then we have for any increasing sequence $(v_k)_{k=0}^{\infty}$ of non-negative integers (in particular for $v_k = k, 2k,$ or $2k+1$) and $M \in \bar{\mathbb{N}}_0$,

$$(i) \qquad \liminf_{m\to\infty} R_m^{-d} \mathbb{E}\Big[\sum_{k=0}^{M} \tilde{S}_{v_k,n_m}(t)\Big] \geq \sum_{k=0}^{M} \frac{s_{d-1}\gamma_L^{v_k+1}}{(v_k+1)!(\alpha(v_k+1)-d)} \int_{(\mathbb{R}^d)^{v_k}} h_t^{v_k}(0,\mathbf{y})\,\mathrm{d}\mathbf{y},$$

$$(ii) \qquad \limsup_{m\to\infty} R_m^{-d} \mathbb{E}\Big[\sum_{k=0}^{M} \tilde{S}_{v_k,n_m}(t)\Big] \leq \sum_{k=0}^{M} \frac{s_{d-1}\gamma_U^{v_k+1}}{(v_k+1)!(\alpha(v_k+1)-d)} \int_{(\mathbb{R}^d)^{v_k}} h_t^{v_k}(0,\mathbf{y})\,\mathrm{d}\mathbf{y},$$

$$(iii) \qquad \sup_m R_m^{-d} \mathrm{Var}\Big(\sum_{k=0}^{M} \tilde{S}_{v_k,n_m}(t)\Big) < \infty.$$

Proof. Let us suppose throughout the entirety of the proof that $M = \infty$ and consider only the case when $\Phi_n = \mathcal{X}_n$ is a binomial process for (*i*) and (*ii*). The only difference between the binomial and Poisson cases being the quantity $\binom{n_m}{v_k+1}$, which becomes $n_m^{v_k+1}/(v_k+1)!$

in the latter scenario—see Lemma 2.2.4(i). We begin by offering an abridged proof of (i) and (ii) (for an extended argument in a similar setup see Proposition 7.2 in [67]). It is clear to see that

$$\begin{aligned}\mathbb{E}\Big[\sum_{k=0}^{\infty}\tilde{S}_{v_k,n_m}(t)\Big] &= \sum_{k=0}^{n_m-1}\mathbb{E}[\tilde{S}_{v_k,n_m}(t)] \\ &= \sum_{k=0}^{n_m-1}\binom{n_m}{v_k+1}\mathbb{E}\Big[h_t^{v_k}(\mathcal{X}_{v_k+1})1\Big\{\min_{1\le i\le v_k+1}\|X_i\|>R_m\Big\}\Big]. \end{aligned}\tag{5.10}$$

From this we have

$$\mathbb{E}\Big[h_t^{v_k}(\mathcal{X}_{v_k+1})1\Big\{\min_{1\le i\le v_k+1}\|X_i\|>R_m\Big\}\Big] \tag{5.11}$$

$$= \int_{(\mathbb{R}^d)^{v_k+1}} h_t^{v_k}(x_0,\dots,x_{v_k})\prod_{i=0}^{v_k}f(x_i)1\Big\{\|x_i\|>R_m\Big\}\,\mathrm{d}\mathbf{x} \tag{5.12}$$

$$= \int_{(\mathbb{R}^d)^{v_k+1}} h_t^{v_k}(0,y_1\dots,y_{v_k})f(x)1\Big\{\|x\|>R_m\Big\}\prod_{i=1}^{v_k}f(x+y_i)1\Big\{\|x+y_i\|>R_m\Big\}\,\mathrm{d}x\,\mathrm{d}\mathbf{y},$$

by the changes of variable $x_0 \mapsto x$ and $x_i \mapsto x+y_i$, $i \ge 1$ and translation invariance. Furthermore, let us make the change of variables $x \mapsto r\theta$ and $r \mapsto R_m\rho$ to get

$$\begin{aligned} R_m^d f(R_m)^{v_k+1}&\int_{(\mathbb{R}^d)^{v_k}}\int_{S^{d-1}}\int_1^{\infty} h_t^{v_k}(0,\mathbf{y})\frac{f(R_m\rho)}{f(R_m)}\rho^{d-1} \\ &\times\prod_{i=1}^{v_k}\frac{f(R_m\|\rho\theta+y_i/R_m\|)}{f(R_m)}1\Big\{\|\rho\theta+y_i/R_m\|>1\Big\}\,\mathrm{d}\rho\,\mathrm{d}\nu_{d-1}(\theta)\,\mathrm{d}\mathbf{y}, \end{aligned}\tag{5.13}$$

Where $\mathbf{y}=(y_1,\dots,y_{v_k})\in(\mathbb{R}^d)^{v_k}$ as usual. Note that Lemma 5.1.1 shows that for a positive $\epsilon<(\alpha-d)\wedge 1$, there exists an M such that for $m \ge M$ we have

$$\frac{f(R_m\rho)}{f(R_m)} \le 2\rho^{-\alpha+\epsilon} \tag{5.14}$$

for all $\rho>1$, which is the condition on ρ in the integral (5.13). By this same argument, we have

$$\prod_{i=1}^{v_k}\frac{f(R_m\|\rho\theta+y_i/R_m\|)}{f(R_m)}1\Big\{\|\rho\theta+y_i/R_m\|>1\Big\} \le 2^{v_k}, \tag{5.15}$$

for $m \geq M$. Putting these bounds together, we get that (5.13) is bounded for large m by

$$C^* \int_{(\mathbb{R}^d)^{v_k}} \int_1^\infty h_t^{v_k}(0, \mathbf{y}) \rho^{d-1-\alpha+\epsilon} \, \mathrm{d}\rho \, \mathrm{d}\mathbf{y} < \infty,$$

which is integrable by $d - 1 - \alpha + \epsilon < 1$ and the standard integral bound for h_t^k, as in Lemma 4.2.1(ii). Hence, the dominated convergence theorem applies and the regular variation of f ensures that the triple integral in (5.13) converges to

$$s_{d-1} \int_{(\mathbb{R}^d)^{v_k}} \int_1^\infty h_t^{v_k}(0, \mathbf{y}) \rho^{d-1-\alpha(v_k+1)} \, \mathrm{d}\rho \, \mathrm{d}\mathbf{y} = \frac{s_{d-1}}{\alpha(v_k+1) - d} \int_{(\mathbb{R}^d)^{v_k}} h_t^{v_k}(0, \mathbf{y}) \, \mathrm{d}\mathbf{y}.$$

Elementary properties of limit infimum furnish that

$$\begin{aligned} \liminf_{n \to \infty} \binom{n_m}{v_k + 1} \mathbb{E}\Big[h_t^{v_k}(\mathcal{X}_{v_k+1}) 1\Big\{ \min_{1 \leq i \leq v_k+1} \|X_i\| > R_m \Big\} \Big] & \\ \geq \frac{s_{d-1} \gamma_L^{v_k+1}}{(v_k+1)!(\alpha(v_k+1) - d)} \int_{(\mathbb{R}^d)^{v_k}} h_t^{v_k}(0, \mathbf{y}) \, \mathrm{d}\mathbf{y}, \end{aligned}$$

where similar properties for lim sup demonstrate that the summand in (ii) is an upper bound for this expectation. Fatou's lemma provides us the desired lower bound for (i) and (ii) is established via an additional application of Potter's bound to demonstrate the dominated convergence assumption holds.

Now we will prove (iii), for $\Phi_n = \mathcal{X}_n$ or $\mathcal{P}_n$. For a point cloud $\mathcal{X} \in \mathcal{F}(\mathbb{R}^d)$ and S_{v_k} defined with respect to *any* spherically symmetric f let

$$\mathfrak{S}_{t,m}(\mathcal{X}) := \sum_{k=0}^\infty S_{v_k}(\mathcal{K}(\mathcal{X} \cap B(0, R_m)^c, t)).$$

We will give bounds on $D_x \mathfrak{S}_{t,m}(\eta)$ for a simple $\mathbf{N}$-valued point process. Abusing notation, we will take η to be not only a point process but also identify it with $\mathrm{supp}(\eta)$. First we have

$$D_x \mathfrak{S}_{t,m}(\eta) = \sum_{k=0}^\infty S_{v_k}\Big(\mathcal{K}\Big((\eta \cup \{x\}) \cap B(0, R_m)^c, t\Big)\Big) - S_{v_k}\Big(\mathcal{K}\Big(\eta \cap B(0, R_m)^c, t\Big)\Big),$$

as for either point process this sum is almost surely finite. For the left hand term in the above sum to contain any new simplices, it must be the case that $\|x\| > R_m$ and that each new simplex contains x. Hence,

$$\begin{aligned} D_x \mathfrak{S}_{t,m}(\eta) &= 1\Big\{\|x\| > R_m\Big\} \sum_{k=0}^{\infty} \sum_{\mathcal{Y} \subset \eta} h_t^{v_k}(\mathcal{Y} \cup \{x\}) 1\Big\{\min_{Y \in \mathcal{Y}} \|Y\| > R_m\Big\} \\ &= 1\Big\{\|x\| > R_m\Big\} \sum_{k=0}^{\infty} \binom{\eta\Big(B(x,t) \cap B(0,R_m)^c\Big)}{v_k} \\ &\leq 1\Big\{\|x\| > R_m\Big\} 2^{\eta(B(x,t) \cap B(0,R_m)^c)} \end{aligned}$$

as $\eta(B(x,t) \cap B(0,R_m)^c) \geq v_k$. Now suppose that X is a (potentially constant) random variable independent of η. Then,

$$\mathbb{E}[|D_X \mathfrak{S}_{t,m}(\eta)|^2] \leq \mathbb{E}\Big[1\Big\{\|X\| > R_m\Big\} \mathbb{E}[4^{\eta(B(X,t) \cap B(0,R_m)^c)} | X]\Big],$$

