J. Appl. Probab. **60,** 314–340 (2023)
*doi:*10.1017/jpr.2022.39

LIMIT THEORY FOR *U*-STATISTICS UNDER GEOMETRIC AND TOPOLOGICAL CONSTRAINTS WITH RARE EVENTS

TAKASHI OWADA,* *Purdue University*

Abstract

We study the geometric and topological features of U-statistics of order k when the k-tuples satisfying geometric and topological constraints do not occur frequently. Using appropriate scaling, we establish the convergence of U-statistics in vague topology, while the structure of a non-degenerate limit measure is also revealed. Our general result shows various limit theorems for geometric and topological statistics, including persistent Betti numbers of Čech complexes, the volume of simplices, a functional of the Morse critical points, and values of the min-type distance function. The required vague convergence can be obtained as a result of the limit theorem for point processes induced by U-statistics. The latter convergence particularly occurs in the $\mathcal{M}_0$-topology.

Keywords: Point process; persistent Betti number; Morse critical point; extreme value theory

2020 Mathematics Subject Classification: Primary 60D05
Secondary 60G55; 55U10; 60G70

1. Introduction

The main focus of this paper is to examine the geometric and topological features of U-statistics when the geometric configuration of a point cloud does not occur frequently. Let $\mathcal{X}_n = \{X_1, \ldots, X_n\} \subset \mathbb{R}^d$, $d \geq 2$, be a random sample, and let (r_n) be a sequence of positive numbers such that $r_n \to 0$ as $n \to \infty$. A geometric graph $G(\mathcal{X}_n, r_n)$ is an undirected graph with a vertex set $\mathcal{X}_n$ and edges $[X_i, X_j]$ for all pairs $X_i, X_j \in \mathcal{X}_n$ such that $\|X_i - X_j\| \leq r_n$, where $\|\cdot\|$ denotes the Euclidean norm. The monograph [30] by Penrose covers a range of related topics, including subgraph counts, the vertex degree, the clique number, and the formation of a giant component. As seen in the monograph, many of the geometric statistics can be represented as U-statistics. Namely, for every $n, k \geq 2$,

$$T_{k,n}^{(1)} := \sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}|=k} H_n(\mathcal{Y}), \tag{1.1}$$

where $|\mathcal{Y}|$ is the cardinality of a point set $\mathcal{Y}$ in $\mathbb{R}^d$ and $H_n \colon (\mathbb{R}^d)^k \to \mathbb{R}$ is defined as

$$H_n(x_1, \ldots, x_k) = H(r_n^{-1}x_1, \ldots, r_n^{-1}x_k), \quad x_1, \ldots, x_k \in \mathbb{R}^d$$

Received 26 May 2021; revision received 15 March 2022.
* Postal address: Department of Statistics, Purdue University, West Lafayette, 47907, USA. Email address: owada@purdue.edu

for some symmetric and translation-invariant map $H\colon (\mathbb{R}^d)^k \to \mathbb{R}$. Additionally, we can also consider a certain variant of (1.1), defined by

$$T_{k,n}^{(2)} := \sum_{\mathcal{Y}\subset\mathcal{X}_n, |\mathcal{Y}|=k} H_n(\mathcal{Y})\mathbb{1}\{\|y-z\| \geq r_n \text{ for all } y\in\mathcal{Y} \text{ and } z \in \mathcal{X}_n \setminus \mathcal{Y}\},$$

where $\mathbb{1}\{\cdot\}$ denotes an indicator function. This is of particular importance when we are examining k-tuples $\mathcal{Y} \subset \mathcal{X}_n$, which not only satisfy geometric conditions implicit in H_n but are also separated from the other points in $\mathcal{X}_n$. In Section 2 we provide a more general definition of $T_{k,n}^{(i)}$ for $i = 1, 2$. If one takes

$$H(x_1, \dots, x_k) = \mathbb{1}\{G(\{x_1, \dots, x_k\}, 1) \cong \Gamma\}, \quad x_i \in \mathbb{R}^d, \tag{1.2}$$

where Γ is a connected graph with k vertices and $\cong$ means graph isomorphism, then $T_{k,n}^{(1)}$ represents the number of subgraphs isomorphic to Γ (with radius r_n) and $T_{k,n}^{(2)}$ counts the number of connected components isomorphic to Γ. In addition to the random geometric graph setup, many of the functionals in stochastic geometry, such as intrinsic volumes of intersection processes, the volumes of simplices, can be treated under the framework of U-statistics [2, 6, 22, 24, 31]. Additionally, $T_{k,n}^{(i)}$ can also arise when examining random geometric complexes. For example, with an appropriate choice of H, $T_{k,n}^{(i)}$ can be used to dictate the behavior of topological invariants of a geometric complex [3, 4, 17, 29].

The limiting behavior of $T_{k,n}^{(i)}$ depends crucially on the decay rate of r_n as $n \to \infty$. If r_n is chosen such that $n^k r_n^{d(k-1)} \to \infty$ as $n \to \infty$, it then follows that $\mathbb{E}\big[T_{k,n}^{(i)}\big] \to \infty$. This implies that the geometric configuration of k-tuples relating to H_n asymptotically occurs infinitely many times. Then $T_{k,n}^{(i)}$ obeys a central limit theorem:

$$\frac{T_{k,n}^{(i)} - \mathbb{E}\big[T_{k,n}^{(i)}\big]}{\sqrt{\operatorname{Var}\big(T_{k,n}^{(i)}\big)}}$$

converges weakly to a standard normal random variable. Last *et al.* [24] and Reitzner and Schulte [31] established the rate of convergence in normal approximation in terms of the Wasserstein distance and the Kolmogorov distance, via the Malliavin–Stein method together with Palm calculus for a Poisson point process. The monograph [23] by Last and Penrose provides details of this line of research. Furthermore, Blaszczyszyn *et al.* [2] derived asymptotic normality of geometric statistics (not necessarily U-statistics) when the input process exhibits fast decay of correlations. In the context of random topology, proving the asymptotic normality of the simplex counts, which themselves are U-statistics, will be a crucial step in deriving the central limit theorem for topological invariants, such as the Euler characteristic and Betti numbers [17, 21, 29, 35]. If r_n decays more slowly, such that $n^k r_n^{d(k-1)} \to c$, $n \to \infty$, for some $c \in (0, \infty)$, the k-tuples that satisfy the geometric conditions in H_n will occur less frequently. Then $\mathbb{E}\big[T_{k,n}^{(i)}\big]$ tends to a finite positive constant as $n \to \infty$. In particular, if one takes H_n as in (1.2), $T_{k,n}^{(i)}$ converges weakly to a Poisson random variable as $n \to \infty$, that is, for all integers $\ell \geq 0$ and $i = 1, 2$,

$$\mathbb{P}\big(T_{k,n}^{(i)} = \ell\big) \to \mathbb{P}(\mathrm{Poi}(\nu_k) = \ell), \quad n \to \infty, \tag{1.3}$$

where 'Poi(ν_k)' stands for a Poisson random variable with mean $\nu_k \in (0, \infty)$. In research relating to random topology, Kahle and Meckes [17] and Owada and Thomas [29] proved that

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

the Betti number of a geometric complex converges weakly to the difference of time-changed homogeneous Poisson processes on the real half-line. Furthermore, Decreusefond *et al.* [6] provided the rate of convergence of a point process induced by U-statistics in terms of the Kantorovich–Rubinstein distance.

