

Convergence of persistence diagrams for topological crackle

TAKASHI OWADA¹ and OMER BOBROWSKI²

¹Department of Statistics, Purdue University, West Lafayette, IN 47907, USA. E-mail: owada@purdue.edu

²Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa, 32000, Israel. E-mail: omer@ee.technion.ac.il

In this paper, we study the persistent homology associated with topological crackle generated by distributions with an unbounded support. Persistent homology is a topological and algebraic structure that tracks the creation and destruction of topological cycles (generalizations of loops or holes) in different dimensions. Topological crackle is a term that refers to topological cycles generated by random points far away from the bulk of other points, when the support is unbounded. We establish weak convergence results for persistence diagrams – a point process representation for persistent homology, where each topological cycle is represented by its (*birth*, *death*) coordinates. In this work, we treat persistence diagrams as random closed sets, so that the resulting weak convergence is defined in terms of the Fell topology. Using this framework, we show that the limiting persistence diagrams can be divided into two parts. The first part is a deterministic limit containing a densely-growing number of persistence pairs with a shorter lifespan. The second part is a two-dimensional Poisson process, representing persistence pairs with a longer lifespan.

Keywords: extreme value theory; fell topology; persistent homology; point process; topological crackle

1. Introduction

One of the main themes of the present paper is *topological crackle*. Originally proposed by [1] as an analogy of audio crackling, topological crackle has been modeled as the layered structure of an increasing number of topological cycles away from the origin. Typically, crackle appears in topological manifold learning problems. For example, suppose we wish to recover the topology of an annulus. Given a random sample $\mathcal{P}$, a common practice is to place balls of radius r around $\mathcal{P}$, and consider their union $B_r(\mathcal{P})$ (Figure 1). If the sample noise is small enough as in Figure 1(b), the union $B_r(\mathcal{P})$ is similar in shape to an annulus, and recovering its topology is feasible [28,29]. However, if the distribution of noise has a heavy tail as in Figure 1(c), many extraneous components away from the center of an annulus will make it hard (or even impossible) to recover its topology. This phenomenon is an example of topological crackle.

Taking a single point at the origin as an underlying manifold, Figure 2 visualizes topological crackle as a layered structure of *Betti numbers* of various dimensions. In particular, the k th Betti number counts the number of k -cycles. Loosely speaking, a k -cycle is a structure that is equivalent to a k -dimensional sphere as a boundary of a $(k + 1)$ -dimensional ball. For each individual layer except most inner and outer ones, there is a unique dimension $k \in \{0, \dots, d - 1\}$ such that the k th Betti number is approximated by a Poisson distribution, while all the other Betti numbers either vanish or diverge [1,32].

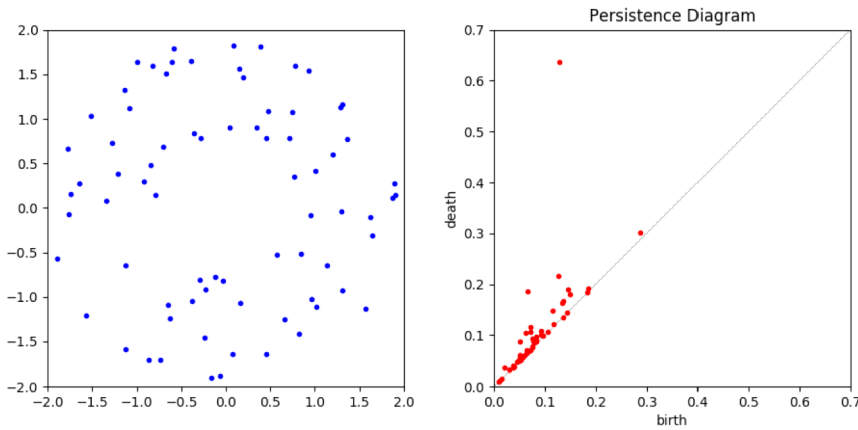


Figure 3. The persistence diagram of a random Čech filtration. On the left figure, the point process is generated on an annulus in $\mathbb{R}^2$. The persistence diagram on the right describes the birth and death times (radii) of all the 1-cycles that appear in this filtration. Notice that most of the points in the persistence diagram are close to the diagonal where birth time equals death time, and one might consider these cycles as “noise.” There is one point that stands out in the diagram, which corresponds to the hole of the annulus. The persistent homology was computed using the GUDHI library [38].

cycle first appears in the filtration and the time at which it is terminated (becomes trivial, or “filled in”), respectively.

Commonly, the output of persistent homology (i.e., a list of birth/death times) is summarized in a plot known as *persistence diagram*; see Figure 3. In this plot, a single point is drawn for any k -cycle, for which the x -axis value represents its birth time, and the y -axis value represents the death time.

The study of homology and persistent homology generated by random data, started in [24], and has been an active research topics over the past decade (see the survey in [10]). Much of this study is dedicated to examining the behavior of noise (i.e., point clouds that contain no intrinsic topological structure), and can be thought of as the study of “null-models” for TDA.

The primary objective of the current study is to establish limit theory of persistence diagram associated with topological crackle. A key notion in our approach is the so-called *Fell topology*, which is the most standard topology on closed sets [27], allowing us to treat persistence diagram as a random closed set in $\mathbb{R}^2$. A more formal discussion on the Fell topology is given in Section 2.3. The main discovery of the present paper is that under some assumptions, the persistence diagram of topological crackle in dimension k for n vertices, denoted $\Phi_n^{(k)}$, has the following limit.

$$\Phi_n^{(k)} \Rightarrow \Phi^{(k,p)} \cup B_{k,p-1}, \quad n \rightarrow \infty, \quad (1.1)$$

where $\Phi^{(k,p)}$ is a spatial Poisson process associated with k -cycles generated on p vertices, and $B_{k,p-1}$ is a related deterministic set, both formally defined in Section 3; see also Figure 6. The convergence “ $\Rightarrow$ ” is in the Fell topology for closed sets.

We remark that the notion of Fell topology is totally new within the context of TDA, while it is more common in EVT. For example, stochastic properties of standard graphical tools in EVT, such as a mean excess plot and a QQ-plot, have been explored via convergence theorems under the Fell topology [14,15,19]. Other publications in EVT, which use Fell topology but are not related to graphical tools, include [3,36].

From the viewpoints of TDA, highly relevant to this paper is the work in [23], studying the distribution of points in the persistence diagram generated by point processes in a d -dimensional box. In [23], persistence diagrams are considered as Radon measures, for which the authors prove the existence of a limit in the form of a deterministic measure. In addition, they provide a law of large numbers and a central limit theorem for the persistent Betti numbers, that is, the number of k -cycles that exist over a given range of radii.

In the case of topological crackle, the analysis in [23] is no longer valid. Alternatively, we shall exploit the idea of [1,31,32], which have studied fixed, rather than persistent, homology generated by distributions with unbounded supports. The first main finding in [1,31,32] was that if the underlying distribution generating data, has a tail at least as heavy as that of an exponential distribution, then k -cycles keep appearing far away from the origin. The second main finding in [1,31,32] was the emergence of a layered structure dividing the Euclidean space as in Figure 2, with each layer occupied by cycles of different dimensions and amounts. Briefly, as we get closer to the origin, the higher-dimensional cycles appear and their number increases. The fact that different regions in space are occupied by different types of structures at different quantities, suggests that one may not look for a single limit theorem for fixed and persistent homology. Instead, we can only provide separate limit theorems for each individual region. More specifically, if the cycles are distributed so densely that their number grows to infinity as the sample size increases, the number of cycles (i.e., Betti number) obeys a central limit theorem [31]. On the other hand, if the spatial distribution of cycles is sparse enough, the Betti numbers will be governed by a Poisson limit theorem [32]. Similar phenomena have been pointed out in a series of works [9,12,23,25,39], in which various limit theorems for topological invariants in different regimes were derived (though they are not directly related to the layered structure described above).

The primary benefit of our approach using the Fell topology is that one can establish limit theorems for the entire persistence diagram, even though the nature of the distribution of persistence pairs (i.e., points on the persistence diagram) differs from region to region. To see this more clearly, let us consider the case where the persistence diagram is approximately divided into two regions, such that the persistence pairs are distributed densely in one region, and in the other region, the distribution is much more sparse. This is roughly the same picture of persistence diagram in our main results. In contrast to the previous works cited above, our approach allows to describe the entire persistence diagram by a “single” limit theorem as in (1.1), which would help us get a whole picture of the limiting persistence diagram. Note that the persistence diagrams we study here only describe cycles generated by the crackle, that is, outside a given region, and does not address the rest of the cycles generated inside and around the core (defined later).

The remainder of this paper is organized as follows. In Section 2, we introduce the terminology used throughout the paper. Section 3 provides the main results of this paper, considering heavy tailed distributions. In Section 4, we discuss the behavior of the model for exponentially decaying distributions. We will classify the results in terms of heaviness of a tail of an underlying distribution. Such classification is typical in EVT. The proofs for both sections are presented in Section 5.

2. Preliminaries

2.1. Geometric complexes

An *abstract simplicial complex* over a set S is a collection of finite subsets $X \subset 2^S$ with the requirement that if $A \in X$ and $B \subset A$ then $B \in X$. A subset in X of size $k + 1$ is called a k -*simplex*, and commonly denoted as $\sigma = [x_0, \dots, x_k]$.

In this work, we discuss abstract simplicial complexes that are generated by a set of points $\mathcal{P} \subset \mathbb{R}^d$, called *geometric complexes*. Among many candidates of geometric complexes (see [21]), the present paper focuses on one of the most studied ones, a *Čech complex*. For construction, we start by fixing a radius $r > 0$, and drawing balls of radius r around the points in $\mathcal{P}$.

Definition 2.1. A *Čech complex* $\mathcal{C}_r(\mathcal{P})$ is defined by the following two conditions:

1. The 0-simplices are the points in $\mathcal{P}$.
2. A k -simplex $[x_0, \dots, x_k]$ is in $\mathcal{C}_r(\mathcal{P})$ if $\bigcap_{j=0}^k B(x_j; r) \neq \emptyset$,

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

One of the key properties of the Čech complex $\mathcal{C}_r(\mathcal{P})$, known as the Nerve lemma (see, e.g., Theorem 10.7 of [8]), asserts that the union of balls $B_r(\mathcal{P}) := \bigcup_{p \in \mathcal{P}} B(p; r)$ and $\mathcal{C}_r(\mathcal{P})$ are homotopy equivalent.

2.2. Persistent homology

In this section, we wish to describe homology and persistent homology in a slightly non-rigorous way, which is enough for the reader to follow the statements and proofs in this paper. We suggest [13,20] as a good introductory reading, while a more rigorous coverage of algebraic topology is in [22].

Let X be a topological space. In this paper, we will consider homology with field coefficients $\mathbb{F}$, in which case homology is essentially a sequence of vector spaces denoted $H_0(X), H_1(X), H_2(X), \dots$. In particular, $H_k(X)$ is the quotient group $\ker \partial_k / \text{im } \partial_{k+1}$, where $\partial_k, \partial_{k+1}$ are boundary maps for X . In other words, the basis of $H_k(X)$ corresponds to a topological invariant generated by nontrivial k -dimensional cycles as a boundary of a $(k + 1)$ -dimensional body (henceforth we simply call it “ k -cycle”). If the dimension k is small, one can understand the concept more intuitively. For example, the basis of $H_0(X)$ corresponds to the connected components in X , and the basis of $H_1(X)$ corresponds to closed loops in X . The basis of $H_2(X)$ corresponds to cavities or “air bubbles” in X . Finally, the k th *Betti number*, denoted $\beta_k(X)$, is the rank of $H_k(X)$, representing the number of k -cycles in X ; see Figure 4.

In addition to the topology of a single space X , homology theory also analyzes mappings between spaces. If $f : X \rightarrow Y$ is a map between topological spaces, then the induced map $f_* : H_k(X) \rightarrow H_k(Y)$ is a linear transformation, describing how k -cycles in X are transformed into k -cycles in Y (or disappear). *Persistent homology* can be thought of as a “multiscale”

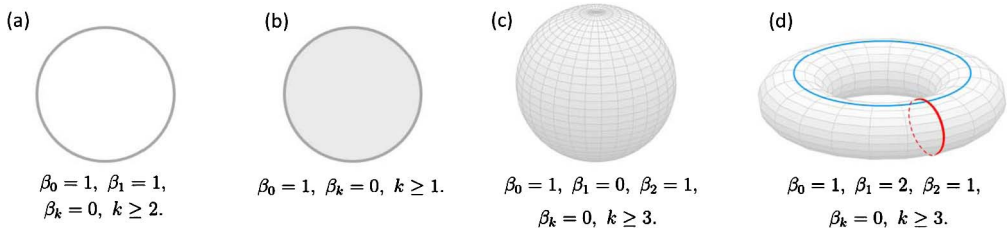


Figure 4. (a) One-dimensional sphere. (b) One-dimensional disk. (c) Two-dimensional sphere. (d) Two-dimensional torus. In (c), any closed loop winding around a sphere will vanish when it moves up and reaches the pole, so $\beta_1 = 0$. In (d), there are two independent closed loops, so $\beta_1 = 2$. This figure was taken from [31].

version of homology, designed to describe topological properties in a sequence of spaces. Let $\{X_t\}_t$ be a filtration of spaces, so that $X_s \subset X_t$ for all $s \leq t$. In this case, one can consider the collection of vector spaces $\{H_k(X_t)\}_t$, together with the corresponding linear transformations $\iota_*^{(s,t)} : H_k(X_s) \rightarrow H_k(X_t)$ for all $s \leq t$ induced by the inclusion map $\iota^{(s,t)} : X_s \rightarrow X_t$. Such a sequence is called a *persistence module* (cf. [13]). Essentially, this sequence allows us to track the evolution of k -cycles as they are formed and terminated throughout the filtration. Any k -cycle in this context is often referred to as a *persistence k -cycle*.

The theory developed for persistence modules allows for the definitions of *barcodes*, which consist of persistence intervals of the form $[\text{birth}, \text{death})$, representing the time (the value of t) when a given cycle first appears and the time when it disappears, respectively. Commonly, the information on the k th persistent homology is graphically provided via k th *persistence diagram*. This is a two-dimensional plot, where each persistence interval of the form $[\text{birth}, \text{death})$ is represented as a single point, with the x -axis representing birth time and the y -axis representing death time. The points on the k th persistence diagram are called the k th *persistence pairs*. Figure 3 shows an example of a first-order persistence diagram generated by a Čech filtration $\{\mathcal{C}_r(\mathcal{P})\}_r$, where $\mathcal{P}$ is a random sample from an annulus.