so that in the case $\eta = \Phi_{n_m}$ we have

$$\begin{aligned} \mathbb{E}[|D_X \mathfrak{S}_{t,m}(\eta)|^2] &\leq \mathbb{E}\Big[1\Big\{\|X\| > R_m\Big\} \mathbb{E}[4^{\Phi_{n_m}(B(X,t) \cap B(0,R_m)^c)} | X]\Big] \\ &\leq \mathbb{E}\Big[1\Big\{\|X\| > R_m\Big\} \exp\Big\{3 n_m \int_{B(X,t)} f(y) 1\Big\{\|y\| > R_m\Big\} \, \mathrm{d}y\Big\}\Big], \end{aligned} \tag{5.16}$$

which follows by the fact that for $X \sim \mathrm{Bin}(n,p)$, $Y \sim \mathrm{Poi}(np)$, and $z \in \mathbb{R}$ that $\mathbb{E}[z^X] = (1 - p + zp)^n \leq e^{(z-1)np}$ and $\mathbb{E}[z^Y] = e^{(z-1)np}$. Fix an $0 < \epsilon < \alpha$. By spherical symmetry and Lemma 5.1.1 there exists some M (not depending on x) such that for $m \geq M$ and any scalar $x \in \mathbb{R}^d$ we have

$$\begin{aligned} n_m \int_{B(x,t)} f(y) 1\Big\{\|y\| > R_m\Big\} \, \mathrm{d}y &\leq n_m f(R_m) \int_{B(x,t)} \frac{f(R_m(\|y\|/R_m))}{f(R_m)} 1\Big\{\|y\|/R_m > 1\Big\} \, \mathrm{d}y \\ &\leq 2\gamma_U \int_{B(x,t)} \left(\frac{\|y\|}{R_m}\right)^{-\alpha+\epsilon} 1\Big\{\|y\|/R_m > 1\Big\} \, \mathrm{d}y \\ &\leq 2\gamma_U \omega_d t^d < \infty. \end{aligned}$$

Thus, if we let $K = \sup_{x \in \mathbb{R}^d} \sup_m \exp\Big\{3n_m \int_{B(x,t)} f(y) 1\Big\{\|y\| > R_m\Big\}\, \mathrm{d}y\Big\}$, we have shown not only that $K < \infty$, but that

$$\mathbb{E}[|D_X \mathfrak{S}_{t,m}(\Phi_{n_m})|^2] \le K P(\|X\| > R_m). \tag{5.17}$$

Applying Lemma 2.2.2, we have for $\Phi_n = \mathcal{P}_n$ that

$$R_m^{-d} \mathrm{Var}\Big(\sum_{k=0}^{\infty} \tilde{S}_{v_k, n_m}(t) \Big) \le K n_m R_m^{-d} P(\|X\| > R_m).$$

Now consider $\Phi_n = \mathcal{X}_n$. By Lemma 2.2.1, we see that

$$R_m^{-d} \mathrm{Var}\Big(\sum_{k=0}^{\infty} \tilde{S}_{v_k, n_m}(t) \Big) \le 2 R_m^{-d} n_m \mathbb{E}\Big[|D_{X_{n_m}} \mathfrak{S}_{t,m}(\mathcal{X}_{n_m} \setminus \{X_{n_m}\})^2 \Big]$$

and we may apply the bound in (5.17) owing to the fact that X_{n_m} is independent of the point cloud $\mathcal{X}_{n_m} \setminus \{X_{n_m}\}$. Hence,

$$R_m^{-d} \mathrm{Var}\Big(\sum_{k=0}^{\infty} \tilde{S}_{v_k, n_m}(t) \Big) \le 2K n_m R_m^{-d} P(\|X\| > R_m).$$

It remains to show that $n_m R_m^{-d} P(\|X\| > R_m)$ is bounded. Indeed, for $m \ge M$ we make our usual change to polar coordinates $x \mapsto r\theta$ and then $r \mapsto R_m \rho$ to get

$$n_m R_m^{-d} P(\|X\| > R_m) = n_m f(R_m) s_{d-1} \int_1^{\infty} \rho^{d-1} \frac{f(R_m \rho)}{f(R_m)}\, \mathrm{d}\rho,$$

which is bounded as $\limsup_{m \to \infty} n_m f(R_m) \le \gamma_U$ and an application of Potter's bound as above. $\square$

Proof of Proposition 5.2.1. Note that the proof here differs slightly from that of [93] handle the case where R_n varies regularly. The argument here is inspired by that of [35].

Let $j_m = \lfloor \gamma^m \rfloor$, for $m \in \mathbb{N}_0$ and $\gamma > 1$. As a result, for every $n \in \mathbb{N}$, there exists a unique $m = m(n)$ such that $j_m \le n < j_{m+1}$. Let us define

$$p_m := \arg\max\{j_m \le \ell \le j_{m+1} : R_\ell\} \tag{5.18}$$

$$q_m := \arg\min\{j_m \le \ell \le j_{m+1} : R_\ell\}. \tag{5.19}$$

It then holds that $R_{p_m} = \max_{j_m \le \ell \le j_{m+1}} R_\ell$ and $R_{q_m} = \min_{j_m \le \ell \le j_{m+1}} R_\ell$. As a result, we get

$$\frac{T_k^m(t)}{R_{p_m}^d} \le \frac{\tilde{S}_{k,n}(t)}{R_n^d} \le \frac{U_k^m(t)}{R_{q_m}^d},$$

with

$$U_k^m(t) := \sum_{\mathcal{Y} \subset \mathcal{X}_{j_{m+1}}} h_t^k(\mathcal{Y}) 1\Big\{ \min_{y \in \mathcal{Y}} \|y\| > R_{q_m} \Big\}, \tag{5.20}$$

$$T_k^m(t) := \sum_{\mathcal{Y} \subset \mathcal{X}_{j_m}} h_t^k(\mathcal{Y}) 1\Big\{ \min_{y \in \mathcal{Y}} \|y\| > R_{p_m} \Big\}. \tag{5.21}$$

Hence, we have that

$$\sum_{k=0}^\infty \frac{T_{2k+i}^m(t)}{R_{p_m}^d} \le \frac{\chi_n^{(i)}(t)}{R_n^d} \le \sum_{k=0}^\infty \frac{U_{2k+i}^m(t)}{R_{q_m}^d}, \quad i = 0, 1. \tag{5.22}$$

As the proof is the same regardless of whatever value i takes, let us take $i = 0$ and drop i from $\chi_n^{(i)}(t)$. Let us define $T_m(t) := \sum_{k=0}^\infty T_{2k}^m(t)$ and $U_m(t) := \sum_{k=0}^\infty U_{2k}^m(t)$. Then, we see that (5.22) implies that

$$\liminf_{n \to \infty} \frac{T_m(t) - \mathbb{E}[T_m(t)]}{R_{p_m}^d} + \liminf_{n \to \infty} \frac{\mathbb{E}[T_m(t)]}{R_{p_m}^d} \le \liminf_{n \to \infty} \frac{\chi_n(t)}{R_n^d}, \tag{5.23}$$

and

$$\limsup_{n \to \infty} \frac{\chi_n(t)}{R_n^d} \le \limsup_{n \to \infty} \frac{\mathbb{E}[U_m(t)]}{R_{q_m}^d} + \limsup_{n \to \infty} \frac{U_m(t) - \mathbb{E}[U_m(t)]}{R_{q_m}^d}. \tag{5.24}$$

As $\liminf_{n \to \infty} j_m f(R_{p_m}) \ge \xi/\gamma$ and $\limsup_{n \to \infty} j_{m+1} f(R_{q_m}) \le \xi\gamma$, applying Lemma 5.2.3($i$) and ($ii$), with $v_k = 2k$, implies that

$$\liminf_{n \to \infty} \frac{\mathbb{E}[T_m(t)]}{R_{p_m}^d} \ge \sum_{k=0}^\infty \gamma^{-(2k+1)} s_{2k}(t), \quad n \to \infty,$$

and

$$\limsup_{n \to \infty} \frac{\mathbb{E}[U_m(t)]}{R_{q_m}^d} \le \sum_{k=0}^\infty \gamma^{2k+1} s_{2k}(t), \quad n \to \infty,$$

respectively. Let us continue by showing that both $R_{p_m}^{-d}\Big(T_m(t) - \mathbb{E}[T_m(t)]\Big) \to 0$ a.s., $n \to \infty$, as well as $R_{q_m}^{-d}\Big(U_m(t) - \mathbb{E}[U_m(t)]\Big) \to 0$ a.s., $n \to \infty$. To begin, we shall demonstrate that there exists an N such that for any $m \geq N$ and $n \geq j_m$,

$$R_n^d \geq C\Big(\gamma^{\frac{d}{\alpha+\delta}}\Big)^{m-1}, \tag{5.25}$$

where C a positive constant depending only on f and ξ and where $\delta > 0$. As $f \in RV_{-\alpha}$, we can represent $f(r) = r^{-\alpha}L(r)$, where L is a slowly varying function. Therefore, we have $R_n^\alpha = L(R_n)/f(R_n)$. For any $\delta > 0$ we have that $x^\delta L(x) \in RV_\delta$, hence $\lim_{x\to\infty} x^\delta L(x) = \infty$ by Proposition 2.6(i) in [82]. Hence, for $b > 0$ there exists an N_0 such that if $m \geq N_0$ then $R_n^\delta L(R_n) \geq b$ for $n \geq j_m$. Therefore, for such an m we have that

$$R_n^{\alpha+\delta} \geq \frac{nb}{nf(R_n)} \quad \Leftrightarrow \quad R_n^d \geq \left(\frac{nb}{nf(R_n)}\right)^{\frac{d}{\alpha+\delta}}, \quad n \geq j_m.$$

As $nf(R_n) \to \xi$, there exists an N_1 where $m \geq N_1 \geq N_0$ implies that $nf(R_n) \leq 2\xi$ for $n \geq j_m$. Finally, there exists a larger $N \geq N_1$ such that $j_m > \gamma^{m-1}$, for all $m \geq N$, which happens just when $\lfloor \gamma^m \rfloor / \gamma^m \geq \gamma^{-1}$. Therefore, (5.25) follows as a result.