The main aim of this paper is to explore the limiting behavior of $T_{k,n}^{(i)}$ when the k-tuples satisfying the geometric conditions in H_n are even less likely to occur. Specifically, we assume that r_n decays to 0 at a faster rate: $n^k r_n^{d(k-1)} \to 0$ as $n \to \infty$. It then follows that

$$\mathbb{P}\big(|T_{k,n}^{(i)}| \ge \epsilon\big) \to 0, \quad n \to \infty$$

for all $\epsilon > 0$. In this setting, we aim to detect a sequence (v_n) that grows to infinity, so that

$$\big(v_n \mathbb{P}\big(T_{k,n}^{(i)} \in \cdot\big),\ n \ge 1\big) \tag{1.4}$$

converges to a non-degenerate limiting measure. Since (1.4) is not a sequence of probability measures, weak convergence as in (1.3) can no longer be used. Alternatively, by exploiting the notion of vague convergence (see [19] and [32]), we show that in the space of Radon measures on $[-\infty, \infty] \setminus \{0\}$,

$$v_n \mathbb{P}\big(T_{k,n}^{(i)} \in \cdot\big) \stackrel{v}{\to} \mu_k, \quad n \to \infty, \tag{1.5}$$

where $\stackrel{v}{\to}$ denotes vague convergence and μ_k is a non-null limit measure with $\mu_k(\{\pm\infty\}) = 0$. From (1.5), one can deduce, from the perspective of vague topology, the exact rate (up to the scale) of the probability that $T_{k,n}^{(i)}$ becomes non-trivial (i.e. non-zero). Furthermore, the limit μ_k is expected to dictate the geometric and topological structure of $T_{k,n}^{(i)}$ which still remains in the limit. In the literature of random topology (not necessarily related to random geometric complexes), one of the key focuses is how rapidly each homology group appears and disappears [5, 9, 16, 18, 34]. In the same spirit, we replace $T_{k,n}^{(i)}$ in (1.5) with the (persistent) Betti numbers and explore the rate of v_n, as well as the structure of μ_k. See Section 3.1 for more details.

From a technical viewpoint, the articles most relevant to this study are those of Fasen and Roy [8] and Hult and Samorodnitsky [14]. In these papers the authors established large deviations for point processes based on a stationary sequence with heavy-tailed marginals and non-trivial dependency. Using the same approach as applied in these papers, the required vague convergence in (1.5) can be derived from the limit theorem for the sequence

$$\left(v_n \mathbb{P}\left(\sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}|=k} \delta_{H_n(\mathcal{Y})} \in \cdot\right),\ n \ge 1\right), \tag{1.6}$$

where δ_x denotes the Dirac measure at $x \in \mathbb{R}$. The point process in (1.6) is a random element into the space of Radon point measures, but this space is not locally compact. Accordingly, the convergence of (1.6) can no longer be treated in the vague topology. Alternatively, we aim to demonstrate the limit theory for (1.6) in the so-called $\mathcal{M}_0$-topology. This notion was first developed by Hult and Lindskog [13] and has been used extensively, especially in extreme value theory, for the study of regular variation of stochastic processes [8, 14, 25, 33]. Proposition 4.1 gives a more precise statement of this result. After completing the limit theorem for (1.6), this paper proceeds to show (1.5) by means of a continuous mapping theorem for $\mathcal{M}_0$-convergence, as well as by using various approximation arguments.

The remainder of this paper is structured as follows. Section 2 presents the limit theorems for $T_{k,n}^{(i)}$ under a more general setup. Section 3 applies our general result to deduce the limit theory for geometric and topological statistics, including persistent Betti numbers of Čech complexes, the volume of simplices, a functional of the Morse critical points, and values of the min-type distance function. All the proofs are deferred to Section 4.

Before commencing the main body of the paper, let us add a few more comments on our setup. First, we assume that the density f of $\mathcal{X}_n$ is a.e. continuous and bounded. We can obtain the same result under a weaker assumption that

$$\int_{(\mathbb{R}^d)^k} f(x)^{2k-2}\,\mathrm{d}x < \infty.$$

However, we have decided to impose stronger assumptions in order to avoid technical arguments relating to moment convergence, which necessarily involves the density f. Second, we observe that the same result can be obtained even if a random sample $\mathcal{X}_n$ is replaced by a Poisson point process $\mathcal{P}_n := \{X_1, \ldots, X_{N_n}\}$, where N_n is Poisson-distributed with mean n, independent of (X_i). In this case, one needs to use Palm calculus (see e.g. [30, Section 1.7]) when computing the moments of $T^{(i)}_{k,n}$. Finally, we remark that establishing a more general limit theory for a (discrete time) process $\mathcal{X}_n$ with non-trivial dependency remains a topic of further research. Indeed, Fasen and Roy [8] and Hult and Samorodnitsky [14] examined a moving average process and derived a series of large deviation results in the form of (1.5) and (1.6). In such cases, the structure of the limit μ_k becomes more complicated, reflecting a significant amount of clusters induced by a moving average process. In the case of Poisson limit theorems, a similar line of research can be found in [27], which studied the asymptotic behavior of Betti numbers generated by a moving average process.

2. Main limit theorem

We take a random sample $\mathcal{X}_n = \{X_1, \ldots, X_n\} \subset \mathbb{R}^d$, $d \geq 2$, with density f, and a sequence of (non-random) radii $r_n \to 0$, $n \to \infty$, such that $n^k r_n^{d(k-1)} \to 0$ for some $k \geq 2$. Assume that f is a.e. continuous and bounded, that is, $\|f\|_\infty := \operatorname{ess\,sup}_{x\in\mathbb{R}^d} f(x) < \infty$. Fix $m \geq 1$ and let $H\colon (\mathbb{R}^d)^k \to \mathbb{R}^m$ be a measurable function satisfying the following conditions.