For the study of a Čech filtration, some structure must be imposed on persistence diagrams. First, notice that any persistence diagram is a subset of

$$\Delta := \{(x, y) : 0 \leq x \leq y\},$$

as death times always come after birth times. Further, given m, k and b_0 , if we consider k -cycles generated on m points whose birth time is b_0 , then there is a maximum value $d_0 \propto b_0 m^{1/k}$ for the possible death time (see Lemma 4.1 in [11]). Denoting $\pi_{k,m} = d_0/b_0$, the scaling invariance of persistent homology implies that all k -cycles generated on m points are restricted to the region

$$\Delta_{k,m} := \{(x, y) : 0 \leq x \leq y \leq \pi_{k,m} x\} \subset \Delta.$$

More precisely $\pi_{k,m}$ is a constant for which forming a persistence k -cycle is possible if $(x, y) \in \Delta_{k,m}$, but it becomes infeasible whenever $y > \pi_{k,m} x$. Notice that m has to be at least $k + 2$ in order to generate any cycles in the k th persistent homology for the Čech filtration. Finally, $\Delta_{k,m}$ is nondecreasing in m , that is, $\Delta_{k,m_1} \subset \Delta_{k,m_2}$ for all $m_1 \leq m_2$; see Figure 5.

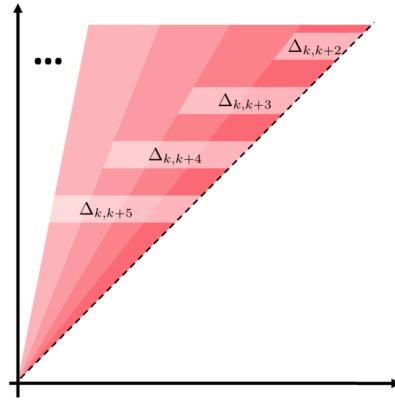


Figure 5. The structure of persistence diagrams generated by a Čech filtration. Note that $\Delta_{k,m}$ represents the region in which the k th persistence pairs generated by subsets of size m , may appear. For a Čech filtration, m has to be at least $k + 2$.

2.3. Fell topology

The novel idea of the current paper is to treat random points in the persistence diagram as closed sets in Δ . To this aim, we introduce *Fell topology*, which is perhaps the most standard topology on closed sets. Let $\mathcal{F}(\Delta)$ be the space of closed sets of Δ . A sequence of closed sets (F_n) converges to another closed set F if and only if the following two conditions hold:

- F hits an open set G , that is, $F \cap G \neq \emptyset$, implies there exists $N \geq 1$ such that for all $n \geq N$, F_n hits G .
- F misses a compact set K , that is, $F \cap K = \emptyset$, implies there exists $N \geq 1$ such that for all $n \geq N$, F_n misses K .

By this property, Fell topology can be recognized as “hit and miss” topology. The main reference used in this paper is [27].

The Fell topology is metrizable, and hence induces a Borel σ -field $\mathcal{B}(\mathcal{F}(\Delta))$. Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we say that $S : \Omega \rightarrow \mathcal{F}(\Delta)$ is a random closed set if

$$\{\omega : S \cap K \neq \emptyset\} \in \mathcal{A}$$

for every compact set K in Δ , that is, observing S , one can always determine if S hits or misses any given compact set. Let us provide some facts about the convergence of a sequence of random closed sets. Given random closed sets (S_n) and S , the weak convergence $S_n \Rightarrow S$ in $\mathcal{F}(\Delta)$ is implied by

$$\mathbb{P}(S_n \cap K \neq \emptyset) \rightarrow \mathbb{P}(S \cap K \neq \emptyset)$$

for every compact subset $K \subset \Delta$. For a measurable set $A \subset \Delta$ and $\epsilon > 0$, denote by

$$(A)^{\epsilon-} = \{(x, y) \in \Delta : d((x, y), A) < \epsilon\} \quad (2.1)$$

an open ϵ -envelop in terms of the Euclidean metric d . We say that S_n converges to S in probability if

$$\mathbb{P}([(S_n \setminus (S)^{\epsilon-}) \cup (S \setminus (S_n)^{\epsilon-})] \cap K \neq \emptyset) \rightarrow 0, \quad n \rightarrow \infty,$$

for every $\epsilon > 0$ (see Definition 6.19 in [27]).

3. Main results – Regularly varying tail case

In this section, we describe in detail the problem studied in this paper, and present the main results.

3.1. Definitions

The present section considers the following family of density functions with regularly varying tail. As is well known in EVT, in the one-dimensional case, the regular variation of a tail completely characterizes the maximum-domain of attraction of a Fréchet distribution [18].

Definition 3.1. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a probability density function. Let S^{d-1} be the unit sphere in $\mathbb{R}^d$.

1. We say that f is *spherically symmetric* if $f(\rho\theta_1) = f(\rho\theta_2)$ for any $\rho \in \mathbb{R}_+$ and $\theta_1, \theta_2 \in S^{d-1}$. For such functions, we define $f(\rho) := f(\rho\theta)$ for any $\theta \in S^{d-1}$.
2. We say that a spherically symmetric f has a *regularly varying tail* if there exists $\alpha > d$ such that

$$\lim_{\rho \rightarrow \infty} \frac{f(\rho t)}{f(\rho)} = t^{-\alpha} \quad \text{for all } t > 0. \quad (3.1)$$

Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables, having a common density function f satisfying the conditions in Definition 3.1. Let $N_n \sim \text{Poisson}(n)$ be a Poisson random variable, independent of (X_i) . Define the following point process:

$$\mathcal{P}_n := \begin{cases} \{X_1, \dots, X_{N_n}\} & \text{if } N_n > 0, \\ \emptyset & \text{if } N_n = 0. \end{cases} \quad (3.2)$$

Then one can show that $\mathcal{P}_n$ is a spatial Poisson process on $\mathbb{R}^d$ with intensity function nf (see, e.g., Chapter 5 in [35]).

Let $R = R(n)$ be a sequence of n growing to infinity, and consider

$$L_R = \{x \in \mathbb{R}^d : \|x\| \geq R\} = (B(0; R))^c.$$

The main objective in this paper is to study the “extreme-value behavior” of the persistent homology for the Čech filtration. More concretely, we study the behavior of persistence cycles far away from the origin, generated by the points in $\mathcal{P}_{n,R} := \mathcal{P}_n \cap L_R$ for large enough n . In

other words, we aim to analyze the limiting distribution of persistent homology for the filtration $\{\mathcal{C}_r(\mathcal{P}_{n,R})\}_{r \geq 0}$.

To that end, we define the following functions and objects. Recall that $\Delta = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq y\}$ is the upper infinite triangle in the first quadrant. Let $k \geq 1$ be an integer which remains fixed throughout the paper. For any finite set $\mathcal{Y} \subset \mathbb{R}^d$, let $\text{PH}_k(\mathcal{Y})$ be the k th persistent homology generated by $\{\mathcal{C}_r(\mathcal{Y})\}_{r \geq 0}$. We now define the finite counting measure on Δ ,

$$\mu_{\mathcal{Y}}^{(k)}(\cdot) := \sum_{\gamma \in \text{PH}_k(\mathcal{Y})} \delta_{(\gamma_b, \gamma_d)}(\cdot), \quad (3.3)$$

where γ represents a persistence cycle in $\text{PH}_k(\mathcal{Y})$ such that (γ_b, γ_d) are its birth and death times (radii), respectively. Moreover, $\delta_{(x,y)}(\cdot)$ is the Dirac measure at (x, y) . In other words, $\mu_{\mathcal{Y}}^{(k)}$ represents all the pairs (γ_b, γ_d) that appear in the k th persistence diagram generated by the set $\mathcal{Y}$. The finiteness of $\mu_{\mathcal{Y}}^{(k)}$ comes from a simple fact that if $|\mathcal{Y}| = m$ ($|\cdot|$ denotes cardinality of a given set), the number of k -cycles supported on m vertices is bounded by the number of k -simplices, which itself is bounded by $\binom{m}{k+1}$.

We need a few more definitions before introducing the main point processes. For a collection $\mathcal{Y}, \mathcal{Z}$ of points in $\mathbb{R}^d$ with $\mathcal{Y} \subset \mathcal{Z}$, define

$$\begin{aligned} h_R(\mathcal{Y}) &:= \mathbb{1}\{\mathcal{Y} \cap B(0; R) = \emptyset\}, \\ g_M(\mathcal{Y}, \mathcal{Z}) &:= \mathbb{1}\{\mathcal{C}_M(\mathcal{Y}) \text{ is a connected component of } \mathcal{C}_M(\mathcal{Z})\}, \\ g_M(\mathcal{Y}) &:= g_M(\mathcal{Y}, \mathcal{Y}) = \mathbb{1}\{\mathcal{C}_M(\mathcal{Y}) \text{ is connected}\}. \end{aligned} \quad (3.4)$$

The point processes we will examine in this paper are

$$\begin{aligned} \Phi_n^{(k,m)}(\cdot) &:= \sum_{\mathcal{Y} \subset \mathcal{P}_n, |\mathcal{Y}|=m} h_R(\mathcal{Y}) g_M(\mathcal{Y}, \mathcal{P}_n) \mu_{\mathcal{Y}}^{(k)}(M \cdot), \\ \Phi_n^{(k)}(\cdot) &:= \sum_{m=k+2}^{\infty} \Phi_n^{(k,m)}(\cdot), \end{aligned} \quad (3.5)$$

where $M = M(n)$ is a sequence of n , which will be explicitly determined below together with $R = R(n)$. Note that $M \equiv \text{constant}$ is permissible as a special case. The process $\Phi_n^{(k,m)}$ represents the persistence k -cycles that are generated by the Poisson process $\mathcal{P}_{n,R}$ defined above, such that the vertices forming these cycles belong to a single connected component of size m in the complex $\mathcal{C}_M(\mathcal{P}_{n,R})$. By the construction of (3.5) and the assumption that the random points generating data have a continuous distribution, the process $\Phi_n^{(k,m)}$ is simple (i.e., $\sup_{x \in \Delta} \Phi_n^{(k,m)}(\{x\}) \leq 1$ a.s.) and finite. Forming a k -cycle in the Čech complex requires at least $k+2$ vertices; hence, the sum defining $\Phi_n^{(k)}$ only starts at $m = k+2$. Furthermore, $\Phi_n^{(k)}$ is almost surely a sum of finite number of point processes, since $\Phi_n^{(k,m)} \equiv 0$ for all $m > |\mathcal{P}_n|$.

Finally, we remark that the same (or even easier) analysis can apply even when $k = 0$. There, many of the objects and functions will be degenerate, and we need to slightly modify the normalizing constants in the main theorem of the next section. In order to avoid blurring the main message of this paper, we only study the case $k \geq 1$.

3.2. Weak convergence

The primary goal in this paper is to prove a weak convergence theorem for $\Phi_n^{(k)}$ as $n \rightarrow \infty$. Since the point process $\Phi_n^{(k,m)}$ in (3.5) is simple and finite, the support of $\Phi_n^{(k,m)}$ is a finite random closed set (see Corollary 8.2 in [27]). By a slight abuse of notation, the letter $\Phi_n^{(k,m)}$ is used to denote both a point process and a random closed set as its support. In the latter treatment, $\Phi_n^{(k)}$ can be denoted as a union of $\Phi_n^{(k,m)}$'s,

$$\Phi_n^{(k)} = \bigcup_{m=k+2}^{\infty} \Phi_n^{(k,m)}.$$

This also represents almost surely a finite random closed set, because $\Phi_n^{(k,m)} = \emptyset$ whenever $m > |\mathcal{P}_n|$.

Consequently, the topology we use for the weak convergence below is Fell topology on closed sets of Δ . All the proofs, including those for corollaries that follow after Theorem 3.2, are deferred to Section 5.2.

Theorem 3.2. *Let f be a probability density function satisfying the conditions in Definition 3.1, and suppose that $R = R(n)$, $M = M(n)$ are chosen such that $R \rightarrow \infty$, $M/R \rightarrow 0$ as $n \rightarrow \infty$. Assume that there exists an integer $p \geq k + 2$ satisfying*

$$n^p M^{d(p-1)} R^d (f(R))^p \rightarrow 1, \quad n \rightarrow \infty. \quad (3.6)$$

Then, as $n \rightarrow \infty$,

$$\Phi_n^{(k,m)} \Rightarrow \begin{cases} \emptyset \text{ in } \mathcal{F}(\Delta) & \text{if } m > p, \\ \Phi^{(k,p)} \text{ in } \mathcal{F}(\Delta) & \text{if } m = p, \\ B_{k,m} \text{ in } \mathcal{F}(\Delta) & \text{if } k+2 \leq m < p, \end{cases} \quad (3.7)$$

and further,

$$\Phi_n^{(k)} \Rightarrow \Phi^{(k,p)} \cup B_{k,p-1} \text{ in } \mathcal{F}(\Delta), \quad (3.8)$$

where $\Rightarrow$ denotes weak convergence.

The weak limits $\Phi^{(k,p)}$ and $B_{k,m}$ are formally defined below; see Figure 6.

- $\Phi^{(k,p)}$ – a (finite) random closed set characterized as a *Poisson random measure* on Δ , whose mean measure is given by

$$\mathbb{E}(|\Phi^{(k,p)} \cap A|) := \frac{s_{d-1}}{p!(\alpha p - d)} \int_{(\mathbb{R}^d)^{p-1}} g_1(0, \mathbf{y}) \mu_{(0,\mathbf{y})}^{(k)}(A) \, d\mathbf{y}, \quad A \subset \Delta, \quad (3.9)$$

where s_{d-1} is the volume of the $(d-1)$ -dimensional unit sphere in $\mathbb{R}^d$, $\mathbf{y} = (y_1, \dots, y_{p-1}) \in (\mathbb{R}^d)^{p-1}$, $(0, \mathbf{y}) = (0, y_1, \dots, y_{p-1}) \in (\mathbb{R}^d)^p$, and so, $g_1(0, \mathbf{y}) = g_1(0, y_1, \dots, y_{p-1})$.