We conclude by using the Borel-Cantelli lemma to show that $R_{q_m}^{-d}\Big(U_m(t) - \mathbb{E}[U_m(t)]\Big) \to 0$ a.s., $n \to \infty$ (the situation with $R_{p_m}^{-d}\Big(T_m(t) - \mathbb{E}[T_m(t)]\Big)$ is analogous, so the proof is omitted). As usual, take C^* to be a positive constant that we allow to vary between lines. Applying the Borel-Cantelli lemma, we see that

$$\begin{aligned}\sum_{m=1}^{\infty} \mathbb{P}(|U_m(t) - \mathbb{E}[U_m(t)]| \geq \epsilon R_{q_m}^d) \\ \leq \sum_{m=1}^{\infty} \frac{\mathrm{Var}(U_m(t))}{\epsilon^2 (R_{q_m}^d)^2} \leq C^* \sum_{m=1}^{\infty} \frac{1}{R_{q_m}^d}\end{aligned}$$

because of Lemma 5.2.3(iii). Now, by (5.25) we have for $m \geq N$,

$$C^* \sum_{m=N+1}^{\infty} \frac{1}{R_{q_m}^d} \leq \sum_{m=N+1}^{\infty} \Big(\gamma^{-\frac{d}{\alpha+\delta}}\Big)^{m-1} < \infty.$$

Therefore, $\lim_{m\to\infty} R_{q_m}^{-d}\Big|U_m(t) - \mathbb{E}[U_m(t)]\Big| = 0$, a.s. and hence for $n \to \infty$ as well. Combining all the above results, we see that,

$$\sum_{k=0}^{\infty} \gamma^{-(2k+1)} s_{2k}(t) \leq \liminf_{n\to\infty} \frac{\chi_n(t)}{R_n^d} \leq \limsup_{n\to\infty} \frac{\chi_n(t)}{R_n^d} \leq \sum_{k=0}^{\infty} \gamma^{2k+1} s_{2k}(t) \quad \text{a.s.}$$

for any $\gamma > 1$. Noting that we may assume γ to be less than any $M > 1$ without loss of generality, we take $\gamma \downarrow 1$ and apply the dominated convergence theorem to the above. This yields that for fixed $t \geq 0$ we have

$$\frac{\chi_n(t)}{R_n^d} \to \sum_{k=0}^{\infty} s_{2k}(t), \quad \text{a.s.}.$$

The situation with $i = 1$ differs only in that we take $v_k = 2k + 1$ in applying Lemma 5.2.3. For each $\tilde{S}_{k,n}(t)$, $k \in \mathbb{N}_0$, similar, simpler proofs hold. □

Having proved Lemma 5.2.3 and Proposition 5.2.1 we can finally prove Theorem 5.2.1, and do so in very concise manner.

Proof of Theorem 5.2.1. We apply Proposition 2.3.1 to (5.9) in Proposition 5.2.1 to get

$$\left(\frac{\chi_n^{(i)}(t)}{R_n^d},\, t \geq 0\right) \to \left(\sum_{k=0}^{\infty} s_{2k+i}(t),\, t \geq 0\right), \quad \text{a.s. in } (D, U), \tag{5.26}$$

for $i = 0, 1$. Furthermore, we know that

$$(\tilde{\chi}_n(t),\, t \geq 0) = \left(\chi_n^{(0)}(t) - \chi_n^{(1)}(t),\, t \geq 0\right), \quad \text{a.s.},$$

so that for each $T \geq 0$ we have

$$\sup_{0\leq t\leq T}\left|\tilde{\chi}_n(t) - \sum_{k=0}^{\infty}(-1)^k s_k(t)\right|$$
$$\sup_{0\leq t\leq T}\left|\chi_n^{(0)}(t) - \sum_{k=0}^{\infty} s_{2k}(t)\right| + \sup_{0\leq t\leq T}\left|\chi_n^{(1)}(t) - \sum_{k=0}^{\infty} s_{2k+1}(t)\right| \to 0,$$

as $n \to \infty$ by (5.26). Hence, the theorem is proved.

Finally, let us explicitly demonstrate that the limit in (5.6) is finite for all $t \geq 0$. By virtue of property 2 in Definition 2.1.5,

$$\Big|\sum_{k=0}^{\infty}(-1)^k s_k(t)\Big| \leq \sum_{k=0}^{\infty} \frac{s_{d-1}\xi^{k+1}}{(k+1)!\Big(\alpha(k+1)-d\Big)} \int_{(\mathbb{R}^d)^k} \prod_{i=1}^{k} \mathbb{1}\Big\{\|y_i\| \leq ct\Big\}\, \mathrm{d}\mathbf{y}$$

$$\leq C^* \sum_{k=0}^{\infty} \frac{\Big((ct)^d \xi \omega_d\Big)^k}{k!} = C^* e^{(ct)^d \xi \omega_d} < \infty.$$

□

5.3 Exponentially-decaying tail case

Here, we consider the limiting behavior of (5.1) for a density with faster tail decay than the previous section. We suppose that our density f is specified by

$$f(x) = C \exp\Big\{-\psi(\|x\|)\Big\},\ x \in \mathbb{R}^d, \tag{5.27}$$

where C is a normalizing constant and $\psi : [0,\infty) \to [0,\infty)$ is regularly varying with exponent $\tau > 0$. Moreover, we assume that ψ is twice differentiable, that $\psi'(x) > 0$ for all $x > 0$, and that ψ' is eventually monotone. Eventually monotone means that there exists $z_0 > 0$ such that ψ' is monotone (nonincreasing or nondecreasing) in $[z_0, \infty)$—if $\tau \leq 1$ we assume ψ' is eventually nonincreasing. The function ψ is a special case of what is called a *von Mises function*, which characterizes the max-domain of attraction for the Gumbel distribution $F(x) = \exp\{-e^{-x}\}$ in the univariate setup.

Let us define the function $a(z) := 1/\psi'(z)$. As a consequence of the conditions of ψ, we have that a is also regularly varying with index $1-\tau$ (see, e.g., Proposition 2.5 in [82]). It follows that $a(z)/z \to 0$ as $z \to \infty$. It is evident that all of the above assumptions imply that if $\tau < 1$ then $a(z) \to \infty$ as $z \to \infty$. Additionally, we have for any $t_n \to \infty$ and $z > 0$ that

$$\lim_{n\to\infty} \frac{a(t_n + a(t_n)z)}{a(t_n)} = 1. \tag{5.28}$$

This can be seen from the representation of $t_n+a(t_n)z = t_n(1+za(t_n)/t_n)$, that $a(t_n)/t_n \to 0$ as $n \to \infty$, and the fact that regularly varying sequences converge uniformly (see Theorem 1.1 in [89] for a general proof).

Here, it is important to note that the occurrence of topological crackle depends on the limiting value ζ of $a(z)$ as $z \to \infty$, as demonstrated in [70]. In particular, [70] showed that crackle occurs if and only if

$$\zeta := \lim_{z\to\infty} a(z) \in (0,\infty]. \tag{5.29}$$

The case of $\zeta = 0$ is dealt with in Corollary 5.3.5. Interestingly, if $\zeta = \infty$ in (5.29), then (5.31) agrees with (5.7) up to multiplicative constants. We can now state the main result for this section.

Theorem 5.3.1. *Let $\mathcal{K}$ be any any right-continuous simplicial construction function and take Φ_n to be either Poisson or binomial. Suppose that f is a probability density satisfying (5.27) with $\tau \in (0,1]$ for $d \geq 3$ or $\tau \in (0,1)$ for $d = 2$. Then if $nf(R_n) \to \xi \in (0,\infty)$ we have the following functional strong law of large numbers, i.e.*

$$\left(\frac{\tilde{\chi}_n(t)}{a(R_n)R_n^{d-1}},\, t \geq 0\right) \to \left(\sum_{k=0}^{\infty}(-1)^k s_k(t),\, t\geq 0\right), \quad \text{a.s. in } (D,U), \tag{5.30}$$

where

$$\begin{aligned} s_k(t) := \frac{\xi^{k+1}}{(k+1)!}\int_0^\infty \int_{S^{d-1}} \int_{(\mathbb{R}^d)^k} & h_t^k(0,y_1,\dots,y_k)\, e^{-(k+1)\rho-\zeta^{-1}\sum_{i=1}^k \langle\theta,y_i\rangle} \\ & \times \prod_{i=1}^k \mathbb{1}\Big\{\rho+\zeta^{-1}\langle\theta,y_i\rangle > 0\Big\}\, \mathrm{d}\mathbf{y}\, \mathrm{d}\nu_{d-1}(\theta)\, \mathrm{d}\rho, \quad t\geq 0, \quad k \geq 1, \end{aligned} \tag{5.31}$$

with $\langle\cdot,\cdot\rangle$ being the Euclidean inner product and $s_0(t) \equiv s_{d-1}\xi$. In particular, the limit in (5.30) is convergent for all $t \geq 0$.