(H1) H is symmetric about permutations, i.e. $H(x_1, \ldots, x_k) = H(x_{\sigma(1)}, \ldots, x_{\sigma(k)})$ for all $x_i \in \mathbb{R}^d$ and every permutation σ of $\{1, \ldots, k\}$.

(H2) H is translation-invariant, i.e. $H(x_1, \ldots, x_k) = H(x_1 + y, \ldots, x_k + y)$ for all $x_i, y \in \mathbb{R}^d$.

(H3) H is locally determined, i.e. there exists $L > 0$ such that $H(x_1, \ldots, x_k) = 0$ whenever $\operatorname{diam}(x_1, \ldots, x_k) \geq L$, where $\operatorname{diam}(x_1, \ldots, x_k) = \max_{1\leq i,j\leq k} \|x_i - x_j\|$.

(H4) H is integrable in the sense of

$$\int_{(\mathbb{R}^d)^{k-1}} \|H(0, y_1, \ldots, y_{k-1})\|\,\mathrm{d}\mathbf{y} < \infty.$$

We also define a scaled version of H by

$$H_n(x_1, \ldots, x_k) := H(r_n^{-1}x_1, \ldots, r_n^{-1}x_k), \qquad x_i \in \mathbb{R}^d. \tag{2.1}$$

Given a subset $\mathcal{Y}$ of k points in $\mathbb{R}^d$, a finite point set $\mathcal{Z} \supset \mathcal{Y}$ in $\mathbb{R}^d$, and $\mathbf{t} = (t_1, \ldots, t_m) \in [0, \infty)^m$, we define

$$c(\mathcal{Y}, \mathcal{Z}; \mathbf{t}) = (\mathbb{1}\{\|y - z\| \geq t_i \text{ for all } y \in \mathcal{Y} \text{ and } z \in \mathcal{Z} \setminus \mathcal{Y}\})_{i=1}^m. \tag{2.2}$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

In particular, the ith component of (2.2) requires that each point in $\mathcal{Y}$ must be distance at least t_i from all the remaining points in $\mathcal{Z} \setminus \mathcal{Y}$. Moreover,

$$G(\mathcal{Y}, \mathcal{Z}; \mathbf{t}) := H(\mathcal{Y}) \circ c(\mathcal{Y}, \mathcal{Z}; \mathbf{t}), \tag{2.3}$$

where $\circ$ means the Hadamard product: for two matrices $A = (a_{ij})$ and $B = (b_{ij})$ of the same dimension $\ell_1 \times \ell_2$, $A \circ B$ represents an $\ell_1 \times \ell_2$ matrix with (i, j) element given by $a_{ij}b_{ij}$. For $\mathcal{Y} = (y_1, \dots, y_k) \in (\mathbb{R}^d)^k$ and $a \in \mathbb{R}$, we write $a\mathcal{Y} = (ay_1, \dots, ay_k)$. We then define

$$c_n(\mathcal{Y}, \mathcal{Z}; \mathbf{t}) := c\big(r_n^{-1}\mathcal{Y}, r_n^{-1}\mathcal{Z}; \mathbf{t}\big) = \big(\mathbb{1}\{\|y - z\| \geq r_n t_i \text{ for all } y \in \mathcal{Y} \text{ and } z \in \mathcal{Z} \setminus \mathcal{Y}\}\big)_{i=1}^m, \tag{2.4}$$

and

$$G_n(\mathcal{Y}, \mathcal{Z}; \mathbf{t}) := G\big(r_n^{-1}\mathcal{Y}, r_n^{-1}\mathcal{Z}; \mathbf{t}\big) = H_n(\mathcal{Y}) \circ c_n(\mathcal{Y}, \mathcal{Z}; \mathbf{t}). \tag{2.5}$$

The primary objective of this paper is to examine the behavior of

$$T_{k,n}^{(1)} := \sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}|=k} H_n(\mathcal{Y}) \quad \text{and} \quad T_{k,n}^{(2)} := \sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}|=k} G_n(\mathcal{Y}, \mathcal{X}_n; \mathbf{t}). \tag{2.6}$$

For the rigorous description of the asymptotic theory of (2.6), one needs the following notations and concepts. Our main references are [19] and [32]. First, let $E := (\overline{\mathbb{R}})^m \setminus \{\mathbf{0}\} = [-\infty, \infty]^m \setminus \{\mathbf{0}\}$ with $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^m$, and let $M_+(E)$ be the space of Radon measures on E, and $M_p(E)$ denotes the space of Radon point measures on E. Note that $M_p(E)$ is a closed subset of $M_+(E)$ in the *vague topology*; see Proposition 3.14 in [32]. Define $C_K^+(E)$ to be the collection of non-negative and continuous functions on E with compact support. For $\eta_n, \eta \in M_+(E)$, we say that η_n converges vaguely to η, denoted by $\eta_n \xrightarrow{v} \eta$ in $M_+(E)$, if it holds that

$$\int_E g(\mathrm{x})\eta_n(\mathrm{dx}) \to \int_E g(\mathrm{x})\eta(\mathrm{dx}) \quad \text{for all } g \in C_K^+(E).$$

Now we can state our main theorem. The proof is deferred to Section 4.1.

Theorem 2.1. *Under the assumptions above, for each* $i = 1, 2$*, we have*

$$\big(n^k r_n^{d(k-1)}\big)^{-1} \mathbb{P}\big(T_{k,n}^{(i)} \in \cdot\big) \xrightarrow{v} C_k \lambda\big\{\mathbf{y} \in (\mathbb{R}^d)^{k-1} : H(0, \mathbf{y}) \in \cdot\big\} \quad \text{in } M_+(E), \quad n \to \infty,$$

where

$$\mathbf{y} = (y_1, \dots, y_{k-1}) \in (\mathbb{R}^d)^{k-1}, \quad C_k := (k!)^{-1} \int_{\mathbb{R}^d} f(x)^k \, \mathrm{d}x,$$

and λ *is the Lebesgue measure on* $(\mathbb{R}^d)^{k-1}$.

The U-statistics $T_{k,n}^{(1)}$ is associated with k-tuples satisfying the geometric conditions implicit in H_n, while $T_{k,n}^{(2)}$ adds an extra constraint that the points in $\mathcal{Y}$ must be distance at least a constant multiple of r_n from the remaining points in $\mathcal{X}_n \setminus \mathcal{Y}$. Despite such a difference, Theorem 2.1 indicates that the behaviors of $T_{k,n}^{(i)}$, $i = 1, 2$ are asymptotically the same. In other words, the extra restriction imposed on $T_{k,n}^{(2)}$ is asymptotically negligible, whenever r_n decays so fast that $n^k r_n^{d(k-1)} \to 0$ as $n \to \infty$.