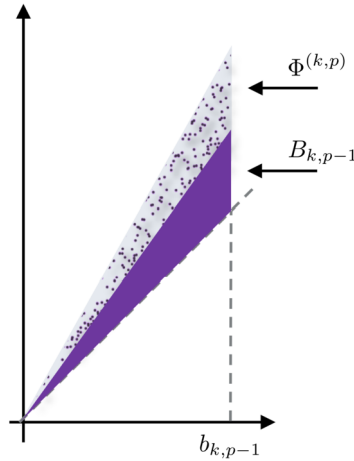


Figure 6. A cartoon for the limiting k th persistence diagram. The solid area represents $B_{k,p-1}$, where the persistence pairs are dense enough, so that the entire region is covered. The grey area is $B_{k,p} \setminus B_{k,p-1}$, where only a finite number of Poisson points, denoted $\Phi^{(k,p)}$, are scattered in the limit.

- $B_{k,m}$ – a nonrandom closed set of Δ defined as follows. Recall that for a given subsets of size m , the k th persistence pairs (γ_b, γ_d) are limited to the region $\Delta_{k,m} = \{(x, y) : 0 \leq x \leq y \leq \pi_{k,m}x\} \subset \Delta$. Next, define

$$b_{k,m} := \sup\{\gamma_b : (\gamma_b, \gamma_d) \in \text{PH}_k(\mathcal{Y}), |\mathcal{Y}| = m, \mathcal{C}_1(\mathcal{Y}) \text{ is connected}\},$$

that is, $b_{k,m}$ is the largest birth time for persistence k -cycles that are generated on m vertices and connected at unit radius. Finally, define

$$B_{k,m} := \Delta_{k,m} \cap ([0, b_{k,m}] \times \mathbb{R}_+).$$

In other words, $B_{k,m}$ is the area in which the k th persistence pairs generated by subsets of m points that are connected at unit radius, may appear. Note that $B_{k,m}$ is increasing in m , that is, $B_{k,m_1} \subset B_{k,m_2}$ for all $m_1 \leq m_2$. In particular, only $B_{k,p-1}$ will contribute to the limiting persistence diagram in (3.8); see Figure 6.

Let us provide some intuition behind our theorem. As detailed in Lemmas 5.1 and 5.2, we can show that for a measurable set $A \subset \Delta$ with $A \cap B_{k,k+2} \neq \emptyset$,

$$\mathbb{E}(|\Phi_n^{(k,p)} \cap A|) \rightarrow \mathbb{E}(|\Phi^{(k,p)} \cap A|) \in (0, \infty), \quad n \rightarrow \infty, \quad (3.10)$$

and as $n \rightarrow \infty$,

$$\mathbb{E}(|\Phi_n^{(k,m)} \cap A|) \rightarrow 0 \quad \text{if } m > p, \quad (3.11)$$

$$\mathbb{E}(|\Phi_n^{(k,m)} \cap A|) \rightarrow \infty \quad \text{if } k+2 \leq m < p. \quad (3.12)$$

Among these three results, the first indicates that there asymptotically exist at most finitely many persistence k -cycles that are generated on p vertices in the complex $\mathcal{C}_M(\mathcal{P}_{n,R})$. Because of the rareness of persistence k -cycles, the set $\Phi_n^{(k,p)}$ will become “Poissonian” in the limit. Namely, as $n \rightarrow \infty$,

$$\Phi_n^{(k,p)} \Rightarrow \Phi^{(k,p)} \text{ in } \mathcal{F}(\Delta). \quad (3.13)$$

Additionally, (3.11) implies that any persistence k -cycles supported on more than p vertices will vanish in the limit. In other words, $\Phi_n^{(k,m)}$ converges to an empty set for all $m > p$, that is,

$$\Phi_n^{(k,m)} \Rightarrow \emptyset \text{ in } \mathcal{F}(\Delta) \quad \text{if } m > p. \quad (3.14)$$

As for the remaining sets $\Phi_n^{(k,m)}$ for $k+2 \leq m < p$, (3.12) implies that there appear infinitely many persistence k -cycles as $n \rightarrow \infty$, that are generated on m vertices in $\mathcal{C}_M(\mathcal{P}_{n,R})$. Accordingly, $\Phi_n^{(k,m)}$ consists of infinitely many k th persistence pairs as $n \rightarrow \infty$, and ultimately, it converges to a deterministic closed set $B_{k,m}$,

$$\Phi_n^{(k,m)} \Rightarrow B_{k,m} \text{ in } \mathcal{F}(\Delta) \quad \text{if } k+2 \leq m < p. \quad (3.15)$$

Finally, combining (3.13), (3.14) and (3.15), along with an increasing property of $B_{k,m}$, one should get that

$$\Phi_n^{(k)} \Rightarrow \Phi^{(k,p)} \cup \bigcup_{m=k+2}^{p-1} B_{k,m} = \Phi^{(k,p)} \cup B_{k,p-1} \text{ in } \mathcal{F}(\Delta).$$

Remark 3.3. Note that (3.6) implicitly rules out a very quick decay of M . If M decays to zero so quickly that $\limsup_{n \rightarrow \infty} n^p M^{d(p-1)} < \infty$, then (3.6) implies that $\liminf_{n \rightarrow \infty} R^d(f(R))^p > 0$, but this contradicts with (3.1). At the same, the condition $M/R \rightarrow 0$ prevents a quick divergence of M . So the limiting behavior of M is controlled on both sides. Moreover, generalizing (3.6) to the case when the limit is a finite and positive constant is easy. It will simply change the mean measure in (3.9) up to constant factors. For simplicity of notation, we shall assume that the limit in (3.6) is 1.

Example 3.4. We consider a simple density with a Pareto tail,

$$f(x) = \frac{C}{1 + \|x\|^\alpha}, \quad (3.16)$$

where C is a normalizing constant. Taking $M \equiv 1$ and solving (3.6) with respect to R , we obtain

$$R = (Cn)^{p/(\alpha p - d)}. \quad (3.17)$$

This sequence grows at a regularly varying rate with index $p/(\alpha p - d)$. Assuming (3.16) together with other conditions in Theorem 3.2, the weak convergence (3.8) holds.

Before concluding this section, we state three corollaries of Theorem 3.2. In the first corollary, we assume that instead of (3.6), R and M satisfy

$$n^p M^{d(p-1)} R^d(f(R))^p \rightarrow 0, \quad n^{p-1} M^{d(p-2)} R^d(f(R))^{p-1} \rightarrow \infty \quad \text{as } n \rightarrow \infty. \quad (3.18)$$

To see the difference between (3.6) and (3.18), we simplify the situation by assuming (3.16) and taking $M \equiv 1$. It is then elementary to show that the R satisfying (3.18) grows faster than the right-hand side of (3.17), that is, $R^{-1}(Cn)^{p/(\alpha p-d)} \rightarrow 0$ as $n \rightarrow \infty$. This means that unlike (3.10), we obtain $\mathbb{E}(|\Phi_n^{(k,p)} \cap A|) \rightarrow 0$ as $n \rightarrow \infty$, in which case the random part $\Phi^{(k,p)}$ vanishes from the limit. A formal statement is given below.

Corollary 3.5. *Suppose that instead of (3.6), R and M satisfy (3.18). Then, as $n \rightarrow \infty$,*

$$\Phi_n^{(k,m)} \Rightarrow \begin{cases} \emptyset \text{ in } \mathcal{F}(\Delta) & \text{if } m \geq p, \\ B_{k,m} \text{ in } \mathcal{F}(\Delta) & \text{if } k+2 \leq m < p, \end{cases}$$

and also,

$$\Phi_n^{(k)} \Rightarrow B_{k,p-1} \text{ in } \mathcal{F}(\Delta).$$

For the second corollary, we again assume the condition at (3.6). We here aim to study the maximal lifespan (i.e., death time – birth time) of persistence k -cycles in the limiting persistence diagram. For the required analyses, we need a continuous functional $T : \mathcal{F}(\Delta) \rightarrow \mathbb{R}_+$ defined by

$$T(F) = \sup_{(x,y) \in F} (y - x). \quad (3.19)$$

This functional captures the maximal vertical distance from the points in $F \subset \Delta$ to the diagonal line. For the remainder of this discussion, fix $t \in (0, b_{k,p-1})$, and define

$$\begin{aligned} I_t &:= \Delta \cap ([0, t] \times \mathbb{R}_+), \\ J_t &:= \{(x, y) \in \Delta_{k,p} \cap I_t : y - x > T(\Delta_{k,p-1} \cap I_t)\}. \end{aligned}$$

In other words, J_t consists of points (x, y) in $\Delta_{k,p} \cap I_t$ such that $y - x$ exceeds the maximal lifespan that can be attained by the points in $\Delta_{k,p-1} \cap I_t$. See Figure 7.

The following corollary describes the limiting behavior of the $T(\Phi_n^{(k)} \cap I_{Mt})$, that is, the maximal lifespan of the persistence k -cycles generated by $\Phi_n^{(k)}$, with the restriction that the birth time is less than Mt . The proof is immediate via continuous mapping theorem. Indeed applying a continuous functional (3.19) to the weak convergence in Theorem 3.2 can yield the required result.

Corollary 3.6. *Under the assumptions of Theorem 3.2, we have*

$$T(\Phi_n^{(k)} \cap I_{Mt}) \Rightarrow Z_t \text{ in } \mathbb{R}_+, \quad n \rightarrow \infty,$$

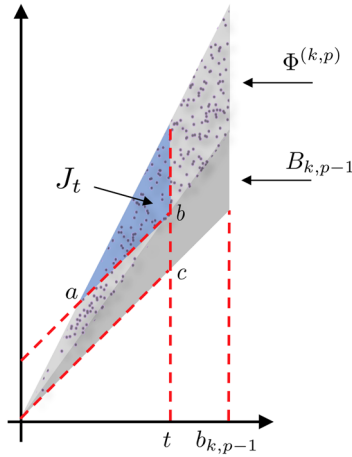


Figure 7. A dashed line segment $[a, b]$ is parallel to the diagonal line. The shaded area is J_t , in which there are at most finitely many Poisson points, representing persistence pairs with longer lifespan. The region $\Delta_{k,p-1} \cap I_t$ is densely covered by persistence pairs with shorter lifespan. If J_t contains points, the maximal lifespan of persistence k -cycles can be attained by one of these points. If J_t does not contain any points, the maximal lifespan is nonrandom and is equal to $T(\Delta_{k,p-1} \cap I_t)$, which is represented by a line segment $[b, c]$.

where

$$Z_t = \begin{cases} T(\Delta_{k,p-1} \cap I_t) & \text{if } \Phi^{(k,p)} \cap J_t = \emptyset, \\ T(\Phi^{(k,p)} \cap J_t) & \text{if } \Phi^{(k,p)} \cap J_t \neq \emptyset. \end{cases}$$

The last statement says that if the limiting Poisson random measure $\Phi^{(k,p)}$ has no points in J_t , the weak limit Z_t takes a purely deterministic value, and the “nonrandom” set $\Delta_{k,p-1} \cap I_t$ yields the maximal lifespan. On the other hand, if $\Phi^{(k,p)}$ has at least one points in J_t , the corresponding lifespan is necessarily longer than $T(\Delta_{k,p-1} \cap I_t)$. Then the actual value of Z_t is random.

For the third corollary, recall that $\bigcup_{m=p}^{\infty} \Phi_n^{(k,m)}$ asymptotically consists of the k th persistence pairs that are generated by a finite number of components of size p . Since the number of k th persistence pairs is almost surely finite, the weak convergence can be reformulated as that in the space $\text{MP}(\Delta)$ of locally finite counting measures on Δ . We here equip $\text{MP}(\Delta)$ with the vague topology (see [34]).

Corollary 3.7. *Under the conditions of Theorem 3.2, we take $\Phi_n^{(k,m)}$ as a point process. Then*

$$\sum_{m=p}^{\infty} \Phi_n^{(k,m)} \Rightarrow \Phi^{(k,p)} \text{ in } \text{MP}(\Delta).$$

4. Exponentially decaying tails case

In the present section, we wish to study the case where the distribution generating random points has an exponentially decaying tail. The results in this case are parallel to those of the previous section except for the normalization and limiting distributions.

To define the density function, we use the *von Mises function*. The following setup is somewhat typical in EVT; see [4,6,30,32].

Definition 4.1 (von Mises function). We say that $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a *von Mises function*, if ψ is C^2 , $\psi'(z) > 0$, and

$$\lim_{z \rightarrow \infty} \psi(z) = \infty, \quad \lim_{z \rightarrow \infty} \frac{d}{dz} \left(\frac{1}{\psi'(z)} \right) = 0.$$

In this section, we study density functions $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ of the form

$$f(x) = L(\|x\|) e^{-\psi(\|x\|)}. \quad (4.1)$$

Let $a(z) := 1/\psi'(z)$, then from Definition 4.1 we have that $a'(z) \rightarrow 0$, $z \rightarrow \infty$. Therefore,

$$\lim_{z \rightarrow \infty} \frac{a(z)}{z} = 0. \quad (4.2)$$

We assume that $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is *flat* for a , that is,

$$\lim_{t \rightarrow \infty} \frac{L(t + a(t)v)}{L(t)} = 1, \quad (4.3)$$

uniformly on $v \in [-K, K]$ for every $K > 0$. Furthermore, we assume that for some $\gamma \geq 0$, $z_0 > 0$ and $C \geq 1$, we have

$$\frac{L(zt)}{L(z)} \leq Ct^\gamma \quad \text{for all } t > 1, z \geq z_0. \quad (4.4)$$

Condition (4.3) together with (4.1) implies that the tail of f is determined by the function ψ (or equivalently a), and is independent of L . Thus, we can classify f in terms of the asymptotics of a . If $a(z)$ converges to a positive, finite constant as $z \rightarrow \infty$, we say that f has an (asymptotic) exponential tail. If $a(z)$ diverges as $z \rightarrow \infty$ we say that f has a *subexponential* tail, and finally, if $a(z) \rightarrow 0$, we say that f has a *superexponential* tail.

As in the previous section, we need to choose a radius $R = R(n)$ and connectivity value $M = M(n)$, for topological crackle to occur. In [32], it was shown that the occurrence of topological crackle depends on the limit

$$c := \lim_{n \rightarrow \infty} \frac{a(R)}{M}.$$

In particular, if $c = 0$, topological crackle never occurs, and random points are densely scattered near the origin, so that placing unit balls around the points constitutes a topologically contractible

object called *core*; see [1]. Since the main focus of the present work is topological crackle, we do not treat the case $c = 0$ and always assume $c \in (0, \infty]$. By definition, if M is a positive constant and $c \in (0, \infty]$, then f never has a superexponential tail.