Remark 5.3.2. It is the author's view that a somewhat tedious argument—resorting to bounds on higher central moments of $\tilde{S}_{v_k,n_m}$ in analogy with Lemma 5.3.4(*iii*)—would establish Theorem 5.3.1 for $d = 2$, $\tau = 1$ when $\zeta > 0$.

Example 5.3.3. We consider a special case of the density in (5.27),

$$f(x) = Ce^{-\|x\|^\tau/\tau}, \quad x \in \mathbb{R}^d, \quad \tau \in (0,1].$$

Define $R_n = \big(\tau \log n + \tau \log C\big)^{1/\tau}$ so that $nf(R_n) = 1$. Then, $a(z) = z^{1-\tau}$, $z > 0$. According to Theorem 5.3.1,

$$\left(\frac{\tilde{\chi}_n(t)}{(\tau \log n)^{\frac{d-\tau}{\tau}}}, \, t \geq 0 \right) \to \left(\sum_{k=0}^\infty (-1)^k s_k(t), \, t \geq 0 \right), \quad \text{a.s. in } D[0,\infty).$$

where $s_k(t)$ is defined in (5.31). Moreover, applying the continuous functions $U_{a,b}$ and I from Example 5.2.2, we have

$$\frac{\sup_{a \leq t \leq b} \big|\chi_n(t)\big|}{(\tau \log n)^{\frac{d-\tau}{\tau}}} \to \sup_{a \leq t \leq b} \left| \sum_{k=0}^\infty (-1)^k s_k(t) \right| \quad \text{a.s.}$$

and

$$\left(\frac{\int_0^t \chi_n(r) \, \mathrm{d}r}{(\tau \log n)^{\frac{d-\tau}{\tau}}}, \ t \geq 0 \right) \to \left(\sum_{k=0}^\infty (-1)^k \int_0^t s_k(r) \, \mathrm{d}r, \ t \geq 0 \right), \quad \text{a.s. in } D[0,\infty).$$

In the case of an exponentially-decaying tail, we would like a result analogous to Proposition 5.2.1. This result is contained in Proposition 5.3.1.

Proposition 5.3.1. *Suppose that f is a probability density satisfying (5.27) and $\tau \in (0,1]$ for $d \geq 3$ or $\tau \in (0,1)$ for $d = 2$, with $nf(R_n) \to \xi \in (0,\infty)$. Then for every $t \geq 0$,*

$$\frac{\tilde{S}_{k,n}(t)}{R_n^d} \to \tilde{S}_k(t), \quad \text{a.s.,} \quad n \to \infty,$$

where $s_k(t)$ is defined at (5.31). Additionally, we have

$$\frac{\chi_n^{(i)}(t)}{R_n^d} \to \sum_{k=0}^\infty s_{2k+i}(t), \quad \text{a.s.,} \quad n \to \infty, \tag{5.32}$$

for $i = 0, 1$—where $\chi_n^{(i)}(t) := \sum_{k=0}^\infty \tilde{S}_{2k+i,n}(t)$.

The following lemma is the analogue in the exponentially-decaying tail case of Lemma 5.2.3.

Lemma 5.3.4. *Let $\mathcal{K}$ be any simplicial construction function, let Φ_n be Poisson or binomial, and suppose that f is spherically symmetric and suppose that f has an exponentially-decaying tail as in (5.27). Let us assume that $(n_m)_{m\geq 1}, (R_m)_{m\geq 1}, (Q_m)_{m\geq 1}$ are sequences tending to infinity as $m \to \infty$, satisfying*

$$0 < \gamma_L \leq \liminf_{n\to\infty} n_m f(R_m) \leq \limsup_{n\to\infty} n_m f(R_m) \leq \gamma_U < \infty,$$

and $R_m \sim Q_m$. Also, suppose that $a(z) \to \zeta \in (0, \infty]$ as $z \to \infty$. Then we have for any increasing sequence $(v_k)_{k=0}^{\infty}$ of non-negative integers (in particular for $v_k = k, 2k,$ or $2k+1$) and $M \in \bar{\mathbb{N}}_0$,

$$(i) \qquad \liminf_{m\to\infty} [a(Q_m)Q_m^{d-1}]^{-1}\mathbb{E}\Big[\sum_{k=0}^{M} \tilde{S}_{v_k,n_m}(t)\Big] \geq \sum_{k=0}^{M} \frac{\gamma_L^{v_k+1}}{(v_k+1)!}\mathcal{I}_{v_k}(t),$$

$$(ii) \qquad \limsup_{m\to\infty} [a(Q_m)Q_m^{d-1}]^{-1}\mathbb{E}\Big[\sum_{k=0}^{M} \tilde{S}_{v_k,n_m}(t)\Big] \leq \sum_{k=0}^{M} \frac{\gamma_U^{v_k+1}}{(v_k+1)!}\mathcal{I}_{v_k}(t),$$

$$(iii) \qquad \sup_m [a(Q_m)Q_m^{d-1}]^{-1}\mathrm{Var}\Big(\sum_{k=0}^{M} \tilde{S}_{v_k,n_m}(t)\Big) < \infty,$$

where $\mathcal{I}_k(t)$ is equal to the integral term in $s_k(t)$ defined at (5.31)

Proof. Much of this proof adheres closely to the argument of Proposition 7.4 in [67]. As in Lemma 5.2.3, we assume that $M = \infty$ and begin by proving (i) and (ii). Similarly, we also only consider $\Phi_n = \mathcal{X}_n$ here, as in Lemma 5.2.3. It is clear that

$$\mathbb{E}\Big[\sum_{k=0}^{\infty} \tilde{S}_{v_k,n_m}(t)\Big] = \sum_{k=0}^{n_m-1} \binom{n_m}{v_k+1} \int_{(\mathbb{R}^d)^{v_k+1}} h_t^{v_k}(0, y_1 \dots, y_{v_k}) f(x) 1\Big\{\|x\| > R_m\Big\} \times \prod_{i=1}^{v_k} f(x+y_i) 1\Big\{\|x+y_i\| > R_m\Big\}\, \mathrm{d}x\, \mathrm{d}\mathbf{y},$$

from the same as the manipulations from (5.10) and (5.12), and because of the translation invariance of $\mathcal{K}$. From here we make the following changes of variable: first, $x \mapsto r\theta$ and then $r \mapsto R_m + a(R_m)\rho$. Hence the integral above becomes

$$a(R_m)R_m^{d-1} \int_{(\mathbb{R}^d)^{v_k}} \int_{S^{d-1}} \int_0^{\infty} h_t^{v_k}(0, \mathbf{y}) f(R_m + a(R_m)\rho) \Big(\frac{a(R_m)\rho}{R_m} + 1\Big)^{d-1}$$

$$\times \prod_{i=1}^{v_k} f\Big(\|(R_m + a(R_m)\rho)\theta + y_i\|\Big) 1\Big\{\|(R_m + a(R_m)\rho)\theta + y_i\| > R_m\Big\} \,\mathrm{d}\rho \,\mathrm{d}\nu_{d-1}(\theta) \,\mathrm{d}\mathbf{y},$$

which is equal to

$$a(R_m)R_m^{d-1} f(R_m)^{v_k+1} \int_{(\mathbb{R}^d)^{v_k}} \int_{S^{d-1}} \int_0^\infty h_t^{v_k}(0,\mathbf{y}) \frac{f(R_m + a(R_m)\rho)}{f(R_m)} \left(\frac{a(R_m)\rho}{R_m} + 1\right)^{d-1} \tag{5.33}$$
$$\times \prod_{i=1}^{v_k} \frac{f\Big(\|(R_m + a(R_m)\rho)\theta + y_i\|\Big)}{f(R_m)} 1\Big\{\|(R_m + a(R_m)\rho)\theta + y_i\| > R_m\Big\} \,\mathrm{d}\rho \,\mathrm{d}\nu_{d-1}(\theta) \,\mathrm{d}\mathbf{y}.$$

Let us first demonstrate the appropriate upper bounds and limits for each term in the integral (5.33) and $\mathbb{E}[\tilde{S}_{v_k,n_m}(t)]$ so that we may identify the limit and show that the dominated convergence theorem can be applied twice, as required. Clearly, $h_t^{v_k}(0,\mathbf{y}) \le \prod_{i=1}^{v_k} 1\Big\{y_i \in B(0,ct)\Big\}$. Furthermore,

$$\left(1 + \frac{a(R_m)\rho}{R_m}\right)^{d-1} \to 1, \quad m \to \infty,$$

by virtue of $a(z)/z \to 0$ as $z \to \infty$, as we demonstrated in the setup. Also, for this term we have

$$\left(1 + \frac{a(R_m)\rho}{R_m}\right)^{d-1} \le 2(\rho \vee 1)^{d-1}, \tag{5.34}$$

for m sufficiently large. We continue with $f(a(R_m)\rho + R_m)/f(R_m)$. For this quantity,

$$\begin{aligned} \frac{f(R_m + a(R_m)\rho)}{f(R_m)} &= \exp\Big\{ -\psi(R_m + a(R_m)\rho) + \psi(R_m)\Big\} \\ &= \exp\left(-\int_0^\rho \frac{a(R_m)}{a\Big(R_m + a(R_m)z\Big)} \,\mathrm{d}z \right) \to e^{-\rho}, \quad n \to \infty, \end{aligned} \tag{5.35}$$

by the property that as $m \to \infty$,

$$\frac{a(R_m)}{a(R_m + a(R_m)z)} \to 1, \tag{5.36}$$

as mentioned at (5.28). We additionally define a sequence $(s_\ell(m), \ell \ge 0, m \ge 1)$ by

$$\psi(R_m + a(R_m)s_\ell(m)) = \psi(R_m) + \ell,$$

which exists by ψ increasing, $\psi(z) \to \infty$, $z \to \infty$ and the Intermediate Value Theorem, for example. Then, Lemma 5.2 in [7] implies that for any $\epsilon \in (0, d^{-1})$ there exists a positive integer $N = N(\epsilon)$ such that

$$s_\ell(m) \leq e^{\ell\epsilon}/\epsilon,$$

for $m \geq N$, $\ell \geq 0$. As a result of this and the fact that ψ is increasing, we can establish a bound for (5.35), for $m \geq N$ as follows:

$$\begin{aligned}
&\exp\Big\{-\psi(R_m + a(R_m)\rho) + \psi(R_m)\Big\} 1\Big\{\rho > 0\Big\} \\
&\quad = \sum_{\ell=0}^{\infty} 1\Big\{s_\ell(m) < \rho \leq s_{\ell+1}(m)\Big\} \exp\Big\{-\psi(R_m + a(R_m)\rho) + \psi(R_m)\Big\} \\
&\quad \leq \sum_{\ell=0}^{\infty} 1\Big\{0 < \rho \leq e^{(\ell+1)\epsilon}/\epsilon\Big\} e^{-\ell},
\end{aligned} \tag{5.37}$$