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

3. Geometric and topological applications

In this section we use Theorem 2.1 to deduce the limit theory for geometric and topological statistics satisfying conditions **(H1)–(H4)**. Throughout this section we assume that $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ is a random sample in $\mathbb{R}^d$, $d \geq 2$, with density f, and let (r_n) be a sequence of connectivity radii with $r_n \to 0$ as $n \to \infty$. Furthermore, f is assumed to be a.e. continuous and bounded. Denote λ to be the Lebesgue measure in a given dimension. All of the proofs are provided in Sections 4.2–4.4. All the examples here are more or less concerned with a *Čech complex* defined on $\mathcal{X}_n$ with connectivity radius r_n.

Definition 3.1. Given a set $\mathcal{X} = \{x_1, \ldots, x_n\}$ of points in $\mathbb{R}^d$ and a positive number $r > 0$, we define a Čech complex $\check{C}(\mathcal{X}, r)$ as follows.

- The 0-simplices are the points in $\mathcal{X}$.
- The p-simplex $[x_{i_0}, \ldots, x_{i_p}]$, $1 \leq i_0 < \cdots < i_p \leq n$, belongs to $\check{C}(\mathcal{X}, r)$ if

$$\bigcap_{\ell=0}^{p} B(x_{i_\ell}, r/2) \neq \emptyset,$$

where $B(x, r)$ is a d-dimensional closed ball of radius r centered at $x \in \mathbb{R}^d$.

3.1. Persistent Betti number

Our first application is concerned with the *persistent Betti number*. Because of the recent development of topological data analysis, the (persistent) Betti number has been intensively studied as a basic topological invariant representing, roughly, the creation and destruction of topological cycles of various dimensions [4, 12, 15, 17, 20, 36, 37]. First we define a family

$$\big(\check{C}\big(r_n^{-1}\mathcal{X}_n, t\big),\ t \geq 0\big) = \big(\check{C}(\mathcal{X}_n, r_n t),\ t \geq 0\big) \tag{3.1}$$

of Čech complexes over a scaled random sample $r_n^{-1}\mathcal{X}_n$. Note that (3.1) constitutes a nested sequence of Čech complexes satisfying monotonicity property $\check{C}(\mathcal{X}_n, r_n s) \subset \check{C}(\mathcal{X}_n, r_n t)$ for all $0 < s \leq t < \infty$.

Now we fix a non-negative integer k and let $Z_k(\check{C}(\mathcal{X}_n, r_n t))$ be the kth cycle group of $\check{C}(\mathcal{X}_n, r_n t)$, and let $B_k(\check{C}(\mathcal{X}_n, r_n t))$ be the kth boundary group of the same complex. Then $H_k(\check{C}(\mathcal{X}_n, r_n t)) := Z_k(\check{C}(\mathcal{X}_n, r_n t))/B_k(\check{C}(\mathcal{X}_n, r_n t))$ is the kth homology group, representing the elements of (non-trivial) k-dimensional cycles, which can be interpreted as the boundary of a $(k+1)$-dimensional body. The kth Betti number

$$\beta_{k,n}(t) := \dim H_k(\check{C}(\mathcal{X}_n, r_n t)) = \dim \frac{Z_k(\check{C}(\mathcal{X}_n, r_n t))}{B_k(\check{C}(\mathcal{X}_n, r_n t))}, \quad t \geq 0, \tag{3.2}$$

denotes the rank of $H_k(\check{C}(\mathcal{X}_n, r_n t))$. Loosely speaking, (3.2) counts the number of k-dimensional cycles in $\check{C}(\mathcal{X}_n, r_n t)$. Moreover, (3.2) can be extended to the kth persistent Betti number, defined by

$$\beta_{k,n}(s, t) := \dim \frac{Z_k(\check{C}(\mathcal{X}_n, r_n s))}{Z_k(\check{C}(\mathcal{X}_n, r_n s)) \cap B_k(\check{C}(\mathcal{X}_n, r_n t))}, \quad 0 \leq s \leq t < \infty. \tag{3.3}$$

More intuitively, (3.3) represents the number of k-dimensional cycles that appear in (3.1) before time s and remain alive at time t. Clearly $\beta_{k,n}(t, t)$ reduces to the ordinary Betti number

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

in (3.2). Readers wishing to have a more rigorous coverage of these algebraic topological notions may refer to [7], [11], and [26].

To provide a precise setup for the theorem below, we restrict the range of k to $\{1, \dots, d-1\}$, while taking $m \ge 1$ and $0 \le s_i \le t_i < \infty$ for $i = 1, \dots, m$. For $(x_1, \dots, x_{k+2}) \in (\mathbb{R}^d)^{k+2}$ and $r > 0$, we define

$$h_r(x_1, \dots, x_{k+2}) := \mathbb{1}\left\{ \left\{ \bigcap_{j=1, j\neq j_0}^{k+2} B(x_j, r/2) \neq \emptyset \text{ for all } j_0 \in \{1, \dots, k+2\} \right\} \cap \left\{ \bigcap_{j=1}^{k+2} B(x_j, r/2) = \emptyset \right\} \right\}. \tag{3.4}$$

Here (3.4) requires that a point set $\{x_1, \dots, x_{k+2}\}$ in $\mathbb{R}^d$ forms a single k-dimensional cycle with connectivity radius r. Furthermore, we let

$$H(\{x_1, \dots, x_{k+2}\}; \mathbf{s}, \mathbf{t}) := \left(h_{s_i}(x_1, \dots, x_{k+2}) h_{t_i}(x_1, \dots, x_{k+2})\right)_{i=1}^m, \tag{3.5}$$

where $\mathbf{s} = (s_1, \dots, s_m)$ and $\mathbf{t} = (t_1, \dots, t_m)$. It is then easy to check that H satisfies conditions **(H1)–(H4)**. The theorem below derives the exact rate (up to the scale) of a probability that the kth persistent Betti number becomes non-zero, when $n^{k+2} r_n^{d(k+1)} \to 0$ as $n \to \infty$.

Theorem 3.1. *Assume that $n^{k+2} r_n^{d(k+1)} \to 0$ as $n \to \infty$. Then, as $n \to \infty$, we have*

$$\begin{aligned}\left(n^{k+2} r_n^{d(k+1)}\right)^{-1} &\mathbb{P}\big((\beta_{k,n}(s_i, t_i),\ i = 1, \dots, m) \in \cdot\big) \\ &\xrightarrow{v} C_{k+2}\lambda\left\{\mathbf{y} \in (\mathbb{R}^d)^{k+1} : H(\{0, \mathbf{y}\}; \mathbf{s}, \mathbf{t}) \in \cdot\right\} \quad \text{in } M_+([0, \infty]^m \setminus \{\mathbf{0}\}),\end{aligned} \tag{3.6}$$

where $\mathbf{y} = (y_1, \dots, y_{k+1}) \in (\mathbb{R}^d)^{k+1}$, $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^m$, and C_{k+2} is given in Theorem 2.1.