We now describe a series of results analogous to those in the previous section. The proof is presented in Section 5.3.

Theorem 4.2. *Let f be a probability density function of the form (4.1), and suppose that $R = R(n)$, $M = M(n)$ are chosen such that $R \rightarrow \infty$, $M/R \rightarrow 0$, $a(R)/M \rightarrow c \in (0, \infty]$ as $n \rightarrow \infty$. Assume that there exists an integer $p \geq k + 2$ such that*

$$n^p M^{d(p-1)} a(R) R^{d-1} (f(R))^p \rightarrow 1, \quad n \rightarrow \infty. \quad (4.5)$$

Then, as $n \rightarrow \infty$,

$$\Phi_n^{(k,m)} \Rightarrow \begin{cases} \emptyset \text{ in } \mathcal{F}(\Delta) & \text{if } m > p, \\ \Phi^{(k,p)} \text{ in } \mathcal{F}(\Delta) & \text{if } m = p, \\ B_{k,m} \text{ in } \mathcal{F}(\Delta) & \text{if } k + 2 \leq m < p, \end{cases}$$

and further,

$$\Phi_n^{(k)} \Rightarrow \Phi^{(k,p)} \cup B_{k,p-1} \text{ in } \mathcal{F}(\Delta),$$

where $\Phi^{(k,p)}$ is defined below, and $B_{k,m}$ is the same nonrandom set as in Theorem 3.2.

Similar to Theorem 3.2, the limit $\Phi^{(k,p)}$ above is a (finite) random closed set characterized as a Poisson random measure. Here, the mean measure of $\Phi^{(k,p)}$ is given by

$$\begin{aligned} & \frac{1}{p!} \int_0^\infty d\rho \int_{S^{d-1}} J(\theta) d\theta \int_{(\mathbb{R}^d)^{p-1}} d\mathbf{y} g_1(0, \mathbf{y}) \mu_{(0, \mathbf{y})}^{(k)}(\cdot) \\ & \times e^{-p\rho - c^{-1} \sum_{i=1}^{p-1} \langle \theta, y_i \rangle} \mathbb{1}_{\{\rho + c^{-1} \langle \theta, y_i \rangle \geq 0, i = 1, \dots, p-1\}}, \end{aligned} \quad (4.6)$$

where $\langle \cdot, \cdot \rangle$ denotes scalar product and

$$J(\theta) = \sin^{d-2}(\theta_1) \sin^{d-3}(\theta_2) \cdots \sin(\theta_{d-2}) \quad (4.7)$$

is the Jacobian. Interestingly, if $c = \infty$, (4.6) coincides with (3.9) up to multiplicative constants, implying that the two limiting Poisson random measures coincide regardless of heaviness of the tail of an underlying distribution. As in the heavy tail case, generalizing the limit value in (4.5) is straightforward; see Remark 3.3.

Notice that the main difference between (3.6) and (4.5) lies only in the growth rate of R . To see this, take $M \equiv 1$, and consider the simple example

$$f(x) = C e^{-\|x\|^\tau / \tau}, \quad 0 < \tau \leq 1.$$

Then $a(z) = z^{1-\tau}$, and the solution to (4.5) is given by

$$R = (\tau \log n + p^{-1}(d - \tau) \log(\tau \log n) + \tau \log C)^{1/\tau},$$

which grows logarithmically, whereas, as seen in Example 3.4, the R in the heavy tail setup grows at a regularly varying rate.

We now present the statements equivalent to those in Corollaries 3.5, 3.6 and 3.7.

Corollary 4.3. *Suppose that instead of (4.5), R and M satisfy*

$$\begin{aligned} n^p M^{d(p-1)} a(R) R^{d-1} (f(R))^p &\rightarrow 0, \quad n \rightarrow \infty, \\ n^{p-1} M^{d(p-2)} a(R) R^{d-1} (f(R))^{p-1} &\rightarrow \infty, \quad n \rightarrow \infty. \end{aligned}$$

Then, as $n \rightarrow \infty$,

$$\Phi_n^{(k,m)} \Rightarrow \begin{cases} \emptyset \text{ in } \mathcal{F}(\Delta) & \text{if } m \geq p, \\ B_{k,m} \text{ in } \mathcal{F}(\Delta) & \text{if } k+2 \leq m < p, \end{cases}$$

and also,

$$\Phi_n^{(k)} \Rightarrow B_{k,p-1} \text{ in } \mathcal{F}(\Delta).$$

Corollary 4.4. *Under the assumptions of Theorem 4.2, we have, as $n \rightarrow \infty$,*

$$T(\Phi_n^{(k)} \cap I_{M_t}) \Rightarrow Z_t \text{ in } \mathbb{R}_+, \quad n \rightarrow \infty,$$

where Z_t is given by

$$Z_t = \begin{cases} T(\Delta_{k,p-1} \cap I_t) & \text{if } \Phi^{(k,p)} \cap J_t = \emptyset, \\ T(\Phi^{(k,p)} \cap J_t) & \text{if } \Phi^{(k,p)} \cap J_t \neq \emptyset. \end{cases}$$

Corollary 4.5. *Under the assumptions of Theorem 4.2, we take $\Phi_n^{(k,m)}$ as a point process. Then*

$$\sum_{m=p}^{\infty} \Phi_n^{(k,m)} \Rightarrow \Phi^{(k,p)} \text{ in } \text{MP}(\Delta).$$

5. Proofs

In this section, we provide the proofs for all the statements in this paper. We split the proofs between the regularly varying and the exponentially decaying tail cases.

5.1. Some notation

The following notation will be used throughout the proofs. For $x \in \mathbb{R}^d$, $\mathbf{y} = (y_1, \dots, y_m) \in (\mathbb{R}^d)^m$, and $r > 0$, define

$$x + r\mathbf{y} := (x + ry_1, \dots, x + ry_m) \in (\mathbb{R}^d)^m.$$

The proofs will involve calculating certain volumes, which we define next. Let

$$\mathcal{B}_r(\mathbf{x}) := \bigcup_{i=1}^m B(x_i; r), \quad \mathbf{x} = (x_1, \dots, x_m) \in (\mathbb{R}^d)^m,$$

be a union of m closed balls of radius r , and let

$$Q_r(\mathbf{x}) := \int_{\mathcal{B}_r(\mathbf{x})} f(z) \, dz, \quad (5.1)$$

be the probability measure of the given union of balls.

We denote by λ_m the Lebesgue measure on $\mathbb{R}^m$. Finally, the notation C^* will represent a generic positive constant, which does not depend on n and may vary between (or even within) the lines.

5.2. Regularly varying tails

Our main goal in this section is to prove the results for the regularly varying tail case. We do not present the proof of Corollary 3.5, since it is very similar to that for Theorem 3.2. Further, the proof of Corollary 3.6 will be skipped, because the statement is nearly obvious. We will use the following auxiliary point process. Recalling the definitions of a counting measure at (3.3) and h_R, g_M in (3.4), we define

$$\begin{aligned} \tilde{\Phi}_n^{(k,m)}(\cdot) &:= \sum_{\mathcal{Y} \subset \mathcal{P}_n, |\mathcal{Y}|=m} h_R(\mathcal{Y}) g_M(\mathcal{Y}) \mu_{\mathcal{Y}}^{(k)}(M \cdot), \\ \tilde{\Phi}_n^{(k)}(\cdot) &:= \sum_{m=k+2}^{\infty} \tilde{\Phi}_n^{(k,m)}(\cdot). \end{aligned} \quad (5.2)$$

The only difference between $\Phi_n^{(k,m)}$ and $\tilde{\Phi}_n^{(k,m)}$ is that the latter does not require the subsets $\mathcal{Y}$ to form a connected component of $\mathcal{C}_M(\mathcal{P}_{n,R})$, that is, $\mathcal{Y}$ does not need to be isolated from the rest of the complex. Consequently, we have $\tilde{\Phi}_n^{(k,m)}(\cdot) \geq \Phi_n^{(k,m)}(\cdot)$. As in the case of (3.5), we may and will denote by (5.2) the corresponding random closed sets. Indeed the proof below uses (3.5) and (5.2) as random closed sets only, except for the argument for Corollary 3.7.

Since the proof of Theorem 3.2 is rather long, we shall outline the main idea of its proof. We start with two lemmas (i.e., Lemmas 5.1 and 5.2 below) to evaluate asymptotic first and second moments associated to $\Phi_n^{(k,p)}$ and $\Phi_n^{(k,p-1)}$. We divide the proof of (3.8) into three parts as below. The arguments in Part I and Part II will immediately imply (3.7) as well.

Part I – Prove the “random” part of the limit, that is,

$$\bigcup_{m=p}^{\infty} \Phi_n^{(k,m)} \Rightarrow \Phi^{(k,p)} \text{ in } \mathcal{F}(\Delta). \quad (5.3)$$

Part II – Prove the “nonrandom” part of the limit, that is,

$$\bigcup_{m=k+2}^{p-1} \Phi_n^{(k,m)} \Rightarrow B_{k,p-1} \text{ in } \mathcal{F}(\Delta). \quad (5.4)$$

Part III – Combine I and II to conclude the statement in the theorem,

$$\Phi_n^{(k)} \Rightarrow \Phi^{(k,p)} \cup B_{k,p-1} \text{ in } \mathcal{F}(\Delta). \quad (5.5)$$

In particular, the argument in Part I is closely related to the proof of [32], the main tool of which is *Stein’s Poisson approximation theorem* (e.g., Theorem 2.1 in [33]). The arguments in Part II and Part III exploit some of the basic properties of the Fell topology. In particular, the deterministic set $B_{k,p-1}$ is handled by the second moment result in Lemma 5.2.

Lemma 5.1. *Let $A \subset \Delta$ be a measurable set, such that $A \cap B_{k,p} \neq \emptyset$. Under the assumptions of Theorem 3.2,*

$$\lim_{n \rightarrow \infty} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) = \lim_{n \rightarrow \infty} \mathbb{E}(|\tilde{\Phi}_n^{(k,p)} \cap A|) = \mathbb{E}(|\Phi^{(k,p)} \cap A|) \in (0, \infty).$$

Lemma 5.2. *Let $A \subset \Delta$ be a measurable set, with $A \cap B_{k,p-1} \neq \emptyset$. Under the assumptions of Theorem 3.2,*

$$\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|) \sim C_1 (nM^d f(R))^{-1}, \quad n \rightarrow \infty,$$

and

$$\text{Var}(|\Phi_n^{(k,p-1)} \cap A|) \leq C_2 (nM^d f(R))^{-1},$$

where C_1 and C_2 are positive constants, independent of n and depending only on d, k, p, A , and a probability density f .

Next, we prove Lemma 5.1, while the proof for Lemma 5.2 is of a similar nature, and postponed to the [Appendix](#).

Proof of Lemma 5.1. We will prove the limit for $\Phi_n^{(k,p)}$ only, since the limit for $\tilde{\Phi}_n^{(k,p)}$ can be proved in the same way. It follows from the Palm theory for Poisson processes (e.g., Section 1.7 in [33]) that

$$\mathbb{E}(|\Phi_n^{(k,p)} \cap A|) = \frac{n^p}{p!} \mathbb{E}[h_R(\mathcal{X}_p) g_M(\mathcal{X}_p, \mathcal{X}_p \cup \mathcal{P}_n) \mu_{\mathcal{X}_p}^{(k)}(MA)], \quad (5.6)$$

where $\mathcal{X}_p = (X_1, \dots, X_p)$ is a set of p i.i.d. points with probability density f , and independent of $\mathcal{P}_n$. Note that for the set $\mathcal{X}_p$ to be disconnected from the rest of the complex $\mathcal{C}_M(\mathcal{P}_n, R)$, we

require that $\mathcal{P}_n \cap \mathcal{B}_{2M}(\mathcal{X}_p) = \emptyset$. Therefore, by the conditioning on $\mathcal{X}_p$ we have

$$\begin{aligned} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) &= \frac{n^p}{p!} \mathbb{E}[h_R(\mathcal{X}_p) g_M(\mathcal{X}_p) \mu_{\mathcal{X}_p}^{(k)}(MA) \mathbb{P}(\mathcal{P}_n \cap \mathcal{B}_{2M}(\mathcal{X}_p) = \emptyset | \mathcal{X}_p)] \\ &= \frac{n^p}{p!} \mathbb{E}[h_R(\mathcal{X}_p) g_M(\mathcal{X}_p) \mu_{\mathcal{X}_p}^{(k)}(MA) e^{-nQ_{2M}(\mathcal{X}_p)}] \\ &= \frac{n^p}{p!} \int_{(\mathbb{R}^d)^p} h_R(\mathbf{x}) g_M(\mathbf{x}) \mu_{\mathbf{x}}^{(k)}(MA) e^{-nQ_{2M}(\mathbf{x})} \prod_{i=1}^p f(x_i) d\mathbf{x}. \end{aligned}$$

Performing the change of variables $x_1 \leftrightarrow x$, $x_i \leftrightarrow x + My_{i-1}$, $i = 2, \dots, p$, we have

$$\begin{aligned} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) &= \frac{n^p}{p!} M^{d(p-1)} \int_{\mathbb{R}^d} dx \int_{(\mathbb{R}^d)^{p-1}} d\mathbf{y} h_R(x, x + M\mathbf{y}) g_M(x, x + M\mathbf{y}) \\ &\quad \times \mu_{(x, x+M\mathbf{y})}^{(k)}(MA) e^{-nQ_{2M}(x, x+M\mathbf{y})} f(x) \prod_{i=1}^{p-1} f(x + My_i) \\ &= \frac{n^p}{p!} M^{d(p-1)} \int_{\mathbb{R}^d} dx \int_{(\mathbb{R}^d)^{p-1}} d\mathbf{y} h_R(x, x + M\mathbf{y}) g_1(0, \mathbf{y}) \mu_{(0, \mathbf{y})}^{(k)}(A) \\ &\quad \times e^{-nQ_{2M}(x, x+M\mathbf{y})} f(x) \prod_{i=1}^{p-1} f(x + My_i), \end{aligned}$$

where the second equality follows from the translation invariance and scaling properties of g_M and $\mu^{(k)}$. Next, we apply a polar coordinate transform $x \leftrightarrow (r, \theta)$ where $r \in [0, \infty)$ and $\theta \in S^{d-1}$, which is followed by another change of variable $r \leftrightarrow R\rho$. Notice also that

$$h_R(R\rho\theta, R\rho\theta + M\mathbf{y}) = \mathbb{1}\{\rho \geq 1\} h_1(\rho\theta + M\mathbf{y}/R).$$

Combining all of these together, we obtain

$$\begin{aligned} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) &= \frac{n^p}{p!} M^{d(p-1)} R^d (f(R))^p \int_1^\infty \rho^{d-1} d\rho \int_{S^{d-1}} J(\theta) d\theta \int_{(\mathbb{R}^d)^{p-1}} d\mathbf{y} \\ &\quad \times h_1(\rho\theta + M\mathbf{y}/R) g_1(0, \mathbf{y}) \mu_{(0, \mathbf{y})}^{(k)}(A) \\ &\quad \times e^{-nQ_{2M}(R\rho\theta, R\rho\theta + M\mathbf{y})} \frac{f(R\rho)}{f(R)} \prod_{i=1}^{p-1} \frac{f(R\|\rho\theta + M\mathbf{y}_i/R\|)}{f(R)}, \end{aligned} \quad (5.7)$$

where $J(\theta)$ is the Jacobian given by (4.7).