We now discuss the final untreated term from the integral (5.33). Let us give a helpful fact about $\|a_m\theta + y_i\|$, where $a_m = R_m + a(R_m)\rho$. We have that

$$\|a_m\theta + y_i\| - \Big(a_m + \langle\theta, y_i\rangle\Big) = \frac{\|y_i\|^2 - \langle\theta, y_i\rangle^2}{\|a_m\theta + y_i\| + a_n + \langle\theta, y_i\rangle} =: \gamma_m(\rho, \theta, y_i),$$

and when $h_t^{v_k}(0, y) = 1$ and $\|(R_m + a(R_m)\rho)\theta + y_i\| > R_m$, the Cauchy-Schwarz inequality gives us

$$|\gamma_m(\rho, \theta, y_i)| \leq \frac{c^2t^2}{R_m - ct} \to 0, \quad n \to \infty, \tag{5.38}$$

the convergence of which is clearly uniform in ρ, θ and y_i, and c is as in Condition 2 of Definition 2.1.5. Letting $A_m = \Big\{y \in \mathbb{R}^d : \|(R_m + a(R_m)\rho)\theta + y\| > R_m\Big\}$, we have that for each $i = 1, \ldots, v_k$,

$$\begin{aligned}
&\frac{f\Big(\|(R_m + a(R_m)\rho)\theta + y_i\|\Big)}{f(R_m)} 1_{A_m}(y_i) \\
&\qquad = \exp\Big\{-\psi\Big(R_m + a(R_m)\rho + \langle\theta, y_i\rangle + \gamma_m(\rho, \theta, y_i)\Big) + \psi(R_m)\Big\} 1_{A_m}(y_i) \\
&\qquad = \exp\left(-\int_0^{\rho + \xi_m(\rho,\theta,y_i)} \frac{a(R_m)}{a\Big(R_m + a(R_m)z\Big)}\,\mathrm{d}z\right) 1_{A_m}(y_i),
\end{aligned} \tag{5.39}$$

where $\xi_m(\rho,\theta,y_i) := a(R_m)^{-1}\Big(\langle\theta,y_i\rangle + \gamma_m(\rho,\theta,y_i)\Big)$. Note that the integral term in (5.39) is bounded by 1. This is due to the fact

$$\|(R_m + a(R_m)\rho)\theta + y_i\| > R_m \;\Leftrightarrow\; \rho + \xi_m(\rho,\theta,y_i) > 0.$$

Additionally, (5.29) and (5.38) yield that

$$\xi_m(\rho,\theta,y_i) \to \zeta^{-1}\langle\theta,y_i\rangle, \quad m\to\infty,$$

for every $\rho>0$, $\theta\in S^{d-1}$, and $y_i\in\mathbb{R}^d$. Hence, as $m\to\infty$,

$$\exp\left\{-\int_0^{\rho+\xi_m(\rho,\theta,y_i)} \frac{a(R_m)}{a\Big(R_m + a(R_m)r\Big)}\,\mathrm{d}r\right\} \to \exp\Big\{-\rho-\zeta^{-1}\langle\theta,y_i\rangle\Big\}, \tag{5.40}$$

and

$$1_{A_m}(y_i) \to 1\Big\{\rho+\zeta^{-1}\langle\theta,y_i\rangle > 0\Big\}.$$

Combining all the bounds derived thus far, the integral in (5.33) is bounded above by

$$\begin{aligned}
&2\int_0^\infty\int_{S^{d-1}}\int_{(\mathbb{R}^d)^{v_k}} h_t(0,\mathbf{y})\,(1\vee\rho)^{d-1}\sum_{\ell=0}^\infty 1\Big\{0<\rho\le\epsilon^{-1}e^{(\ell+1)\epsilon}\Big\}e^{-\ell}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\nu_{d-1}(\theta)\,\mathrm{d}\rho\\
&= C^*\int_0^\infty\sum_{\ell=0}^\infty 1\Big\{0<\rho\le\epsilon^{-1}e^{(\ell+1)\epsilon}\Big\}e^{-\ell}(1\vee\rho)^{d-1}\,\mathrm{d}\rho\\
&\le C^*\Big(\frac{e^\epsilon}{\epsilon}\Big)^d\sum_{\ell=0}^\infty e^{-(1-\epsilon d)\ell} < \infty,
\end{aligned} \tag{5.41}$$

as $\epsilon^{-1}e^{(\ell+1)\epsilon}\ge 1$ and $\epsilon d<1$. Now, by the dominated convergence theorem, we can see that the integral in (5.33) converges to

$$\int_0^\infty\int_{S^{d-1}}\int_{(\mathbb{R}^d)^{v_k}} h_t^{v_k}(0,\mathbf{y})\,e^{-(v_k+1)\rho-\zeta^{-1}\sum_{i=1}^{v_k}\langle\theta,y_i\rangle}\prod_{i=1}^{v_k}1\Big\{\rho+\zeta^{-1}\langle\theta,y_i\rangle>0\Big\}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\nu_{d-1}(\theta)\,\mathrm{d}\rho.$$

By the fact that regularly varying sequences converge uniformly, we have $a \in RV_{1-\tau}$ implies that $a(R_m) \sim a(Q_m)$. Finally, because $\liminf_{m\to\infty} \binom{n_m}{v_k+1} f(R_m)^{v_k+1} \geq \gamma_L^{v_k+1}/(v_k+1)!$, we have that

$$\begin{aligned}
\liminf_{m\to\infty}[a(Q_m)Q_m^{d-1}]^{-1}\mathbb{E}\Big[\sum_{k=0}^{\infty}\tilde{S}_{v_k,n_m}(t)\Big] &= \liminf_{m\to\infty}\sum_{k=0}^{n_m-1}\binom{n_m}{v_k+1}f(R_m)^{v_k+1}\frac{a(R_m)R_m^{d-1}}{a(Q_m)Q_m^{d-1}} \\
&\quad\times\int_{(\mathbb{R}^d)^{v_k+1}} h_t^{v_k}(0,y_1\dots,y_{v_k})f(x)1\Big\{\|x\|>R_m\Big\} \\
&\quad\times\prod_{i=1}^{v_k} f(x+y_i)1\Big\{\|x+y_i\|>R_m\Big\}\,\mathrm{d}x\,\mathrm{d}\mathbf{y}, \\
&\geq \sum_{k=0}^{\infty}\frac{\gamma_L^{v_k+1}}{(v_k+1)!}\mathcal{I}_{v_k}(t),
\end{aligned}$$

by Fatou's lemma. The situation with (*ii*) is similarly established for with an application of the dominated convergence theorem, by noting summability of factorial terms and DCT condition on the integral at (5.41).

For the proof of the variance, we note that the density $f(x) = C\exp(-\psi(\|x\|))$ is decreasing in $\|x\|$ due to the conditions on ψ. Hence, the same bound as in (5.17) holds. Applying either the Efron-Stein inequality or the Poincaré inequality yields

$$\begin{aligned}
&[a(Q_m)Q_m^{d-1}]^{-1}\mathrm{Var}\Big(\sum_{k=0}^{\infty}\tilde{S}_{v_k,n_m}(t)\Big) \\
&\qquad\leq \frac{2Kn_m}{a(Q_m)Q_m^{d-1}}\int_{\mathbb{R}^d} f(x)1\Big\{\|x\|>R_m\Big\}\,\mathrm{d}x \\
&\qquad= \frac{C^* n_m f(R_m)a(R_m)R_m^{d-1}}{a(Q_m)Q_m^{d-1}}\int_0^{\infty}\frac{f(R_m+a(R_m)\rho)}{f(R_m)}\left(\frac{a(R_m)\rho}{R_m}+1\right)^{d-1}\mathrm{d}\rho
\end{aligned} \tag{5.42}$$

by the changes of variable $x \mapsto r\theta$ and $r \mapsto R_m + a(R_m)\rho$. The bound (*iii*) then follows from using (5.35), (5.34), and the conditions of the Lemma to bound (5.42).

□

Though the main theme of this study is topological crackle, if $\zeta = 0$ in the above (in particular when $\tau > 1$) we may still establish some interesting results for some topological functionals of an extreme sample cloud with density f as in (5.27). That crackle does not occur when $\tau > 1$ is well known from [4]—but Corollary 5.3.5 establishes a more homologically oriented result for light-tailed densities than previous seen and implies results pertaining to the vanishing of the Euler characteristic process and the sum of all Betti numbers, what Adamaszek calls the *total Betti number* in [1]. Note that in Corollary 5.3.5 we make no assumptions on τ beyond the fact that τ must be such that $a(z) \to 0$ as $z \to \infty$.