Additionally, for $u_i \ge 0$, $u_i \neq 1$, $i = 1, \dots, m$, with $\max_{1 \le i \le m} u_i > 0$, we have, as $n \to \infty$,

$$\begin{aligned}&\frac{\mathbb{P}(\beta_{k,n}(s_i, t_i) \ge u_i,\ i = 1, \dots, m)}{\binom{n}{k+2}\mathbb{P}(h_{r_n}(X_1, \dots, X_{k+2}) = 1)} \\ &\to \left(\int_{(\mathbb{R}^d)^{k+1}} h_1(0, \mathbf{y})\, \mathrm{d}\mathbf{y}\right)^{-1} \lambda\left\{\mathbf{y} \in (\mathbb{R}^d)^{k+1} : h_{s_i}(0, \mathbf{y})\, h_{t_i}(0, \mathbf{y}) \ge u_i,\ i = 1, \dots, m\right\}.\end{aligned} \tag{3.7}$$

Note that the limit in (3.7) is equal to 0 whenever $\max_{1 \le i \le m} u_i > 1$. As a direct consequence of (3.7), we obtain that for every $a_i \in \{0, 1\}$, $i = 1, \dots, m$ with $\sum_{i=1}^m a_i \ge 1$,

$$\begin{aligned}&\frac{\mathbb{P}(\beta_{k,n}(s_i, t_i) = a_i,\ i = 1, \dots, m)}{\binom{n}{k+2}\mathbb{P}(h_{r_n}(X_1, \dots, X_{k+2}) = 1)} \\ &\to \frac{\int_{(\mathbb{R}^d)^{k+1}} \prod_{i=1}^m \{a_i h_{s_i}(0, \mathbf{y}) h_{t_i}(0, \mathbf{y}) + (1 - a_i)(1 - h_{s_i}(0, \mathbf{y}) h_{t_i}(0, \mathbf{y}))\}\, \mathrm{d}\mathbf{y}}{\int_{(\mathbb{R}^d)^{k+1}} h_1(0, \mathbf{y})\, \mathrm{d}\mathbf{y}}, \qquad n \to \infty.\end{aligned}$$

We observe that $h_{s_i}(0, \mathbf{y}) h_{t_i}(0, \mathbf{y}) = 1$ if and only if the point set $\{0, \mathbf{y}\} = \{0, y_1, \dots, y_{k+1}\} \in (\mathbb{R}^d)^{k+2}$ forms a single k-cycle before time s_i, such that this cycle is still alive at time t_i.

3.2. Volume of simplices

We next consider an application to the volume functional of simplices. Fix $d \geq 2$ and $1 \leq k \leq d$. For $(x_1, \ldots, x_{k+1}) \in (\mathbb{R}^d)^{k+1}$, let

$$[x_1, \ldots, x_{k+1}] = \left\{ \sum_{i=1}^{k+1} a_i x_i \colon a_i \geq 0, \ \sum_{i=1}^{k+1} a_i = 1 \right\}$$

be the k-simplex spanned by $x_1, \ldots, x_{k+1}$. Furthermore, $V_k([x_1, \ldots, x_{k+1}])$ denotes its k-dimensional volume. By slightly abusing notation, we write $V_k(\mathcal{Y}) = V_k([y_1, \ldots, y_{k+1}])$ for $\mathcal{Y} = (y_1, \ldots, y_{k+1}) \in (\mathbb{R}^d)^{k+1}$. The objective of this section is to explore the asymptotic behavior of

$$F_{k,n} := \left(\sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}|=k+1} V_k(\mathcal{Y})^{b_i} \mathbb{1}\left\{ \bigcap_{y \in \mathcal{Y}} B(y, r_n T_i/2) \neq \emptyset \right\} \right)_{i=1}^m,$$

where $b_i \geq 0$ and $T_i > 0$ for $i = 1, \ldots, m$. If one takes $b_i = 0$, the ith component of $F_{k,n}$ represents the k-simplex counts of a Čech complex $\check{C}(\mathcal{X}_n, r_n T_i)$. In the case of $b_i = 1$, the ith component of $F_{k,n}$ represents the total volume of these k-simplices. Furthermore, if $k = 1$ and $b_i = 1$, the ith component of $F_{1,n}$ is the total edge length in a random geometric graph with radius $r_n T_i$.

The corollary below investigates the probability that each component of the scaled $F_{k,n}$ exceeds a positive constant when $n^{k+1} r_n^{dk} \to 0$ as $n \to \infty$.

Corollary 3.1. *Assume that $n^{k+1} r_n^{dk} \to 0$ as $n \to \infty$. Then, as $n \to \infty$,*

$$\begin{aligned} &\left(n^{k+1} r_n^{dk}\right)^{-1} \mathbb{P}\left(\left(r_n^{-kb_i}\right)_{i=1}^m \circ F_{k,n} \in \cdot\right) \\ &\quad \xrightarrow{v} C_{k+1} \lambda\{\mathbf{y} \in (\mathbb{R}^d)^k \colon H(0, \mathbf{y}) \in \cdot\} \quad \text{in } M_+([0, \infty]^m \setminus \{\mathbf{0}\}), \end{aligned} \tag{3.8}$$

where $\mathbf{y} = (y_1, \ldots, y_k) \in (\mathbb{R}^d)^k$, $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^m$, and

$$H(0, \mathbf{y}) := \left(V_k([0, y_1, \ldots, y_k])^{b_i} \mathbb{1}\left\{ B(0, T_i/2) \cap \bigcap_{j=1}^k B(y_j, T_i/2) \neq \emptyset \right\} \right)_{i=1}^m. \tag{3.9}$$

Furthermore, for all $u_i > 0$, $i = 1, \ldots, m$, we have as $n \to \infty$,

$$\begin{aligned} &\frac{\mathbb{P}\left(\left(r_n^{-kb_i}\right)_{i=1}^m \circ F_{k,n} \in \prod_{i=1}^m [u_i, \infty)\right)}{\binom{n}{k+1} \mathbb{P}\left(\bigcap_{j=1}^{k+1} B(X_j, r_n/2) \neq \emptyset\right)} \\ &\to \frac{\lambda\left\{\mathbf{y} \in (\mathbb{R}^d)^k \colon V_k([0, \mathbf{y}])^{b_i} \geq u_i, \ B(0, T_i/2) \cap \bigcap_{j=1}^k B(y_j, T_i/2) \neq \emptyset, \ i = 1, \ldots, m\right\}}{\lambda\left\{\mathbf{y} \in (\mathbb{R}^d)^k \colon B(0, 1/2) \cap \bigcap_{j=1}^k B(y_j, 1/2) \neq \emptyset\right\}}. \end{aligned} \tag{3.10}$$