Our next goal is to find the limit of the individual terms inside the integral. First, notice that since $M/R \rightarrow 0$ we have that $h_1(\rho\theta + M\mathbf{y}/R) \rightarrow 1$ for all $\rho \geq 1$, $\theta \in S^{d-1}$, and

$\mathbf{y} = (y_1, \dots, y_{p-1}) \in (\mathbb{R}^d)^{p-1}$. Next, appealing to the regular variation of f in (3.1), we have

$$\frac{f(R\rho)}{f(R)} \prod_{i=1}^{p-1} \frac{f(R\|\rho\theta + My_i/R\|)}{f(R)} \rightarrow \rho^{-\alpha p}$$

for all $\rho \geq 1$, $\theta \in S^{d-1}$, and $\mathbf{y} \in (\mathbb{R}^d)^{p-1}$.

Finally, we verify that the exponential term in (5.7) converges to one. To evaluate Q_{2M} , we apply the change of variable $z \leftrightarrow R\rho\theta + Mv$ in (5.1). This yields

$$\begin{aligned} nQ_{2M}(R\rho\theta, R\rho\theta + M\mathbf{y}) &= nM^d f(R) \int_{\mathcal{B}_2(0, \mathbf{y})} \frac{f(R\|\rho\theta + Mv/R\|)}{f(R)} dv \\ &\leq nM^d f(R) \sup_{v \in \mathcal{B}_2(0, \mathbf{y})} \frac{f(R\|\rho\theta + Mv/R\|)}{f(R)} \lambda_d(\mathcal{B}_2(0, \mathbf{y})). \end{aligned}$$

Observe that for all $\rho \geq 1$, $\theta \in S^{d-1}$, and $v \in \mathcal{B}_2(0, \mathbf{y})$ such that $\mathcal{C}_1(0, \mathbf{y})$ is connected, we have, for large enough n ,

$$\|\rho\theta + Mv/R\| \geq \frac{\rho}{2} \geq \frac{1}{2}.$$

Therefore, the Potter bound for regularly varying functions (e.g., Theorem 1.5.6 in [7] or Proposition 2.6 in [35]) gives, for every $0 < \zeta < \alpha - d$,

$$\begin{aligned} \sup_{v \in \mathcal{B}_2(0, \mathbf{y})} \frac{f(R\|\rho\theta + Mv/R\|)}{f(R)} &\leq C^* \sup_{v \in \mathcal{B}_2(0, \mathbf{y})} \max\{\|\rho\theta + Mv/R\|^{-\alpha+\zeta}, \|\rho\theta + Mv/R\|^{-\alpha-\zeta}\} \\ &\leq C^*. \end{aligned}$$

Thus, for all $\rho, \theta, \mathbf{y}$ we have

$$nQ_{2M}(R\rho\theta, R\rho\theta + M\mathbf{y}) \leq C^* nM^d f(R).$$

Recalling (3.6), together with the assumption $M/R \rightarrow 0$, ensures that $nM^d f(R) \rightarrow 0$, from which we can conclude that $e^{-nQ_{2M}(R\rho\theta, R\rho\theta + M\mathbf{y})} \rightarrow 1$.

Assuming that the dominated convergence theorem applies (as justified next), while using (3.6), we can conclude that

$$\begin{aligned} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) &\rightarrow \frac{1}{p!} \int_1^\infty \rho^{d-1-\alpha p} d\rho \int_{S^{d-1}} J(\theta) d\theta \int_{(\mathbb{R}^d)^{p-1}} g_1(0, \mathbf{y}) \mu_{(0, \mathbf{y})}^{(k)}(A) d\mathbf{y} \\ &= \frac{S_{d-1}}{p!(\alpha p - d)} \int_{(\mathbb{R}^d)^{p-1}} g_1(0, \mathbf{y}) \mu_{(0, \mathbf{y})}^{(k)}(A) d\mathbf{y} \\ &= \mathbb{E}(\Phi^{(k,p)}(A)), \quad n \rightarrow \infty. \end{aligned}$$

It now remains to establish an integrable upper bound for an integrand in (5.7), in order to apply the dominated convergence theorem. First, the exponential term in (5.7) is obviously bounded by one. As for the ratio of the densities, applying the Potter bound repeatedly we derive that, for

every $0 < \zeta < \alpha - d$, there exists a $C > 0$ (we have introduced a specific constant C , not a generic one, for later use) such that, for sufficiently large n ,

$$\frac{f(R\rho)}{f(R)} \mathbb{1}\{\rho \geq 1\} \leq C \max\{\rho^{-\alpha+\zeta}, \rho^{-\alpha-\zeta}\} \mathbb{1}\{\rho \geq 1\} = C\rho^{-\alpha+\zeta} \mathbb{1}\{\rho \geq 1\}, \quad (5.8)$$

and for each $i = 1, \dots, p-1$,

$$\begin{aligned} & \frac{f(R\|\rho\theta + My_i/R\|)}{f(R)} \mathbb{1}\{\|\rho\theta + My_i/R\| \geq 1\} \\ & \leq C \max\{\|\rho\theta + My_i/R\|^{-\alpha+\zeta}, \|\rho\theta + My_i/R\|^{-\alpha-\zeta}\} \mathbb{1}\{\|\rho\theta + My_i/R\| \geq 1\} \leq C. \end{aligned} \quad (5.9)$$

Since

$$\int_1^\infty \rho^{d-1-\alpha+\zeta} d\rho \int_{(\mathbb{R}^d)^{p-1}} g_1(0, \mathbf{y}) \mu_{(0, \mathbf{y})}^{(k)}(A) d\mathbf{y} < \infty,$$

we now obtain the required integrable bound. $\square$

Proof of Theorem 3.2. Below we present the proofs of (5.3), (5.4) and (5.5) in Parts I–III.

Proof of (5.3) in Part I:

By virtue of Theorem 6.5 in [27], it is enough to verify that for every compact subset $A \subset \Delta$ with $A \cap B_{k,p} \neq \emptyset$, we have

$$\mathbb{P}\left(\bigcup_{m=p}^\infty \Phi_n^{(k,m)} \cap A \neq \emptyset\right) \rightarrow \mathbb{P}(\Phi^{(k,p)} \cap A \neq \emptyset).$$

We can proceed as follows:

$$\begin{aligned} & \left| \mathbb{P}\left(\bigcup_{m=p}^\infty \Phi_n^{(k,m)} \cap A \neq \emptyset\right) - \mathbb{P}(\Phi^{(k,p)} \cap A \neq \emptyset) \right| \\ & \leq |\mathbb{P}(\Phi_n^{(k,p)} \cap A = \emptyset) - \mathbb{P}(\Phi^{(k,p)} \cap A = \emptyset)| + \mathbb{P}\left(\bigcup_{m=p+1}^\infty \{\Phi_n^{(k,m)} \cap A \neq \emptyset\}\right) \\ & \leq |\mathbb{P}(\tilde{\Phi}_n^{(k,p)} \cap A = \emptyset) - \mathbb{P}(\Phi^{(k,p)} \cap A = \emptyset)| + \mathbb{P}((\tilde{\Phi}_n^{(k,p)} \setminus \Phi_n^{(k,p)}) \cap A \neq \emptyset) \\ & \quad + \mathbb{P}\left(\bigcup_{m=p+1}^\infty \{\Phi_n^{(k,m)} \cap A \neq \emptyset\}\right) \\ & \leq |\mathbb{P}(\tilde{\Phi}_n^{(k,p)} \cap A = \emptyset) - \mathbb{P}(\Phi^{(k,p)} \cap A = \emptyset)| + \mathbb{E}[|\tilde{\Phi}_n^{(k,p)} \cap A| - |\Phi_n^{(k,p)} \cap A|] \\ & \quad + \sum_{m=p+1}^\infty \mathbb{E}(|\Phi_n^{(k,m)} \cap A|), \\ & =: T_1 + T_2 + T_3, \end{aligned} \quad (5.10)$$

where the last step follows from Markov's inequality. To complete the proof, we thus need to show that $T_i \rightarrow 0$ for $i = 1, 2, 3$. First, $T_2 \rightarrow 0$ follows as a direct consequence of Lemma 5.1.

Next, we show that $T_1 \rightarrow 0$. To this end, we introduce an i.i.d. random sample version of $\tilde{\Phi}_n^{(k,p)}$. More specifically, let

$$\mathcal{I}_n := \begin{cases} \{(i_1, \dots, i_p) \in \mathbb{N}_+^p : 1 \leq i_1 < \dots < i_p \leq n\} & \text{if } n \geq p, \\ \emptyset & \text{if } n < p, \end{cases}$$

and

$$\eta_{\mathbf{i},n} := h_R(\mathcal{X}_{\mathbf{i}})g_M(\mathcal{X}_{\mathbf{i}})\mu_{\mathcal{X}_{\mathbf{i}}}^{(k)}(MA),$$

where $\mathcal{X}_{\mathbf{i}} = (X_{i_1}, \dots, X_{i_p})$ for $\mathbf{i} = (i_1, \dots, i_p) \in \mathcal{I}_n$. Notice that the X_i 's here are the same as those in (3.2) to generate $\mathcal{P}_n$. In other words, $\eta_{\mathbf{i},n}$ is the total number of the k th persistence pairs lying in MA and generated by the subset $\mathcal{X}_{\mathbf{i}}$, with the restriction that $\mathcal{X}_{\mathbf{i}}$ is connected and each point in $\mathcal{X}_{\mathbf{i}}$ lies outside $B(0; R)$. We now claim that

$$\mathbb{P}\left(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n} = 0\right) - \mathbb{P}(\tilde{\Phi}_n^{(k,p)} \cap A = \emptyset) \rightarrow 0, \quad n \rightarrow \infty.$$

For any integer-valued random variables Y_1 and Y_2 defined on the same probability space we have

$$|\mathbb{P}(Y_1 = 0) - \mathbb{P}(Y_2 = 0)| \leq \mathbb{E}(|Y_1 - Y_2|).$$

Therefore,

$$\begin{aligned} \left| \mathbb{P}\left(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n} = 0\right) - \mathbb{P}(\tilde{\Phi}_n^{(k,p)} \cap A = \emptyset) \right| &\leq \mathbb{E}\left[\left|\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n} - |\tilde{\Phi}_n^{(k,p)} \cap A|\right|\right] \\ &= \sum_{m=0}^{\infty} \mathbb{P}(|\mathcal{P}_n| = m) \mathbb{E}\left[\left|\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n} - \sum_{\mathbf{i} \in \mathcal{I}_m} \eta_{\mathbf{i},n}\right|\right] \\ &= \sum_{m=0}^{\infty} \mathbb{P}(|\mathcal{P}_n| = m) \left| \binom{n}{p} - \binom{m}{p} \right| \mathbb{E}(\eta_{\mathbf{i},n}), \quad (5.11) \end{aligned}$$

where $\mathbf{i}$ in the last line is an arbitrary element of $\mathcal{I}_n$. Returning to (5.6), we find that the expectation portion of the right-hand side in (5.6) is asymptotically equal to $\mathbb{E}(\eta_{\mathbf{i},n})$. Additionally, Lemma 5.1 ensures that $\mathbb{E}(|\Phi_n^{(k,p)} \cap A|)$ tends to a positive constant as $n \rightarrow \infty$. Hence the right-most term at (5.11) can be bounded by

$$C^* \sum_{m=0}^{\infty} \mathbb{P}(|\mathcal{P}_n| = m) n^{-p} \left| \binom{n}{p} - \binom{m}{p} \right|, \quad (5.12)$$

which itself is further bounded by

$$C^* \left\{ n^{-p} \left| \binom{n}{p} - \frac{n^p}{p!} \right| + \frac{1}{p!} \mathbb{E} \left[\left| \left(\frac{|\mathcal{P}_n|}{n} \right)^p - 1 \right| \right] + n^{-p} \sum_{m=0}^{\infty} \mathbb{P}(|\mathcal{P}_n| = m) \left| \binom{m}{p} - \frac{m^p}{p!} \right| \right\}.$$

It is now straightforward to show that each of the three terms converges to 0 as $n \rightarrow \infty$.

To prove $T_1 \rightarrow 0$, it now suffices to show that

$$\mathbb{P} \left(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n} = 0 \right) - \mathbb{P}(\Phi^{(k,p)} \cap A = \emptyset) \rightarrow 0, \quad n \rightarrow \infty. \quad (5.13)$$

To this end, our argument relies on the so-called *total variation distance*, which is defined for two random variables Y_1, Y_2 as

$$d_{\text{TV}}(Y_1, Y_2) := \sup_{A \subset \mathbb{R}} |\mathbb{P}(Y_1 \in A) - \mathbb{P}(Y_2 \in A)|.$$

Denoting $Z \sim \text{Poisson}(\mathbb{E}(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n}))$, and using the triangle inequality, we have

$$\begin{aligned} & \left| \mathbb{P} \left(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n} = 0 \right) - \mathbb{P}(\Phi^{(k,p)} \cap A = \emptyset) \right| \\ & \leq d_{\text{TV}} \left(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n}, Z \right) + |\mathbb{P}(Z = 0) - \mathbb{P}(\Phi^{(k,p)} \cap A = \emptyset)|. \end{aligned}$$

Since Z and $\Phi^{(k,p)}(A)$ are both Poisson, an elementary calculation shows that

$$\begin{aligned} |\mathbb{P}(Z = 0) - \mathbb{P}(\Phi^{(k,p)} \cap A = \emptyset)| & \leq \left| \mathbb{E} \left(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n} \right) - \mathbb{E}(|\Phi^{(k,p)} \cap A|) \right| \\ & \leq |\mathbb{E}(|\tilde{\Phi}_n^{(k,p)} \cap A|) - \mathbb{E}(|\Phi^{(k,p)} \cap A|)| + o(1) \rightarrow 0, \end{aligned}$$

where we have used (5.11) and Lemma 5.1.