Corollary 5.3.5. *Assuming the conditions of Lemma 5.3.4, suppose that $\zeta = 0$, then for every $t \geq 0$,*

$$[a(Q_m)Q_m^{d-1}]^{-1}\mathbb{E}\Big[\sum_{k=0}^{M} \tilde{S}_{v_k,n_m}(t)\Big] \to 0, \quad n \to \infty. \tag{5.43}$$

As a result, if $nf(R_n) \to \xi \in (0, \infty)$, then for any $t \geq 0$,

(i)
$$\frac{\tilde{\chi}_n(t)}{a(R_n)R_n^{d-1}} \xrightarrow{P} 0, \quad n \to \infty$$

and

(ii)
$$\frac{\sum_{k=0}^{\infty} \tilde{\beta}_{k,n}(t)}{a(R_n)R_n^{d-1}} \xrightarrow{P} 0, \quad n \to \infty$$

where $\tilde{\beta}_{k,n}(t) := \beta_k(\mathcal{K}(\Phi_n \cap B(0, R_n)^c, t))$ for $t \geq 0$.

Proof. We begin by proving (5.43). In the case $a(z) \to 0$, then the limit in (5.40) is 0. Note that the same dominated convergence assumptions established in Lemma 5.3.4 hold so that it suffices to show that the quantity in (5.39) goes to zero as $m \to \infty$. To see this note that if $\langle \theta, y_i \rangle < 0$, then $\rho + \xi_m(\rho, \theta, y_i) \to -\infty$ and so $1_{A_m}(y_i) \to 0$ as $m \to \infty$. We can then consider the case when $\langle \theta, y_i \rangle > 0$ as $\langle \theta, y_i \rangle = 0$ only on a set of ν_{d-1} measure 0. If this is the case then $\xi(\rho, \theta, y_i) \to \infty$ as $m \to \infty$ and the limit in (5.40) is zero.

We now may use (5.43) to prove (i) and (ii). The above implies that if $\zeta = 0$ then

$$[a(R_n)R_n^{d-1}]^{-1}\mathbb{E}\Big[\sum_{k=0}^{\infty} \tilde{S}_{k,n}(t)\Big] \to 0.$$

Now, for any ϵ, the Markov inequality implies that

$$\mathbb{P}\Big(|\tilde{\chi}_n(t)| \geq \epsilon a(R_n)R_n^{d-1}\Big) \vee \mathbb{P}\Big(\sum_{k=0}^{\infty} \tilde{\beta}_{k,n}(t) \geq \epsilon a(R_n)R_n^{d-1}\Big) \leq \frac{\mathbb{E}\Big[\sum_{k=0}^{\infty} \tilde{S}_{k,n}(t)\Big]}{\epsilon a(R_n)R_n^{d-1}}.$$

So the conclusion follows by noting that $\beta_{k,n}(t) \leq S_{k,n}(t)$. □

The results (i) and (ii) in Corollary 5.3.5 are highly similar to the result of Theorem 5.2 in [70], though here we showed a result for the weak core regime, albeit in much less generality. The complete picture of the homology of the whole sample cloud in the case of these unbounded densities is mostly understood, but demonstrating that the whole Gaussian point cloud becomes contractible for any fixed radius of balls centered around the points—and that this is not the case for point clouds distributed according to f at (5.27) with $\zeta > 0$—would be a worthwhile and interesting result.

Assume the conditions of Proposition 5.3.1. Now by the bound

$$\sum_{k=0}^{\infty} \tilde{\beta}_{k,n}(t) \leq \sum_{k=0}^{\infty} \tilde{S}_{k,n}(t),$$

and any positive sequence $(\omega_n)_{n\geq 1}$ such that $a(R_n)R_n^{d-1}/\omega_n \to 0$, $n \to \infty$ we have that

$$\left(\omega_n^{-1} \sum_{k=0}^{\infty} \tilde{\beta}_{k,n}(t),\ t \geq 0\right) \to 0, \quad \text{a.s. in } (D, U),$$

as $n \to \infty$, by Proposition 5.3.1 and an application of Proposition 2.3.1 as well. The same result holds in the heavy tail case.

We continue now with the proof of Proposition 5.3.1.

Proof of Proposition 5.3.1. This proof is essentially similar in character to the proof of Proposition 5.2.1, but involves some slightly more complex machinery. Let us redefine $j_m = \lfloor e^{m^\gamma} \rfloor$ for $m \in \mathbb{N}_0$ and $0 < \gamma < 1$. By the conditions on d and τ we may take

$$\gamma \in \Big(\frac{\tau}{d-\tau}, 1\Big). \tag{5.44}$$

Let p_m and q_m remain as in (5.18) and (5.19) respectively. Additionally, we introduce

$$t_m := \arg\max\{j_m \le \ell \le j_{m+1} : a(R_\ell)R_\ell^{d-1}\},$$
$$u_m := \arg\min\{j_m \le \ell \le j_{m+1} : a(R_\ell)R_\ell^{d-1}\}.$$

Let $U_k^m(t)$ and $T_k^m(t)$ be defined as at (5.20) and (5.21) respectively. Thus, we see that

$$\frac{T_k^m(t)}{a(R_{t_m})R_{t_m}^{d-1}} \le \frac{\tilde{S}_{k,n}(t)}{a(R_n)R_n^{d-1}} \le \frac{U_k^m(t)}{a(R_{u_m})R_{u_m}^{d-1}},$$

and hence have a similar setup as (5.22). As in the proof of Proposition 5.2.1 we prove the strong law of large numbers for $\chi_n^{(i)}(t)$, when $i = 0$—which we will denote $\chi_n(t)$—and let $T_m(t) = \sum_{k=0}^\infty T_{2k}^m(t)$ and $U_m(t) = \sum_{k=0}^\infty U_{2k}^m(t)$ be defined in analogy with the identically named quantities in Proposition 5.2.1. Hence we get

$$\liminf_{m\to\infty} \frac{T_m(t) - \mathbb{E}[T_m(t)]}{a(R_{t_m})R_{t_m}^{d-1}} + \liminf_{m\to\infty} \frac{\mathbb{E}[T_m(t)]}{a(R_{t_m})R_{t_m}^{d-1}} \le \liminf_{n\to\infty} \frac{\chi_n(t)}{a(R_n)R_n^{d-1}}, \tag{5.45}$$

and

$$\limsup_{n\to\infty} \frac{\chi_n(t)}{a(R_n)R_n^{d-1}} \le \limsup_{m\to\infty} \frac{\mathbb{E}[U_m(t)]}{a(R_{u_m})R_{u_m}^{d-1}} + \limsup_{m\to\infty} \frac{U_m(t) - \mathbb{E}[U_m(t)]}{a(R_{u_m})R_{u_m}^{d-1}}. \tag{5.46}$$

Let us consider the expectation term in (5.46) first. Setting $n_m = j_{m+1}$, $R_m = R_{q_m}$, $Q_m = R_{u_m}$ and $v_k = 2k$, we see that Lemma 5.3.4(*ii*) implies that

$$\limsup_{m\to\infty} \frac{\mathbb{E}[U_m(t)]}{a(R_{u_m})R_{u_m}^{d-1}} \le \sum_{k=0}^\infty s_{2k}(t)$$

if we can show that $R_{q_m} \sim R_{u_m}$. Note that $\limsup_{m\to\infty} j_{m+1} f(R_{q_m}) = \xi$ because of the limit relation $\lim_{m\to\infty} j_{m+1}/j_m = 1$. To begin, we see that

$$Cne^{-\psi(R_n)} \sim \xi \;\Rightarrow\; \psi(R_n) \sim \log n.$$

As $\psi \in RV_\tau$ and is increasing then $\psi^{\leftarrow} \in RV_{1/\tau}$. Thus, as $\psi^{\leftarrow}(\psi(R_n)) \sim R_n$ then by Proposition 2.6 (iii) in [82] we have that

$$R_n \sim \psi^{\leftarrow}(\log n). \tag{5.47}$$

From this and the regular variation of $\psi^{\leftarrow}$, it easily follows that $R_{u_m} \sim R_{q_m}$. The asympotic relation $a(R_{u_m}) \sim a(R_{q_m})$ follows by the uniform convergence property of regularly varying functions. We once again conclude by applying the Borel-Cantelli lemma, giving us

$$\frac{U_m(t) - \mathbb{E}[U_m(t)]}{a(R_{u_m})R_{u_m}^{d-1}} \to 0, \quad \text{a.s.,} \quad n \to \infty. \tag{5.48}$$

We prove only the case of $U_m(t)$ as the situation with $T_m(t)$ is a symmetric argument. Hence,

$$\begin{aligned} &\sum_{m=1}^{\infty} \mathbb{P}(|U_m(t) - \mathbb{E}[U_m(t)]| \geq \epsilon a(R_{u_m})R_{u_m}^{d-1}) \\ &\qquad \leq \sum_{m=1}^{\infty} \frac{\text{Var}(U_m(t))}{\epsilon^2 (a(R_{u_m})R_{u_m}^{d-1})^2} \leq C^* \sum_{m=1}^{\infty} \frac{1}{a(R_{u_m})R_{u_m}^{d-1}}, \end{aligned} \tag{5.49}$$

so that it suffices to show that (5.49) is finite. By the constraint on γ in (5.44) we have the existence of constants $\delta_i > 0$, $i = 1, 2$ there exist $\delta_i > 0$, $i = 1, 2$, so that

$$\gamma(d - \tau - \delta_1)\Big(\frac{1}{\tau} - \delta_2\Big) > 1.$$

Then, $a \in \text{RV}_{1-\tau}$ implies that

$$a(R_{u_m})R_{u_m}^{d-1} \geq C^* R_{u_m}^{d-\tau-\delta_1}$$

for all $m \geq 1$. Note that by (5.47),

$$R_{u_m} \geq C^* \psi^{\leftarrow}\Big(\log u_m\Big) \geq C^* \psi^{\leftarrow}(\log j_m) \geq C^* m^{\gamma(1/\tau - \delta_2)}$$

again for all $m \geq 1$. Therefore,

$$a(R_{u_m})R_{u_m}^{d-1} \geq C^* m^{\gamma(d-\tau-\delta_1)(1/\tau-\delta_2)},$$

and

$$\sum_{m=1}^{\infty} \frac{1}{a(R_{u_m})R_{u_m}^{d-1}} \leq C^* \sum_{m=1}^{\infty} \frac{1}{m^{\gamma(d-\tau-\delta_1)(1/\tau-\delta_2)}} < \infty.$$