3.3. Morse critical points and values of min-type distance function

To understand the topology of random Čech complexes, Bobrowski and Adler [3] proposed an approach based on an extension of Morse theory to 'min-type' distance functions. For a

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

finite set $\mathcal{Z}$ of points in $\mathbb{R}^d$, we define a distance function $d_{\mathcal{Z}} : \mathbb{R}^d \to [0, \infty)$ by

$$d_{\mathcal{Z}}(x) := \min_{z \in \mathcal{Z}} \|x - z\|, \quad x \in \mathbb{R}^d. \tag{3.11}$$

Since $d_{\mathcal{Z}}$ is not differentiable, the classical definition of critical points does not apply to $d_{\mathcal{Z}}$. Nevertheless, one can still extend a notion of critical points, as well as their Morse critical index, to the min-type distance function as in (3.11) by means of an approach in [10]. More precisely, following the notations and definitions in [3], we say that $c \in \mathbb{R}^d$ is a critical point of $d_{\mathcal{Z}}$ with index $1 \le k \le d$, if there exists a set $\mathcal{Y} \subset \mathcal{Z}$ of $k+1$ points, such that:

(i) the points in $\mathcal{Y}$ are in general position,

(ii) $d_{\mathcal{Z}}(c) = \|c - y\|$ for all $y \in \mathcal{Y}$, while $d_{\mathcal{Z}}(c) < \|c - z\|$ for all $z \in \mathcal{Z} \setminus \mathcal{Y}$,

(iii) $c \in \mathrm{conv}^\circ(\mathcal{Y})$, where $\mathrm{conv}^\circ(\mathcal{Y})$ represents an interior of a convex hull spanned by the points in $\mathcal{Y}$.

By virtue of the nerve lemma (see e.g. [1, Theorem 10.7]), for each $r > 0$, the sublevel set $d_{\mathcal{X}_n}(-\infty, r]$ is homotopy equivalent to a Čech complex $\check{C}(\mathcal{X}_n, 2r)$. By the standard application of Morse theory as well as the nerve lemma, Bobrowski and Adler [3] justified that given a sequence $r_n \to 0$, $n \to \infty$, the number of critical points of $d_{\mathcal{X}_n}$ with index k, such that their critical values are less than r_n, behaves very similarly to $\beta_{k-1}(\check{C}(\mathcal{X}_n, 2r_n))$. A similar analysis was conducted in [36], in the case when a set of points are sampled from a stationary point process. Additionally, Bobrowski and Mukherjee [4] studied a more general case for which random points are supported on an ℓ-dimensional manifold $\mathcal{M} \subset \mathbb{R}^d$ $(\ell < d)$.

In this setting, we aim to study the asymptotic theory of

$$S_{k,n} := \left(\sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}|=k+1} R(\mathcal{Y})^{b_i} \mathbb{1}\{\gamma(\mathcal{Y}) \in \mathrm{conv}^\circ(\mathcal{Y}),\ R(\mathcal{Y}) \le r_n T_i,\ \mathcal{U}(\mathcal{Y}) \cap \mathcal{X}_n = \emptyset\} \right)_{i=1}^m,$$

where $b_i \ge 0$ and $T_i > 0$, $i = 1, \dots, m$. Moreover, $\gamma(\mathcal{Y})$ denotes a critical point of $d_{\mathcal{X}_n}$ with index k, generated by the points in $\mathcal{Y}$, $R(\mathcal{Y})$ is its critical value, and $\mathcal{U}(\mathcal{Y})$ is an open ball in $\mathbb{R}^d$ with radius $R(\mathcal{Y})$ centered at $\gamma(\mathcal{Y})$. If $b_i = 0$, the ith component of $S_{k,n}$ represents the number of critical points of index k with critical values less than $r_n T_i$. In the case of $b_i = 1$, the ith component of $S_{k,n}$ represents the sum of those critical values. The corollary below gives the rate of a probability that the appropriately scaled $S_{k,n}$ is asymptotically non-trivial when $n^{k+1} r_n^{dk} \to 0$ as $n \to \infty$.

Corollary 3.2. *Assume that $n^{k+1} r_n^{dk} \to 0$ as $n \to \infty$. Then, as $n \to \infty$,*

$$\begin{aligned} &\left(n^{k+1} r_n^{dk}\right)^{-1} \mathbb{P}\left(\left(r_n^{-b_i}\right)_{i=1}^m \circ S_{k,n} \in \cdot \right) \\ &\quad \xrightarrow{v} C_{k+1} \lambda\left\{ \mathbf{y} \in (\mathbb{R}^d)^k : H(0, \mathbf{y}) \in \cdot \right\} \quad \text{in } M_+([0, \infty]^m \setminus \{\mathbf{0}\}), \end{aligned} \tag{3.12}$$

where $\mathbf{y} = (y_1, \dots, y_k) \in (\mathbb{R}^d)^k$ and

$$\begin{aligned} H(0, \mathbf{y}) := \big(&R(0, y_1, \dots, y_k)^{b_i} \mathbb{1}\{\gamma(0, y_1, \dots, y_k) \in \mathrm{conv}^\circ(0, y_1, \dots, y_k), \\ &R(0, y_1, \dots, y_k) \le T_i\}\big)_{i=1}^m. \end{aligned} \tag{3.13}$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

Moreover, for all $0 < u_i \le T_i^{b_i}$, $i = 1, \dots, m$,

$$\frac{\mathbb{P}\big(\big(r_n^{-b_i}\big)_{i=1}^m \circ S_{k,n} \in \prod_{i=1}^m [u_i, \infty)\big)}{\binom{n}{k+1}\mathbb{P}(R(X_1, \dots, X_{k+1}) \le r_n)}$$
$$\to \frac{\lambda\big\{\mathbf{y} \in (\mathbb{R}^d)^k \colon \gamma(0, \mathbf{y}) \in \mathrm{conv}^{\circ}(0, \mathbf{y}),\ u_i^{1/b_i} \le R(0, \mathbf{y}) \le T_i,\ i = 1, \dots, m\big\}}{\lambda\{\mathbf{y} \in (\mathbb{R}^d)^k \colon R(0, \mathbf{y}) \le 1\}}. \tag{3.14}$$

4. Proofs

4.1. Proof of Theorem 2.1

The main machinery for our proof is a certain asymptotic result of point processes induced by the statistics in (2.6). More precisely, we consider the point processes

$$N_{k,n}^{(1)} := \sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}| = k} \delta_{H_n(\mathcal{Y})} \quad \text{and} \quad N_{k,n}^{(2)} := \sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}| = k} \delta_{G_n(\mathcal{Y}, \mathcal{X}_n; \mathbf{t})}, \tag{4.1}$$

where δ_z denotes the Dirac measure at $z \in \mathbb{R}^m$.