In order to bound $d_{\text{TV}}(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n}, Z)$ we will use Stein–Chen Poisson approximation method. As preparation, however, we need to define a certain graph on $\mathcal{I}_n$ as follows. For $\mathbf{i}, \mathbf{j} \in \mathcal{I}_n$, write $\mathbf{i} \sim \mathbf{j}$ if and only if they have at least one common element, that is, $|\mathbf{i} \cap \mathbf{j}| > 0$. Then $(\mathcal{I}_n, \sim)$ constitutes a *dependency graph*, that is, for every $I_1, I_2 \subset \mathcal{I}_n$ with no edges connecting I_1 and I_2 , we have that $(\eta_{\mathbf{i},n}, \mathbf{i} \in I_1)$ and $(\eta_{\mathbf{i},n}, \mathbf{i} \in I_2)$ are independent. Under this setup, Stein’s Poisson approximation theorem yields

$$d_{\text{TV}} \left(\sum_{\mathbf{i} \in \mathcal{I}_n} \eta_{\mathbf{i},n}, Z \right) \leq 3 \left[\sum_{\mathbf{i} \in \mathcal{I}_n} \sum_{\mathbf{j} \in N_{\mathbf{i}}} \mathbb{E}(\eta_{\mathbf{i},n}) \mathbb{E}(\eta_{\mathbf{j},n}) + \sum_{\mathbf{i} \in \mathcal{I}_n} \sum_{\mathbf{j} \in N_{\mathbf{i}} \setminus \{\mathbf{i}\}} \mathbb{E}(\eta_{\mathbf{i},n} \eta_{\mathbf{j},n}) \right],$$

where $N_{\mathbf{i}} = \{\mathbf{j} \in \mathcal{I}_n : \mathbf{i} \sim \mathbf{j}\} \cup \{\mathbf{i}\}$.

From the argument before (5.12), we know that for sufficiently large n ,

$$\mathbb{E}(\eta_{\mathbf{i},n}) \leq 2 \binom{n}{p}^{-1} \mathbb{E}(|\Phi^{(k,p)} \cap A|).$$

Therefore, as $n \rightarrow \infty$,

$$\sum_{\mathbf{i} \in \mathcal{I}_n} \sum_{\mathbf{j} \in N_{\mathbf{i}}} \mathbb{E}(\eta_{\mathbf{i},n}) \mathbb{E}(\eta_{\mathbf{j},n}) \leq \binom{n}{p} \left(\binom{n}{p} - \binom{n-p}{p} \right) 4 \binom{n}{p}^{-2} (\mathbb{E}(|\Phi^{(k,p)} \cap A|))^2 \rightarrow 0.$$

For $\mathbf{i}, \mathbf{j} \in \mathcal{I}_n$ with $\ell := |\mathbf{i} \cap \mathbf{j}| \in \{1, \dots, p-1\}$, by the same change of variables as in the proof of Lemma 5.1, we have

$$\begin{aligned} \mathbb{E}(\eta_{\mathbf{i},n} \eta_{\mathbf{j},n}) &= M^{d(2p-\ell-1)} R^d (f(R))^{2p-\ell} \int_1^\infty \rho^{d-1} d\rho \int_{S^{d-1}} J(\theta) d\theta \int_{(\mathbb{R}^d)^{2p-\ell-1}} d\mathbf{y} \\ &\quad \times h_1(\rho\theta + M\mathbf{y}/R) g_1(0, \mathbf{y}) \mu_{(0,\mathbf{y})}^{(k)}(A) \\ &\quad \times e^{-nQ_{2M}(R\rho\theta, R\rho\theta + M\mathbf{y})} \frac{f(R\rho)}{f(R)} \prod_{i=1}^{2p-\ell-1} \frac{f(R\|\rho\theta + My_i/R\|)}{f(R)} \\ &= O(M^{d(2p-\ell-1)} R^d (f(R))^{2p-\ell}). \end{aligned}$$

It follows from (3.6) that

$$\begin{aligned} \sum_{\mathbf{i} \in \mathcal{I}_n} \sum_{\mathbf{j} \in N_{\mathbf{i}} \setminus \{\mathbf{i}\}} \mathbb{E}(\eta_{\mathbf{i},n} \eta_{\mathbf{j},n}) &= \sum_{\ell=1}^{p-1} \binom{n}{p} \binom{p}{\ell} \binom{n-p}{p-\ell} \mathbb{E}(\eta_{\mathbf{i},n} \eta_{\mathbf{j},n}) \mathbb{1}\{|\mathbf{i} \cap \mathbf{j}| = \ell\} \\ &\leq C^* \sum_{\ell=1}^{p-1} n^{2p-\ell} M^{d(2p-\ell-1)} R^d (f(R))^{2p-\ell} \\ &\leq C^* n M^d f(R) \rightarrow 0, \quad n \rightarrow \infty, \end{aligned}$$

and hence, (5.13) is established.

Finally, we turn our attention to showing that $T_3 \rightarrow 0$ in (5.10). For every $m \geq p+1$, repeating the argument of Lemma 5.1,

$$\begin{aligned} \mathbb{E}(|\Phi_n^{(k,m)} \cap A|) &= \frac{n^m}{m!} M^{d(m-1)} R^d (f(R))^m \int_1^\infty \rho^{d-1} d\rho \int_{S^{d-1}} J(\theta) d\theta \int_{(\mathbb{R}^d)^{m-1}} d\mathbf{y} \\ &\quad \times h_1(\rho\theta + M\mathbf{y}/R) g_1(0, \mathbf{y}) \mu_{(0,\mathbf{y})}^{(k)}(A) \\ &\quad \times e^{-nQ_{2M}(R\rho\theta, R\rho\theta + M\mathbf{y})} \frac{f(R\rho)}{f(R)} \prod_{i=1}^{m-1} \frac{f(R\|\rho\theta + My_i/R\|)}{f(R)}. \end{aligned}$$

Using (5.8) and (5.9), together with the fact that for large enough n ,

$$n^m M^{d(m-1)} R^d (f(R))^m \leq 2(nM^d f(R))^{m-p},$$

we have

$$\begin{aligned} \mathbb{E}(|\Phi_n^{(k,m)} \cap A|) &\leq \frac{2C^m}{m!} (nM^d f(R))^{m-p} \\ &\quad \times \int_1^\infty \rho^{d-1-\alpha+\zeta} d\rho \int_{S^{d-1}} J(\theta) d\theta \int_{(\mathbb{R}^d)^{m-1}} g_1(0, \mathbf{y}) \mu_{(0,\mathbf{y})}^{(k)}(A) d\mathbf{y} \\ &\leq \frac{2s_{d-1}}{\alpha-d-\zeta} \frac{C^m}{m!} \binom{m}{k+1} (nM^d f(R))^{m-p} \int_{(\mathbb{R}^d)^{m-1}} g_1(0, \mathbf{y}) d\mathbf{y}, \end{aligned}$$

where at the second step we have applied $\mu_{(0,\mathbf{y})}^{(k)}(A) \leq \binom{m}{k+1}$. Next, the well-known fact that there exist m^{m-2} spanning trees on a set of m vertices, yields

$$\int_{(\mathbb{R}^d)^{m-1}} g_1(0, \mathbf{y}) d\mathbf{y} \leq m^{m-2} \omega_d^{m-1},$$

where ω_d represents the volume of a unit ball in $\mathbb{R}^d$. To show that $T_3 \rightarrow 0$, it therefore remains to verify that

$$\sum_{m=p+1}^\infty \frac{C^m}{m!} \binom{m}{k+1} (nM^d f(R))^{m-p} m^{m-2} \omega_d^{m-1} \rightarrow 0, \quad n \rightarrow \infty.$$

From Stirling's formula (i.e., $m! \geq (m/e)^m$ for large m enough), we can bound the left-hand side above by a constant multiple of

$$\sum_{m=p+1}^\infty m^{k-1} (eC\omega_d n M^d f(R))^{m-p},$$

which clearly vanishes as $n \rightarrow \infty$. We thus proved that $T_i \rightarrow 0$ for $i = 1, 2, 3$ in (5.10), so we can conclude Part I.

Proof of (5.4) in Part II:

Recall that our goal here is to prove the nonrandom part of the limit, that is,

$$\widehat{\Phi}_n^{(k,p)} := \bigcup_{m=k+2}^{p-1} \Phi_n^{(k,m)} \Rightarrow B_{k,p-1} \text{ in } \mathcal{F}(\Delta). \quad (5.14)$$

First, for a measurable set $A \subset \Delta$ and $\epsilon > 0$, denote by $(A)^{\epsilon-}$ an open ϵ -envelop in terms of the Euclidean metric (see (2.1)). By the definition of convergence in probability under the Fell topology (see Definition 6.19 in [27]), we need to show that

$$\mathbb{P}([(\widehat{\Phi}_n^{(k,p)} \setminus (B_{k,p-1})^{\epsilon-}) \cup (B_{k,p-1} \setminus (\widehat{\Phi}_n^{(k,p)})^{\epsilon-})] \cap K \neq \emptyset) \rightarrow 0$$

for every $\epsilon > 0$ and for every compact set K in Δ . Since by construction, $\widehat{\Phi}_n^{(k,p)} \subset B_{k,p-1}$, we have

$$\widehat{\Phi}_n^{(k,p)} \setminus (B_{k,p-1})^{\epsilon-} = \emptyset \quad \text{a.s.}$$

It thus remains to prove that

$$\mathbb{P}([B_{k,p-1} \setminus (\widehat{\Phi}_n^{(k,p)})^{\epsilon-}] \cap K \neq \emptyset) \rightarrow 0, \quad n \rightarrow \infty.$$

Note that $(\widehat{\Phi}_n^{(k,p)})^{\epsilon-}$ is the union of open balls of radius ϵ centered about the points in $\widehat{\Phi}_n^{(k,p)}$. Since K is a closed and bounded set, we can take, without loss of generality, $K = ([a, b] \times \mathbb{R}_+) \cap B_{k,p-1}$ for some $0 \leq a < b < \infty$. Let $\mathcal{Q}_{p-1}$ be a collection of cubes in $\mathbb{R}^2$ with side length $\epsilon/\sqrt{2}$ such that each cube intersects with K and the union of these cubes covers K . Then

$$\begin{aligned} & \mathbb{P}([B_{k,p-1} \setminus (\widehat{\Phi}_n^{(k,p)})^{\epsilon-}] \cap K \neq \emptyset) \\ & \leq \mathbb{P}\left(\bigcup_{Q \in \mathcal{Q}_{p-1}} \bigcap_{m=k+2}^{p-1} \{\Phi_n^{(k,m)} \cap Q = \emptyset\}\right) \\ & \leq \sum_{Q \in \mathcal{Q}_{p-1}} \mathbb{P}(\Phi_n^{(k,p-1)} \cap Q = \emptyset) \\ & \leq \sum_{Q \in \mathcal{Q}_{p-1}} \mathbb{P}(|\Phi_n^{(k,p-1)} \cap Q| - \mathbb{E}(|\Phi_n^{(k,p-1)} \cap Q|)| \geq \mathbb{E}(|\Phi_n^{(k,p-1)} \cap Q|)) \\ & \leq \sum_{Q \in \mathcal{Q}_{p-1}} \frac{\text{Var}(|\Phi_n^{(k,p-1)} \cap Q|)}{[\mathbb{E}(|\Phi_n^{(k,p-1)} \cap Q|)]^2}. \end{aligned}$$

Hence, from Lemma 5.2 we can bound the rightmost term by a constant multiple of $nM^d f(R)$. Since $nM^d f(R) \rightarrow 0$, the desired result follows.

Proof of (5.5) in Part III:

Here, we wish to combine I and II to conclude the statement in the theorem,

$$\Phi_n^{(k)} \Rightarrow \Phi^{(k,p)} \cup B_{k,p-1} \text{ in } \mathcal{F}(\Delta).$$

Since $\mathcal{F}(\Delta)$ is metrizable in the Fell topology (see [2]), (5.14) implies that there exists a metric on $\mathcal{F}(\Delta)$, denoted ρ , such that

$$\rho(\widehat{\Phi}_n^{(k,p)}, B_{k,p-1}) \xrightarrow{p} 0. \quad (5.15)$$

Now, combining the convergences (5.3) and (5.15) gives (see Proposition 3.1 in [35]),

$$\left(\bigcup_{m=p}^{\infty} \Phi_n^{(k,m)}, \widehat{\Phi}_n^{(k,p)}\right) \Rightarrow (\Phi^{(k,p)}, B_{k,p-1}) \text{ in } \mathcal{F}(\Delta) \times \mathcal{F}(\Delta),$$

where $\mathcal{F}(\Delta) \times \mathcal{F}(\Delta)$ is equipped with the product topology. Finally, using the fact that $(F_1, F_2) \in \mathcal{F}(\Delta) \times \mathcal{F}(\Delta) \mapsto F_1 \cup F_2 \in \mathcal{F}(\Delta)$ is continuous (see p. 7 in [26]), we can conclude from the continuous mapping theorem that

$$\bigcup_{m=k+2}^{\infty} \Phi_n^{(k,m)} \Rightarrow \Phi^{(k,p)} \cup B_{k,p-1} \text{ in } \mathcal{F}(\Delta).$$

□

Before finishing this section, we provide a proof for Corollary 3.7.