We can argue in the same fashion for (5.45) and $(T_m(t) - \mathbb{E}[T_m(t)])/(a(R_{t_m})R_{t_m}^{d-1})$, which yields the proof, save for the strong law for $\tilde{S}_{k,n}(t)$. The proof for $\tilde{S}_{k,n}(t)$ as with the heavy tail case, follows directly from simpler arguments familiar to those above. □

Proof of Theorem 5.3.1. The proof is the same as Theorem 5.2.1, save for the fact we use Proposition 5.3.1 this time. To show finiteness of the limit in (5.30) we use property 2 of Definition 2.1.5 to see that

$$\begin{aligned} \Big| \sum_{k=0}^{\infty} (-1)^k s_k(t) \Big| &\leq \sum_{k=0}^{\infty} \frac{s_{d-1}\xi^{k+1}}{(k+1)!} \int_0^{\infty} \int_{(\mathbb{R}^d)^k} \prod_{i=1}^{k} 1\Big\{ \|y_i\| \leq ct \Big\}\, e^{-\rho}\, \mathrm{d}\mathbf{y}\, \mathrm{d}\rho \\ &\leq C^* \sum_{k=0}^{\infty} \frac{\Big((ct)^d \xi \omega_d \Big)^k}{k!} = C^* e^{(ct)^d \xi \omega_d} < \infty. \end{aligned}$$

□

REFERENCES

[1] Michał Adamaszek. "Extremal problems related to Betti numbers of flag complexes". In: *Discrete Applied Mathematics* 173 (2014), pp. 8–15.

[2] Robert J. Adler. "Some new random field tools for spatial analysis". In: *Stochastic Environmental Research and Risk Assessment* 22.6 (Oct. 2008), pp. 809–822.

[3] Robert J. Adler et al. "Persistent homology for random fields and complexes". In: *Borrowing strength: theory powering applications–a Festschrift for Lawrence D. Brown*. Institute of Mathematical Statistics, 2010, pp. 124–143.

[4] Robert J. Adler, Omer Bobrowski, and Shmuel Weinberger. "Crackle: The homology of noise". In: *Discrete and Computational Geometry* 52.4 (2014), pp. 680–704.

[5] V.I. Arnold, ed. *Arnold's Problems*. Translated and revised edition of the 2000 Russian original, with a preface by V. Philippov, A. Yakivchik, and M. Peters. Springer-Verlag Berlin Heidelberg, 2005.

[6] Guus Balkema and Paul Embrechts. *High Risk Scenarios and Extremes: A Geometric Approach*. European Mathematical Society, 2007.

[7] Guus Balkema and Paul Embrechts. *Multivariate excess distributions*. June 2004. URL: https://people.math.ethz.ch/~embrecht/ftp/guuspe08Jun04.pdf.

[8] Peter J. Bickel and Leo Breiman. "Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test". In: *The Annals of Probability* (1983).

[9] Peter J. Bickel and Michael J. Wichura. "Convergence criteria for multiparameter stochastic processes and some applications". In: *The Annals of Mathematical Statistics* (1971), pp. 1656–1670.

[10] Patrick Billingsley. *Convergence of Probability Measures*. 2nd edition. John Wiley & Sons, 1999.

[11] Christophe A. N. Biscio et al. "Testing goodness of fit for point processes via topological data analysis". In: *Electronic Journal of Statistics* 14.1 (2020), pp. 1024–1074.

[12] Bartłomiej Błaszczyszyn and D. Yogeshwaran. "Clustering and percolation of point processes". In: *Electronic Journal of Probability* (2013).

[13] Omer Bobrowski. "Algebraic Topology of Random Fields and Complexes". PhD. Technion - Israel Institute of Technology, 2012.

[14] Omer Bobrowski and Robert J. Adler. “Distance functions, critical points, and the topology of random Čech complexes”. In: *Homology, Homotopy and Applications* 16.2 (July 2014), pp. 311–344.

[15] Omer Bobrowski and Matthew Strom Borman. “Euler integration of Gaussian random fields and persistent homology”. In: *Journal of Topology and Analysis* 4.1 (2012), pp. 49–70.

[16] Omer Bobrowski and Matthew Kahle. “Topology of random geometric complexes: a survey”. In: *Journal of Applied and Computational Topology* 1.3-4 (June 2018), pp. 331–364.

[17] Omer Bobrowski and Sayan Mukherjee. “The topology of probability distributions on manifolds”. In: *Probability Theory and Related Fields* 161.3-4 (2015), pp. 651–686.

[18] Omer Bobrowski and Primoz Skraba. “Homological percolation and the Euler characteristic”. In: *Physical Review E* (2020).

[19] Omer Bobrowski and Primoz Skraba. “Homological percolation: The formation of giant k-Cycles”. In: *International Mathematics Research Notices* (2020).

[20] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and Topological Inference*. Vol. 57. Cambridge University Press, 2018.

[21] Karol Borsuk. “On the imbedding of systems of compacta in simplicial complexes”. In: *Fundamenta Mathematicae* 35.1 (1948), pp. 217–234.

[22] Peter Bubenik. “Statistical topological data analysis using persistence landscapes”. In: *Journal of Machine Learning Research* (2015).

[23] Gunnar Carlsson. “Topological pattern recognition for point cloud data”. In: *Acta Numerica* 23 (2014), pp. 289–368.

[24] Gunnar Carlsson. “Topology and data”. In: *Bulletin of the American Mathematical Society* 46.2 (2009), pp. 255–308.

[25] Frédéric Chazal et al. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer International Publishing, 2016.

[26] Laurent Decreusefond, Matthias Schulte, and Christoph Thäle. “Functional poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry”. In: *Annals of Probability* 44.3 (June 2016), pp. 2147–2197.

[27] Laurent Decreusefond et al. “Simplicial homology of random configurations”. In: *Advances in Applied Probability* 46.2 (Mar. 2014), pp. 325–347.

[28] Luc Devroye and Gary L. Wise. “Detection of abnormal behavior via nonparametric estimation of the support”. In: *SIAM Journal on Applied Mathematics* 38.3 (1980), pp. 480–488.

[29] Persi Diaconis. “Talk on application of topology”. Video. MSRI Workshop on Application of Topology in Science and Engineering. Sept. 2006.

[30] Monroe D. Donsker. “An invariance principle for certain probability limit theorems”. In: *Four papers on probability*. Memoirs of the American Mathematical Society 6. Providence, Rhode Island: AMS, 1951, pp. 50–61.

[31] Rick Durrett. *Probability: Theory and Examples*. 4th. Cambridge university press, 2010.

[32] Herbert Edelsbrunner and John Harer. “Persistent homology: a survey”. In: *Contemporary Mathematics* (2008), pp. 1–26.

[33] Herbert Edelsbrunner and John L. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010.

[34] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. “Topological persistence and simplification”. In: *Discrete and Computational Geometry* 28.4 (2002), pp. 511–533.

[35] Nasrollah Etemadi. “An elementary proof of the strong law of large numbers”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 55.1 (Feb. 1981), pp. 119–122.

[36] Brittany Terese Fasy et al. *Introduction to the R package TDA*. Nov. 2014. arXiv: 1411.1830. URL: http://arxiv.org/abs/1411.1830.

[37] Patrizio Frosini. “Measuring shapes by size functions”. In: *Intelligent Robots and Computer Vision X: Algorithms and Techniques*. Ed. by David P. Casasent. Vol. 1607. International Society for Optics and Photonics. SPIE, 1992, pp. 122–133.

[38] Vladimir Gershkovich and Hyam Rubinstein. “Morse theory for min-type functions”. In: *Asian Journal of Mathematics* 1.4 (1997), pp. 696–715.

[39] Robert Ghrist. “Barcodes: The persistent topology of data”. In: *Bulletin of the American Mathematical Society* 45.01 (Oct. 2007), pp. 61–76.

[40] Robert Ghrist and Abubakr Muhammad. “Coverage and hole-detection in sensor networks via homology”. In: *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.* 1. IEEE, 2005, pp. 254–260.

[41] E. N. Gilbert. “Random plane networks”. In: *Journal of the Society for Industrial and Applied Mathematics* (1961).

[42] Akshay Goel, Khanh Duy Trinh, and Kenkichi Tsunoda. “Strong law of large numbers for Betti Numbers in the thermodynamic regime”. In: *Journal of Statistical Physics* 174.4 (Feb. 2019), pp. 865–892.

[43] J. C. Gower and G. J. S. Ross. “Minimum spanning trees and single linkage cluster analysis”. In: *Applied Statistics* 18.1 (1969), p. 54.

[44] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2001.

[45] Yasuaki Hiraoka and Tomoyuki Shirai. “Minimum spanning acycle and lifetime of persistent homology in the Linial-Meshulam process”. In: *Random Structures & Algorithms* 51.2 (Sept. 2017), pp. 315–340.