For the rigorous description of the asymptotic behavior of (4.1), we need the following concepts. The main references here are [13], [14], and [25]. Recall first that the vague topology on $M_p(E)$ is metrizable as a complete, separable metric space. The metric that induces the vague topology is called the vague metric, and its explicit form is given in the proof of Proposition 3.17 of [32]. Let $\emptyset \in M_p(E)$ be the null measure that assigns zeros to all Borel-measurable sets in E, and let $B_{\emptyset,r}$ denote an open ball of radius $r > 0$ centered at $\emptyset$ in the vague metric. Let $\mathcal{M}_0 = \mathcal{M}_0(M_p(E))$ denote the space of Borel measures on $M_p(E)$, the restriction of which to $M_p(E) \setminus B_{\emptyset,r}$ is finite for all $r > 0$. Moreover, define $\mathcal{C}_0 = \mathcal{C}_0(M_p(E))$ to be the space of continuous and bounded real-valued functions on $M_p(E)$ that vanish in the neighborhood of $\emptyset$. Given $\eta_n, \eta \in \mathcal{M}_0$, we say that η_n converges to η in the $\mathcal{M}_0$-topology, denoted by $\eta_n \to \eta$ in $\mathcal{M}_0$, if it holds that

$$\int_{M_p(E)} g(\xi)\eta_n(\mathrm{d}\xi) \to \int_{M_p(E)} g(\xi)\eta(\mathrm{d}\xi) \quad \text{for all } g \in \mathcal{C}_0.$$

The proposition below reveals the required asymptotics of (4.1). The result may be of independent interest. It can actually parallel Theorem 4.1 of [14] and Theorems 3.1 and 4.1 of [8], the authors of which studied large deviations for point processes based on a stationary sequence with heavy-tailed marginals and non-trivial dependency. As in the case of Theorem 2.1, the limits of $N_{k,n}^{(i)}$, $i = 1, 2$ coincide with one another, due to the fact that the indicator c_n at (2.4) tends to 1 as $n \to \infty$.

Proposition 4.1. *Under the assumptions in Theorem 2.1, for each* $i = 1, 2$, *we have*

$$\big(n^k r_n^{d(k-1)}\big)^{-1}\mathbb{P}\big(N_{k,n}^{(i)} \in \cdot\big) \to C_k \lambda\big\{\mathbf{y} \in (\mathbb{R}^d)^{k-1} \colon \delta_{H(0,\mathbf{y})} \in \cdot\big\} \quad \textit{in } \mathcal{M}_0, \quad n \to \infty. \tag{4.2}$$

Before commencing the proof, we observe that $M_p(E)$ is not locally compact; thus, unlike Theorem 2.1, the convergence in Proposition 4.1 cannot be treated in terms of vague topology. In contrast, the theory of $\mathcal{M}_0$-topology requires only that the underlying space be complete

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

and separable. Since $M_p(E)$ is complete and separable (see Proposition 3.17 in [32]), one can exploit $\mathcal{M}_0$-topology as an appropriate topology for the convergence in Proposition 4.1.

Proof of Proposition 4.1. Since the proofs of the two statements are very similar in nature, we prove the case $i=2$ only. Given $U_1, U_2 \in C_K^+(E)$ and $\epsilon_1, \epsilon_2 > 0$, define $F_{U_1,U_2,\epsilon_1,\epsilon_2}: M_p(E) \to [0, 1]$ by

$$F_{U_1,U_2,\epsilon_1,\epsilon_2}(\xi) = \big(1 - \mathrm{e}^{-(\xi(U_1)-\epsilon_1)_+}\big)\big(1 - \mathrm{e}^{-(\xi(U_2)-\epsilon_2)_+}\big),$$

where $(a)_+ = a$ if $a \ge 0$ and 0 otherwise, and

$$\xi(U_\ell) = \int_E U_\ell(x)\xi(\mathrm{d}x) \quad \text{for } \ell = 1, 2.$$

It is elementary to check that $F_{U_1,U_2,\epsilon_1,\epsilon_2} \in \mathcal{C}_0$.

For ease of description, we introduce several shorthand notations: for $\ell \ge 1$ and $n \ge 1$, let

$$\mathcal{I}_{\ell,n} = \{\mathbf{i} = (i_1, \dots, i_\ell) \in \mathbb{N}^\ell : 1 \le i_1 < \cdots < i_\ell \le n\} \tag{4.3}$$

be the collection of ordered ℓ-tuples of positive integers. Given a random sample $\mathcal{X}_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^d$, we write

$$\mathcal{X}_{\mathbf{i}} = (X_{i_1}, \dots, X_{i_\ell}), \quad \mathbf{i} = (i_1, \dots, i_\ell) \in \mathcal{I}_{\ell,n}. \tag{4.4}$$

Using these notations, we denote

$$\eta_n(\cdot) := \big(n^k r_n^{d(k-1)}\big)^{-1}\mathbb{P}\big(N_{k,n}^{(2)} \in \cdot\big) = \big(n^k r_n^{d(k-1)}\big)^{-1}\mathbb{P}\Bigg(\sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \delta_{G_n(\mathcal{X}_{\mathbf{i}},\mathcal{X}_n;\,\mathbf{t})} \in \cdot\Bigg), \tag{4.5}$$

$$\eta(\cdot) := C_k\lambda\big\{\mathbf{y} \in (\mathbb{R}^d)^{k-1} : \delta_{H(0,\mathbf{y})} \in \cdot\big\}. \tag{4.6}$$

According to Theorem A.2 in [14], the required statement follows if one can show that

$$\eta_n(F_{U_1,U_2,\epsilon_1,\epsilon_2}) \to \eta(F_{U_1,U_2,\epsilon_1,\epsilon_2}) \quad \text{as } n \to \infty,$$

for every $U_1, U_2 \in C_K^+(E)$ and $\epsilon_1, \epsilon_2 > 0$. For each $\ell = 1, 2$, U_ℓ has compact support in E, so there exists $\zeta > 0$ such that

$$\bigcup_{\ell=1}^{2} \mathrm{supp}(U_\ell) \cap \mathbb{R}^m \subset \{x \in \mathbb{R}^m : \|x\| > \zeta\} \tag{4.7}$$