Proof of Corollary 3.7. To show convergence of $\sum_{m=p}^{\infty} \Phi_n^{(k,m)}$ in $\text{MP}(\Delta)$, we will use Kallenberg's theorem (see Proposition 3.22 in [34]), for which we need to show that for any measurable set $A \subset \Delta$,

$$\begin{aligned} \mathbb{E} \left[\sum_{m=p}^{\infty} \Phi_n^{(k,m)}(A) \right] &\rightarrow \mathbb{E}(\Phi^{(k,p)}(A)), \\ \mathbb{P} \left(\sum_{m=p}^{\infty} \Phi_n^{(k,m)}(A) = 0 \right) &\rightarrow \mathbb{P}(\Phi^{(k,p)}(A) = 0). \end{aligned}$$

The first limit is a direct result of Lemma 5.1 and $T_3 \rightarrow 0$ in (5.10). For the second limit, we have

$$\begin{aligned} \mathbb{P} \left(\sum_{m=p}^{\infty} \Phi_n^{(k,m)}(A) = 0 \right) &= \mathbb{P}(\Phi_n^{(k,p)}(A) = 0) + o(1) \\ &= \mathbb{P}(\tilde{\Phi}_n^{(k,p)}(A) = 0) + o(1) \rightarrow \mathbb{P}(\Phi^{(k,p)}(A) = 0), \quad n \rightarrow \infty, \end{aligned}$$

where we have used $T_i \rightarrow 0$, $i = 1, 2, 3$ in (5.10). □

5.3. Exponentially decaying tails

The proof for the exponentially decaying tail case goes mostly parallel to that in the previous subsection. In particular, regardless of heaviness of the tail of f , the weak limits in Theorems 3.2 and 4.2 are characterized by a Poisson random measure, the only difference lying in the limiting mean measures. Therefore, the current subsection only presents the results on the moment asymptotics corresponding to Lemmas 5.1 and 5.2. All the arguments that follow are essentially the same as the heavy tail case, so we omit them.

Lemma 5.3. *Let $A \subset \Delta$ be a measurable set, such that $A \cap B_{k,p} \neq \emptyset$. Under the conditions of Theorem 4.2,*

$$\lim_{n \rightarrow \infty} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) = \lim_{n \rightarrow \infty} \mathbb{E}(|\tilde{\Phi}_n^{(k,p)} \cap A|) = \mathbb{E}(|\Phi^{(k,p)} \cap A|) \in (0, \infty).$$

If $A \cap B_{k,p-1} \neq \emptyset$, then

$$\begin{aligned}\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|) &\sim C_3 (nM^d f(R))^{-1}, \quad n \rightarrow \infty, \quad \text{and} \\ \text{Var}(|\Phi_n^{(k,p-1)} \cap A|) &\leq C_4 (nM^d f(R))^{-1},\end{aligned}$$

where C_3 and C_4 are positive constants, independent of n and depending only on d, k, p, A and the probability density f .

The proof for this lemma is postponed to the [Appendix](#).

Appendix: Proofs for Lemmas 5.2 and 5.3

Proof of Lemma 5.2. Repeating the arguments in the proof of Lemma 5.1, we can write

$$\begin{aligned}\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|) &= \frac{n^{p-1}}{(p-1)!} M^{d(p-2)} R^d (f(R))^{p-1} \int_1^\infty \rho^{d-1} d\rho \int_{S^{d-1}} J(\theta) d\theta \int_{(\mathbb{R}^d)^{p-2}} d\mathbf{y} \\ &\quad \times h_1(\rho\theta + M\mathbf{y}/R) g_1(0, \mathbf{y}) \mu_{(0,\mathbf{y})}^{(k)}(A) \\ &\quad \times e^{-nQ_{2M}(R\rho\theta, R\rho\theta + M\mathbf{y})} \frac{f(R\rho)}{f(R)} \prod_{i=1}^{p-2} \frac{f(R\|\rho\theta + My_i/R\|)}{f(R)}.\end{aligned}$$

From here, proceeding the same way as in the previous proof, we can conclude that the triple integral above converges to a positive constant. We also use (3.6) to get that, for some $C_1 > 0$,

$$\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|) \sim C_1 (nM^d f(R))^{-1}, \quad n \rightarrow \infty. \quad (\text{A.16})$$

For the result on variance, we begin with writing

$$\begin{aligned}\mathbb{E}[|\Phi_n^{(k,p-1)} \cap A|^2] &= \sum_{\ell=0}^{p-1} \mathbb{E} \left[\sum_{\substack{\mathcal{Y} \subset \mathcal{P}_n \\ |\mathcal{Y}|=|\mathcal{Y}'|=p-1}} \sum_{|\mathcal{Y} \cap \mathcal{Y}'|=\ell} h_R(\mathcal{Y} \cup \mathcal{Y}') g_M(\mathcal{Y}, \mathcal{P}_n) g_M(\mathcal{Y}', \mathcal{P}_n) \right. \\ &\quad \left. \times \mu_{\mathcal{Y}}^{(k)}(A) \mu_{\mathcal{Y}'}^{(k)}(A) \right] \\ &=: \sum_{\ell=0}^{p-1} I_\ell.\end{aligned}$$

For $\ell = p-1$, we know from (A.16) that $I_{p-1} \sim C_1 (nM^d f(R))^{-1}$ as $n \rightarrow \infty$. For every $\ell \in \{1, \dots, p-2\}$, the condition $|\mathcal{Y} \cap \mathcal{Y}'| = \ell$ requires $g_M(\mathcal{Y}, \mathcal{P}_n) g_M(\mathcal{Y}', \mathcal{P}_n) = 0$, as it is impossible

for $\mathcal{Y}$ and $\mathcal{Y}'$ to be a connected component simultaneously. Therefore, we have

$$\begin{aligned}\mathrm{Var}(|\Phi_n^{(k,p-1)} \cap A|) &= \sum_{\ell=0}^{p-1} I_\ell - [\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|)]^2 \\ &\sim C^*(nM^d f(R))^{-1} + I_0 - [\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|)]^2.\end{aligned}$$

To finish the proof, we thus need to show that

$$I_0 - [\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|)]^2 \rightarrow 0, \quad n \rightarrow \infty.$$

Applying Palm theory yields

$$I_0 = \frac{n^{2(p-1)}}{((p-1)!)^2} \mathbb{E}[h_R(\mathcal{Y}_{12})g_M(\mathcal{Y}_1, \mathcal{Y}_{12} \cup \mathcal{P}_n)g_M(\mathcal{Y}_2, \mathcal{Y}_{12} \cup \mathcal{P}_n)\mu_{\mathcal{Y}_1}^{(k)}(A)\mu_{\mathcal{Y}_2}^{(k)}(A)],$$

where $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint sets of $(p-1)$ i.i.d. points, respectively, and $\mathcal{Y}_{12} := \mathcal{Y}_1 \cup \mathcal{Y}_2$ is independent of $\mathcal{P}_n$. Conditioning on $\mathcal{Y}_{12}$, we have

$$\begin{aligned}I_0 &= \frac{n^{2(p-1)}}{((p-1)!)^2} \mathbb{E}[h_R(\mathcal{Y}_{12})g_M(\mathcal{Y}_1)g_M(\mathcal{Y}_2)\mu_{\mathcal{Y}_1}^{(k)}(A)\mu_{\mathcal{Y}_2}^{(k)}(A) \\ &\quad \times \mathbb{1}\{\mathcal{B}_M(\mathcal{Y}_1) \cap \mathcal{B}_M(\mathcal{Y}_2) = \emptyset\}e^{-nQ_{2M}(\mathcal{Y}_{12})}].\end{aligned}$$

On the other hand,

$$\begin{aligned}&[\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|)]^2 \\ &= \frac{n^{2(p-1)}}{((p-1)!)^2} \mathbb{E}[h_R(\mathcal{Y}_{12})g_M(\mathcal{Y}_1)g_M(\mathcal{Y}_2)\mu_{\mathcal{Y}_1}^{(k)}(A)\mu_{\mathcal{Y}_2}^{(k)}(A)e^{-n(Q_{2M}(\mathcal{Y}_1)+Q_{2M}(\mathcal{Y}_2))}].\end{aligned}$$

Combining them together, we have

$$I_0 - [\mathbb{E}(|\Phi_n^{(k,p-1)} \cap A|)]^2 \leq \mathbb{E}(\Xi_n),$$

where

$$\begin{aligned}\Xi_n &= \frac{n^{2(p-1)}}{((p-1)!)^2} h_R(\mathcal{Y}_{12})g_M(\mathcal{Y}_1)g_M(\mathcal{Y}_2) \\ &\quad \times \mu_{\mathcal{Y}_1}^{(k)}(A)\mu_{\mathcal{Y}_2}^{(k)}(A)(e^{-nQ_{2M}(\mathcal{Y}_{12})} - e^{-n(Q_{2M}(\mathcal{Y}_1)+Q_{2M}(\mathcal{Y}_2))}).\end{aligned}$$

Furthermore, $\mathbb{E}(\Xi_n)$ can be split into two parts,

$$\mathbb{E}[\Xi_n(\mathbb{1}\{\mathcal{B}_{2M}(\mathcal{Y}_1) \cap \mathcal{B}_{2M}(\mathcal{Y}_2) = \emptyset\} + \mathbb{1}\{\mathcal{B}_{2M}(\mathcal{Y}_1) \cap \mathcal{B}_{2M}(\mathcal{Y}_2) \neq \emptyset\})].$$

Note that whenever $\mathcal{B}_{2M}(\mathcal{Y}_1) \cap \mathcal{B}_{2M}(\mathcal{Y}_2) = \emptyset$, we have $\mathcal{Q}_{2M}(\mathcal{Y}_{12}) = \mathcal{Q}_{2M}(\mathcal{Y}_1) + \mathcal{Q}_{2M}(\mathcal{Y}_2)$, in which case $\Xi_n = 0$. So it suffices to consider the other part only. Bounding an exponential term by one,

$$\mathbb{E}[\Xi_n] \leq \frac{n^{2(p-1)}}{((p-1)!)^2} \mathbb{E}[h_R(\mathcal{Y}_{12}) g_M(\mathcal{Y}_1) g_M(\mathcal{Y}_2) \mu_{\mathcal{Y}_1}^{(k)}(A) \mu_{\mathcal{Y}_2}^{(k)}(A) \mathbb{1}\{\mathcal{B}_{2M}(\mathcal{Y}_1) \cap \mathcal{B}_{2M}(\mathcal{Y}_2) \neq \emptyset\}].$$

Notice that

$$g_M(\mathcal{Y}_1) g_M(\mathcal{Y}_2) \mathbb{1}\{\mathcal{B}_{2M}(\mathcal{Y}_1) \cap \mathcal{B}_{2M}(\mathcal{Y}_2) \neq \emptyset\} \leq g_{2M}(\mathcal{Y}_{12}).$$

This, together with the fact that $\mu_{\mathcal{Y}_i}^{(k)}(A) \leq \binom{p-1}{k+1}$, yields

$$\mathbb{E}[\Xi_n] \leq \frac{n^{2(p-1)}}{((p-1)!)^2} \left(\frac{p-1}{k+1} \right)^2 \mathbb{E}[h_R(\mathcal{Y}_{12}) g_{2M}(\mathcal{Y}_{12})]. \quad (\text{A.17})$$

Calculating the expectation portion as in the proof of Lemma 5.1 and using (3.6), we find that the right-hand side in (A.17) equals

$$O(n^{2(p-1)} M^{d(2p-3)} R^d (f(R))^{2(p-1)}) = O((nM^d f(R))^{p-2}) \rightarrow 0, \quad n \rightarrow \infty.$$

Now, the entire proof has been completed. $\square$

Proof of Lemma 5.3. Among these claims in the above lemma, we shall prove the first limit only, that is,

$$\lim_{n \rightarrow \infty} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) = \mathbb{E}(|\Phi^{(k,p)} \cap A|). \quad (\text{A.18})$$

The rest of the proofs will be similar, and hence omitted.

Using Palm theory, we have

$$\mathbb{E}(|\Phi_n^{(k,p)} \cap A|) = \frac{n^p}{p!} \mathbb{E}[h_R(\mathcal{X}_p) g_M(\mathcal{X}_p, \mathcal{X}_p \cup \mathcal{P}_n) \mu_{\mathcal{X}_p}^{(k)}(MA)],$$

where $\mathcal{X}_p = (X_1, \dots, X_p)$ denotes i.i.d. points, independent of $\mathcal{P}_n$. Conditioning on $\mathcal{X}_p$ and changing variables $x_1 \leftrightarrow x$, $x_i \leftrightarrow x + M y_{i-1}$, $i = 2, \dots, p$, we obtain

$$\begin{aligned} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) &= \frac{n^p}{p!} M^{d(p-1)} \int_{\mathbb{R}^d} dx \int_{(\mathbb{R}^d)^{p-1}} dy h_R(x, x + M \mathbf{y}) g_1(0, \mathbf{y}) \mu_{(0, \mathbf{y})}^{(k)}(A) \\ &\quad \times e^{-n Q_{2M}(x, x + M \mathbf{y})} f(x) \prod_{i=1}^{p-1} f(x + M y_i). \end{aligned}$$

Changing into polar coordinate change $x \leftrightarrow (r, \theta)$, along with an additional change of variable $\rho = a(R)^{-1}(r - R)$ we have

$$\begin{aligned} \mathbb{E}(|\Phi_n^{(k,p)} \cap A|) &= \frac{n^p}{p!} M^{d(p-1)} a(R) R^{d-1} (f(R))^p \int_0^\infty d\rho \int_{S^{d-1}} J(\theta) d\theta \int_{(\mathbb{R}^d)^{p-1}} d\mathbf{y} \\ &\quad \times \left(1 + \frac{a(R)}{R} \rho\right)^{d-1} h_R((R + a(R)\rho)\theta, (R + a(R)\rho)\theta + M\mathbf{y}) \\ &\quad \times g_1(0, \mathbf{y}) \mu_{(0, \mathbf{y})}^{(k)}(A) e^{-n \mathcal{Q}_{2M}((R + a(R)\rho)\theta, (R + a(R)\rho)\theta + M\mathbf{y})} \\ &\quad \times \frac{f(R + a(R)\rho)}{f(R)} \prod_{i=1}^{p-1} \frac{f(\|(R + a(R)\rho)\theta + M\mathbf{y}_i\|)}{f(R)}. \end{aligned} \quad (\text{A.19})$$