[46] Yasuaki Hiraoka, Tomoyuki Shirai, and Khanh Duy Trinh. “Limit theorems for persistence diagrams”. In: *The Annals of Applied Probability* 28.5 (Oct. 2018), pp. 2740–2780.

[47] Daniel Hug, Günter Last, and Matthias Schulte. “Second-order properties and central limit theorems for geometric functionals of Boolean models”. In: *Annals of Applied Probability* 26.1 (Aug. 2016), pp. 73–135.

[48] Umar Islambekov, Monisha Yuvaraj, and Yulia R. Gel. “Harnessing the power of topological data analysis to detect change points”. In: *Environmetrics* 31.1 (2020), pp. 1–11.

[49] Matthew Kahle. “Random geometric complexes”. In: *Discrete and Computational Geometry* 45.3 (2011), pp. 553–573.

[50] Matthew Kahle. “Topology of random clique complexes”. In: *Discrete Mathematics* (2009).

[51] Matthew Kahle and Elizabeth Meckes. “Erratum to ‘Limit theorems for Betti numbers of random simplicial complexes’”. In: *Homology, Homotopy and Applications* 18.1 (2016), pp. 129–142.

[52] Matthew Kahle and Elizabeth Meckes. "Limit theorems for Betti numbers of random simplicial complexes". In: *Homology, Homotopy and Applications* 15.1 (July 2013), pp. 343–374.

[53] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 1991.

[54] Johannes Krebs. "On limit theorems for persistent Betti numbers from dependent data". In: *Stochastic Processes and their Applications* 139 (May 2021), pp. 139–174.

[55] Johannes Krebs and Christian Hirsch. "Functional central limit theorems for persistent Betti numbers on cylindrical networks". In: *Scandinavian Journal of Statistics* (Mar. 2021).

[56] Johannes Krebs and Wolfgang Polonik. *On the asymptotic normality of persistent Betti numbers*. Mar. 2019. arXiv: 1903.03280. URL: http://arxiv.org/abs/1903.03280.

[57] Johannes Krebs, Benjamin Roycraft, and Wolfgang Polonik. *On approximation theorems for the Euler characteristic with applications to the bootstrap*. May 2020. arXiv: 2005.07557. URL: http://arxiv.org/abs/2005.07557.

[58] Günter Last and Mathew D. Penrose. *Lectures on the Poisson Process*. Vol. 7. Cambridge University Press, 2017.

[59] Günter Last and Mathew D. Penrose. "Poisson process Fock space representation, chaos expansion and covariance inequalities". In: *Probability Theory and Related Fields* (2011).

[60] Ronald Meester and Rahul Roy. *Continuum Percolation*. Vol. 119. Cambridge University Press, 1996.

[61] Yuriy Mileyko, Sayan Mukherjee, and John Harer. "Probability measures on the space of persistence diagrams". In: *Inverse Problems* 27.12 (2011).

[62] James R. Munkres. *Topology*. 2nd. Pearson Modern Classic. Pearson, 2018.

[63] Richard A. Neher, Klaus Mecke, and Herbert Wagner. "Topological estimation of percolation thresholds". In: *Journal of Statistical Mechanics: Theory and Experiment* (2008).

[64] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. "A topological view of unsupervised learning from noisy data". In: *SIAM Journal on Computing* 40.3 (2011), pp. 646–663.

[65] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. "Finding the homology of submanifolds with high confidence from random samples". In: *Discrete & Computational Geometry* 39.1-3 (Mar. 2008), pp. 419–441.

[66] Takashi Owada. *Convergence of persistence diagram in the subcritical regime*. Mar. 2021. arXiv: 2103.12943. URL: http://arxiv.org/abs/2103.12943.

[67] Takashi Owada. "Functional central limit theorem for subgraph counting processes". In: *Electronic Journal of Probability* 22 (May 2017), pp. 1–38.

[68] Takashi Owada. "Limit theorems for Betti numbers of extreme sample clouds with application to persistence barcodes". In: *The Annals of Applied Probability* 28.5 (2018), pp. 2814–2854.

[69] Takashi Owada. "Topological crackle of heavy-tailed moving average processes". In: *Stochastic Processes and their Applications* 129.12 (2019), pp. 4965–4997.

[70] Takashi Owada and Robert J. Adler. "Limit theorems for point processes under geometric constraints (and topological crackle)". In: *Annals of Probability* 45.3 (2017), pp. 2004–2055.

[71] Takashi Owada and Omer Bobrowski. "Convergence of persistence diagrams for topological crackle". In: *Bernoulli* 26.3 (Aug. 2020), pp. 2275–2310.

[72] Takashi Owada and Andrew M. Thomas. "Limit theorems for process-level Betti numbers for sparse and critical regimes". In: *Advances in Applied Probability* 52.1 (2020), pp. 1–31.

[73] Takashi Owada and Zifu Wei. *Functional strong law of large numbers for Betti numbers in the tail*. Mar. 2021. arXiv: 2103.05799. URL: http://arxiv.org/abs/2103.05799.

[74] Giovanni Peccati and Murad S. Taqqu. *Wiener Chaos: Moments, Cumulants and Diagrams*. Vol. 1. Springer-Verlag Italia, 2011.

[75] Mathew D. Penrose. "Central limit theorems for k-nearest neighbour distances". In: *Stochastic processes and their applications* 85.2 (2000), pp. 295–320.

[76] Mathew D. Penrose. *Random Geometric Graphs*. Vol. 5. Oxford University Press, 2003.

[77] Mathew D. Penrose and J. E. Yukich. "Central limit theorems for some graphs in computational geometry". In: *Annals of Applied Probability* (2001).

[78] Mathew D. Penrose and J. E. Yukich. “Weak laws of large numbers in geometric probability”. In: *Annals of Applied Probability* 13.1 (Jan. 2003), pp. 277–303.

[79] Jose A. Perea. “A brief history of persistence”. In: *Morfismos* 23.1 (2019), pp. 1–16.

[80] David Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, 1984.

[81] Sidney I. Resnick. *Extreme Values, Regular Variation, and Point Processes.* Springer Science & Business Media, 1987.

[82] Sidney I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Stochastic Modeling.* Springer Science & Business Media, 2007.

[83] Eitan Richardson and Michael Werman. “Efficient classification using the Euler characteristic”. In: *Pattern Recognition Letters* (2014).

[84] David S. Richeson. *Euler's Gem: The Polyhedron Formula and the Birth of Topology.* Princeton University Press, 2008.

[85] Vanessa Robins. “Betti number signatures of homogeneous Poisson point processes”. In: *Physical Review E* 74.6 (Dec. 2006), p. 061107.

[86] Vanessa Robins. “Computational topology for point data: Betti numbers of α-shapes”. In: (2002), pp. 261–274.

[87] Vanessa Robins. “Towards computing homology from finite approximations”. In: *Topology Proceedings* (1999).

[88] Rolf Schneider and Wolfgang Weil. *Stochastic and Integral Geometry.* Springer Science & Business Media, 2008.

[89] Eugene Seneta. *Regularly Varying Functions.* Vol. 508. Lecture Notes in Mathematics. Berlin: Springer, 1976.

[90] Nikhil Singh et al. “Topological descriptors of histology images”. In: *Machine Learning in Medical Imaging.* Ed. by Guorong Wu, Daoqiang Zhang, and Luping Zhou. Cham: Springer International Publishing, 2014, pp. 231–239.

[91] J. Michael Steele. “An Efron-Stein inequality for nonsymmetric statistics”. In: *The Annals of Statistics* 14.2 (June 1986), pp. 753–758.

[92] Andrew M. Thomas and Takashi Owada. “Functional limit theorems for the Euler characteristic process in the critical regime”. In: *Advances in Applied Probability* 53.1 (Mar. 2021), pp. 57–80.

[93] Andrew M. Thomas and Takashi Owada. "Functional strong laws of large numbers for Euler characteristic processes of extreme sample clouds". In: *Extremes* (May 2021).

[94] Khanh Duy Trinh. "A remark on the convergence of Betti numbers in the thermodynamic regime". In: *Pacific Journal of Mathematics for Industry* 9.4 (Mar. 2017).

[95] Khanh Duy Trinh. "On central limit theorems in stochastic geometry for add-one cost stabilizing functionals". In: *Electronic Communications in Probability* 24 (2019).

[96] Larry Wasserman. "Topological data analysis". In: *Annual Review of Statistics and Its Application* 5.1 (Sept. 2016), pp. 501–532.

[97] Ward Whitt. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Applications to Queues*. Springer Science & Business Media, 2002.

[98] K. J. Worsley et al. "A three-dimensional statistical analysis for CBF activation studies in human brain". In: *Journal of Cerebral Blood Flow and Metabolism* (1992).

[99] D. Yogeshwaran and Robert J. Adler. "On the topology of random complexes built over stationary point processes". In: *Annals of Applied Probability* 25.6 (2015), pp. 3338–3380.

[100] D. Yogeshwaran, Eliran Subag, and Robert J. Adler. "Random geometric complexes in the thermodynamic regime". In: *Probability Theory and Related Fields* 167 (2017), pp. 107–142.

VITA

Andrew Michael Thomas was born in 1992 and grew up in Eden Prairie, Minnesota. An inveterate dilettante, he found that probability and statistics was the only real thing that could ever hold his focus. He received a B.S. in Statistics and a B.A. in Linguistics with a Mathematics minor at the University of Minnesota, Twin Cities in May of 2014. He joined the Department of Statistics at Purdue University as a graduate student in the Fall of 2015. He received his M.S. in Mathematical Statistics in August of 2017. Upon receiving his Ph.D. in June 2021, he will join the Center for Applied Mathematics at Cornell University as a postdoctoral associate.