Using the Taylor expansion, we have

$$\|(R + a(R)\rho)\theta + M\mathbf{y}_i\| = R + a(R)\rho + M(\langle \theta, \mathbf{y}_i \rangle + \gamma_n(\rho, \theta, \mathbf{y}_i)), \quad (\text{A.20})$$

where $\gamma_n(\rho, \theta, \mathbf{y}_i) \rightarrow 0$ uniformly for $\rho > 0$, $\theta \in S^{d-1}$, and $\mathbf{y}_i$ in a bounded set in $\mathbb{R}^d$. Denoting

$$\xi_n(\rho, \theta, \mathbf{y}_i) = \frac{\langle \theta, \mathbf{y}_i \rangle + \gamma_n(\rho, \theta, \mathbf{y}_i)}{a(R)/M},$$

the right-hand side in (A.20) is equal to $R + a(R)(\rho + \xi_n(\rho, \theta, \mathbf{y}_i))$. Due to the uniform convergence of $\gamma_n(\rho, \theta, \mathbf{y}_i)$, it is easy to show that for every $M > 0$,

$$\sup_{\substack{\rho > 0, n \geq 1, \theta \in S^{d-1} \\ \mathbf{y}_i \in [-M, M]^d}} |\xi_n(\rho, \theta, \mathbf{y}_i)| < \infty, \quad (\text{A.21})$$

and further,

$$\xi_n(\rho, \theta, \mathbf{y}_i) \rightarrow c^{-1} \langle \theta, \mathbf{y}_i \rangle \quad \text{as } n \rightarrow \infty. \quad (\text{A.22})$$

In the following, we shall compute the limits for each term in (A.19) under the integral sign, and then establish an appropriate integrable bound for the application of the dominated convergence theorem. From (4.2), we have that $(1 + a(R)\rho/R) \rightarrow 1$ for all $\rho > 0$, and for sufficiently large n , this is bounded by $2(\rho \vee 1)^{d-1}$. Subsequently, from (A.20) and (A.22), we have that

$$h_R((R + a(R)\rho)\theta, (R + a(R)\rho)\theta + M\mathbf{y}) \rightarrow \mathbb{1}\{\rho + c^{-1} \langle \theta, \mathbf{y}_i \rangle \geq 0, i = 1, \dots, p-1\}.$$

As for the ratio terms for the density f , using (4.1) we write

$$\frac{f(R + a(R)\rho)}{f(R)} = \frac{L(R + a(R)\rho)}{L(R)} e^{-[\psi(R + a(R)\rho) - \psi(R)]} \quad (\text{A.23})$$

so that $L(R + a(R)\rho)/L(R) \rightarrow 1$ for all $\rho > 0$, and

$$e^{-[\psi(R + a(R)\rho) - \psi(R)]} = \exp\left\{-\int_0^\rho \frac{a(R)}{a(R + a(R)r)} dr\right\} \rightarrow e^{-\rho}, \quad n \rightarrow \infty.$$

For the last convergence, we applied an elementary result in p. 142 of [18], which asserts that $a(x)/a(x + a(x)r) \rightarrow 1$ as $x \rightarrow \infty$, uniformly for r in any bounded interval. In order to give a proper upper bound for (A.23), we let

$$q_m(n) = \frac{\psi^{-1}(\psi(R) + m) - R}{a(R)}, \quad m \geq 1,$$

equivalently, $\psi(R + a(R)q_m(n)) = \psi(R) + m$. Accordingly to Lemma 5.2 in [5], for every $0 < \epsilon < (d + \gamma p)^{-1}$ (γ is a parameter at (4.4)), there exists $N \geq 1$ such that $q_m(n) \leq e^{m\epsilon}/\epsilon$ for all $n \geq N$ and $m \geq 1$. Since ψ is increasing, we have

$$\begin{aligned} e^{-[\psi(R+a(R)\rho)-\psi(R)]} \mathbb{1}\{\rho > 0\} &= \sum_{m=0}^{\infty} \mathbb{1}\{q_m(n) < \rho \leq q_{m+1}(n)\} e^{-[\psi(R+a(R)\rho)-\psi(R)]} \\ &\leq \sum_{m=0}^{\infty} \mathbb{1}\{0 < \rho \leq e^{(m+1)\epsilon}/\epsilon\} e^{-m} \end{aligned}$$

for all $n \geq N$. For the derivation of the bound for L , we use (4.4), that is,

$$\frac{L(R + a(R)\rho)}{L(R)} \mathbb{1}\{\rho > 0\} \leq 2C(\rho \vee 1)^\gamma \mathbb{1}\{\rho > 0\}$$

for sufficiently large n . Combining these bounds,

$$\frac{f(R + a(R)\rho)}{f(R)} \leq 2C(\rho \vee 1)^\gamma \sum_{m=0}^{\infty} \mathbb{1}\{0 < \rho \leq e^{(m+1)\epsilon}/\epsilon\} e^{-m}.$$

We next deal with the product terms of the probability densities in (A.19). For each $i = 1, \dots, p-1$, it follows from (A.20), (A.21) and (A.22) that

$$\begin{aligned} \frac{f(\|(R + a(R)\rho)\theta + My_i\|)}{f(R)} &= \frac{L(R + a(R)(\rho + \xi_n(\rho, \theta, y_i)))}{L(R)} \\ &\quad \times \exp\left\{-\int_0^{\rho + \xi_n(\rho, \theta, y_i)} \frac{a(R)}{a(R + a(R)r)} dr\right\} \\ &\rightarrow e^{-\rho - c^{-1}\langle \theta, y_i \rangle} \end{aligned}$$

for all $\rho > 0$, $\theta \in S^{d-1}$, and $y_i \in \mathbb{R}^d$. As for the suitable integrable bound, simply dropping the exponential term, we have

$$\frac{f(\|(R + a(R)\rho)\theta + My_i\|)}{f(R)} \leq C'(\rho \vee 1)^{\gamma(p-1)}$$

for some constant $C' > 0$. For the exponential term in (A.19),

$$\begin{aligned} n\mathcal{Q}_{2M}((R + a(R)\rho)\theta, (R + a(R)\rho)\theta + M\mathbf{y}) \\ = nM^d f(R) \int_{B_2(0, \mathbf{y})} \frac{f(\|(R + a(R)\rho)\theta + Mv\|)}{f(R)} dv \rightarrow 0, \end{aligned}$$

since (4.5) implies $nM^d f(R) \rightarrow 0$, $n \rightarrow \infty$.

Combining all convergence results together, while assuming the applicability of the dominated convergence theorem, we get (A.18) as desired. Finally, apply all the bounds derived thus far, and note that

$$\int_0^\infty \sum_{m=0}^\infty \mathbb{1}\{0 < \rho \leq e^{(m+1)\epsilon}/\epsilon\} e^{-m(\rho \vee 1)^{d+\gamma p-1}} d\rho \leq \left(\frac{e^\epsilon}{\epsilon}\right)^{d+\gamma p} \sum_{m=0}^\infty e^{-[1-\epsilon(d+\gamma p)]m} < \infty,$$

since $0 < \epsilon < (d + \gamma p)^{-1}$. Therefore, the dominated convergence theorem is applicable. $\square$

Acknowledgments

TO's research is partially supported by the NSF Grant DMS-1811428.

OB was supported by the Israel Science Foundation Grant #1965/19.

References

- [1] Adler, R.J., Bobrowski, O. and Weinberger, S. (2014). Crackle: The homology of noise. *Discrete Comput. Geom.* **52** 680–704. [MR3279544](#) <https://doi.org/10.1007/s00454-014-9621-6>
- [2] Attouch, H. (1984). *Variational Convergence for Functions and Operators. Applicable Mathematics Series*. Boston, MA: Pitman. [MR0773850](#)
- [3] Bai, S., Owada, T. and Wang, Y. (2019). A functional non-central limit theorem for multiple-stable processes with long-range dependence. Available at [arXiv:1902.00628](#).
- [4] Balkema, A.A., Embrechts, P. and Nolde, N. (2010). Meta densities and the shape of their sample clouds. *J. Multivariate Anal.* **101** 1738–1754. [MR2610743](#) <https://doi.org/10.1016/j.jmva.2010.02.010>
- [5] Balkema, G. and Embrechts, P. (2004). Multivariate excess distributions. Available at www.math.ethz.ch/~embrecht/ftp/guuspe08Jun04.pdf.
- [6] Balkema, G. and Embrechts, P. (2007). *High Risk Scenarios and Extremes: A Geometric Approach. Zurich Lectures in Advanced Mathematics*. Zürich: European Mathematical Society (EMS). [MR2372552](#) <https://doi.org/10.4171/035>
- [7] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. [MR0898871](#) <https://doi.org/10.1017/CBO9780511721434>
- [8] Björner, A. (1995). Topological methods. In *Handbook of Combinatorics, Vols. 1, 2* 1819–1872. Amsterdam: Elsevier. [MR1373690](#)
- [9] Bobrowski, O. and Adler, R.J. (2014). Distance functions, critical points, and the topology of random Čech complexes. *Homology, Homotopy Appl.* **16** 311–344. [MR3280987](#) <https://doi.org/10.4310/HHA.2014.v16.n2.a18>

- [10] Bobrowski, O. and Kahle, M. (2018). Topology of random geometric complexes: A survey. *J. Appl. Comput. Topol.* **1** 331–364. [MR3975557](#) <https://doi.org/10.1007/s41468-017-0010-0>
- [11] Bobrowski, O., Kahle, M. and Skraba, P. (2017). Maximally persistent cycles in random geometric complexes. *Ann. Appl. Probab.* **27** 2032–2060. [MR3693519](#) <https://doi.org/10.1214/16-AAP1232>
- [12] Bobrowski, O. and Mukherjee, S. (2015). The topology of probability distributions on manifolds. *Probab. Theory Related Fields* **161** 651–686. [MR3334278](#) <https://doi.org/10.1007/s00440-014-0556-x>
- [13] Carlsson, G. (2009). Topology and data. *Bull. Amer. Math. Soc. (N.S.)* **46** 255–308. [MR2476414](#) <https://doi.org/10.1090/S0273-0979-09-01249-X>
- [14] Das, B. and Ghosh, S. (2013). Weak limits for exploratory plots in the analysis of extremes. *Bernoulli* **19** 308–343. [MR3019497](#) <https://doi.org/10.3150/11-BEJ401>
- [15] Das, B. and Resnick, S.I. (2008). QQ plots, random sets and data from a heavy tailed distribution. *Stoch. Models* **24** 103–132. [MR2384693](#) <https://doi.org/10.1080/15326340701828308>
- [16] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. New York: Springer. [MR2234156](#) <https://doi.org/10.1007/0-387-34471-3>
- [17] Decreusefond, L., Schulte, M. and Thäle, C. (2016). Functional Poisson approximation in Kantorovich–Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Probab.* **44** 2147–2197. [MR3502603](#) <https://doi.org/10.1214/15-AOP1020>
- [18] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance. Applications of Mathematics (New York)* **33**. Berlin: Springer. [MR1458613](#) <https://doi.org/10.1007/978-3-642-33483-2>
- [19] Ghosh, S. and Resnick, S. (2010). A discussion on mean excess plots. *Stochastic Process. Appl.* **120** 1492–1517. [MR2653263](#) <https://doi.org/10.1016/j.spa.2010.04.002>
- [20] Ghrist, R. (2008). Barcodes: The persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)* **45** 61–75. [MR2358377](#) <https://doi.org/10.1090/S0273-0979-07-01191-3>
- [21] Ghrist, R. (2014). *Elementary Applied Topology*. Charleston, SC: Createspace.
- [22] Hatcher, A. (2002). *Algebraic Topology*. Cambridge: Cambridge Univ. Press. [MR1867354](#)
- [23] Hiraoka, Y., Shirai, T. and Trinh, K.D. (2018). Limit theorems for persistence diagrams. *Ann. Appl. Probab.* **28** 2740–2780. [MR3847972](#) <https://doi.org/10.1214/17-AAP1371>
- [24] Kahle, M. (2011). Random geometric complexes. *Discrete Comput. Geom.* **45** 553–573. [MR2770552](#) <https://doi.org/10.1007/s00454-010-9319-3>
- [25] Kahle, M. and Meckes, E. (2013). Limit theorems for Betti numbers of random simplicial complexes. *Homology, Homotopy Appl.* **15** 343–374. [MR3079211](#) <https://doi.org/10.4310/HHA.2013.v15.n1.a17>
- [26] Matheron, G. (1975). *Random Sets and Integral Geometry*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley. [MR0385969](#)
- [27] Molchanov, I. (2005). *Theory of Random Sets. Probability and Its Applications (New York)*. London: Springer. [MR2132405](#)
- [28] Niyogi, P., Smale, S. and Weinberger, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419–441. [MR2383768](#) <https://doi.org/10.1007/s00454-008-9053-2>
- [29] Niyogi, P., Smale, S. and Weinberger, S. (2011). A topological view of unsupervised learning from noisy data. *SIAM J. Comput.* **40** 646–663. [MR2810909](#) <https://doi.org/10.1137/090762932>
- [30] Owada, T. (2017). Functional central limit theorem for subgraph counting processes. *Electron. J. Probab.* **22** Art. ID 17. [MR3622887](#) <https://doi.org/10.1214/17-EJP30>
- [31] Owada, T. (2018). Limit theorems for Betti numbers of extreme sample clouds with application to persistence barcodes. *Ann. Appl. Probab.* **28** 2814–2854. [MR3847974](#) <https://doi.org/10.1214/17-AAP1375>

- [32] Owada, T. and Adler, R.J. (2017). Limit theorems for point processes under geometric constraints (and topological crackle). *Ann. Probab.* **45** 2004–2055. [MR3650420](#) <https://doi.org/10.1214/16-AOP1106>
- [33] Penrose, M. (2003). *Random Geometric Graphs*. *Oxford Studies in Probability* **5**. Oxford: Oxford Univ. Press. [MR1986198](#) <https://doi.org/10.1093/acprof:oso/9780198506263.001.0001>
- [34] Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. *Applied Probability. A Series of the Applied Probability Trust* **4**. New York: Springer. [MR0900810](#) <https://doi.org/10.1007/978-0-387-75953-1>
- [35] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. [MR2271424](#)
- [36] Samorodnitsky, G. and Wang, Y. (2019). Extremal theory for long range dependent infinitely divisible processes. *Ann. Probab.* **47** 2529–2562. [MR3980927](#) <https://doi.org/10.1214/18-AOP1318>
- [37] Schulte, M. and Thäle, C. (2012). The scaling limit of Poisson-driven order statistics with applications in geometric probability. *Stochastic Process. Appl.* **122** 4096–4120. [MR2971726](#) <https://doi.org/10.1016/j.spa.2012.08.011>
- [38] The GUDHI Project (2015). GUDHI user and reference manual. GUDHI Editorial Board.
- [39] Yogeshwaran, D., Subag, E. and Adler, R.J. (2017). Random geometric complexes in the thermodynamic regime. *Probab. Theory Related Fields* **167** 107–142. [MR3602843](#) <https://doi.org/10.1007/s00440-015-0678-9>

Received October 2019 and revised January 2020