($\mathrm{supp}(U_\ell)$ denotes the support of U_ℓ). Define

$$\Theta_n := \prod_{\ell=1}^{2}\Bigg(1 - \exp\Bigg\{-\Bigg[\sum_{\mathbf{i}\in\mathcal{I}_{k,n}} U_\ell(G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t})) - \epsilon_\ell\Bigg]_+\Bigg\}\Bigg);$$

then we have

$$\begin{aligned}\eta_n(F_{U_1,U_2,\epsilon_1,\epsilon_2}) &= \big(n^k r_n^{d(k-1)}\big)^{-1}\mathbb{E}[\Theta_n] \\ &= \big(n^k r_n^{d(k-1)}\big)^{-1}\mathbb{E}\Bigg[\Theta_n \mathbb{1}\Bigg\{\bigcup_{\mathbf{i}\in\mathcal{I}_{k,n}} \{\|G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t})\| > \zeta\}\Bigg\}\Bigg]\end{aligned}$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

$$= \left(n^k r_n^{d(k-1)}\right)^{-1} \mathbb{E}\left[\Theta_n \mathbb{1}\left\{\bigcup_{\mathbf{i}\in\mathcal{I}_{k,n}} \{\|G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\| > \zeta,\right.\right.$$

$$\left.\left. \|G_n(\mathcal{X}_\mathbf{j}, \mathcal{X}_n; \mathbf{t})\| \le \zeta \text{ for all } \mathbf{j}\in\mathcal{I}_{k,n} \text{ with } \mathbf{j}\ne\mathbf{i}\}\right\}\right]$$

$$+ \left(n^k r_n^{d(k-1)}\right)^{-1} \mathbb{E}\left[\Theta_n \mathbb{1}\left\{\bigcup_{\mathbf{i}\in\mathcal{I}_{k,n}} \bigcup_{\substack{\mathbf{j}\in\mathcal{I}_{k,n},\\ \mathbf{j}\ne\mathbf{i}}} \{\|G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\| > \zeta,\ \|G_n(\mathcal{X}_\mathbf{j}, \mathcal{X}_n; \mathbf{t})\| > \zeta\}\right\}\right]$$

$$=: A_n + B_n. \tag{4.8}$$

We first show that B_n tends to 0 as $n\to\infty$. Since $0\le\Theta_n\le 1$ and $\|G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\| \le \|H_n(\mathcal{X}_\mathbf{i})\|$ for all $\mathbf{i}\in\mathcal{I}_{k,n}$, we have

$$B_n \le \left(n^k r_n^{d(k-1)}\right)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \sum_{\substack{\mathbf{j}\in\mathcal{I}_{k,n},\\ \mathbf{j}\ne\mathbf{i}}} \mathbb{P}(\|H_n(\mathcal{X}_\mathbf{i})\| > \zeta,\ \|H_n(\mathcal{X}_\mathbf{j})\| > \zeta)$$

$$= \left(n^k r_n^{d(k-1)}\right)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \sum_{\substack{\mathbf{j}\in\mathcal{I}_{k,n},\\ |\mathbf{i}\cap\mathbf{j}|=0}} \mathbb{P}(\|H_n(\mathcal{X}_\mathbf{i})\| > \zeta)\,\mathbb{P}(\|H_n(\mathcal{X}_\mathbf{j})\| > \zeta)$$

$$+ \left(n^k r_n^{d(k-1)}\right)^{-1} \sum_{\ell=1}^{k-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \sum_{\substack{\mathbf{j}\in\mathcal{I}_{k,n},\\ |\mathbf{i}\cap\mathbf{j}|=\ell}} \mathbb{P}(\|H_n(\mathcal{X}_\mathbf{i})\| > \zeta,\ \|H_n(\mathcal{X}_\mathbf{j})\| > \zeta)$$

$$\le \left(n^k r_n^{d(k-1)}\right)^{-1} n^{2k} \mathbb{P}(\|H_n(X_1,\ldots,X_k)\| > \zeta)^2$$

$$+ \left(n^k r_n^{d(k-1)}\right)^{-1} \sum_{\ell=1}^{k-1} n^{2k-\ell} \mathbb{P}(\|H_n(X_1,\ldots,X_k)\| > \zeta,$$

$$\|H_n(X_1,\ldots,X_\ell,X_{k+1},\ldots,X_{2k-\ell})\| > \zeta)$$

$$= \left(n^k r_n^{d(k-1)}\right)^{-1} n^{2k} \left(\int_{(\mathbb{R}^d)^k} \mathbb{1}\{\|H_n(x_1,\ldots,x_k)\| > \zeta\} \prod_{i=1}^k f(x_i)\,\mathrm{d}\mathbf{x}\right)^2$$

$$+ \left(n^k r_n^{d(k-1)}\right)^{-1} \sum_{\ell=1}^{k-1} n^{2k-\ell} \int_{(\mathbb{R}^d)^{2k-\ell}} \mathbb{1}\{\|H_n(x_1,\ldots,x_k)\| > \zeta,$$

$$\|H_n(x_1,\ldots,x_\ell,x_{k+1},\ldots,x_{2k-\ell})\| > \zeta\} \prod_{i=1}^{2k-\ell} f(x_i)\,\mathrm{d}\mathbf{x}$$

$$=: C_n + D_n.$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

Performing the change of variables by $x_i = x + r_n y_{i-1}$, $i = 1, \ldots, k$ (with $y_0 \equiv 0$) together with the translation invariance of H as well as (2.1),

$$C_n = \big(n^k r_n^{d(k-1)}\big)^{-1} n^{2k} \Bigg(r_n^{d(k-1)} \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{k-1}} \mathbb{1}\{\|H(0, y_1, \ldots, y_{k-1})\| > \zeta\} \times f(x) \prod_{i=1}^{k-1} f(x + r_n y_i)\, \mathrm{d}\mathbf{y}\, \mathrm{d}x \Bigg)^2$$

$$\le (\|f\|_\infty)^{2k-2} n^k r_n^{d(k-1)} \bigg(\int_{(\mathbb{R}^d)^{k-1}} \mathbb{1}\{\|H(0, y_1, \ldots, y_{k-1})\| > \zeta\}\, \mathrm{d}\mathbf{y} \bigg)^2.$$

By property **(H3)** of H, the integral in the last term is finite. Since $n^k r_n^{d(k-1)} \to 0$ as $n \to \infty$, we obtain $C_n \to 0$, $n \to \infty$. Next, turning to D_n, we change the variables by $x_i = x + r_n y_{i-1}$, $i = 1, \ldots, 2k - \ell$ (with $y_0 \equiv 0$), to obtain that

$$D_n = \sum_{\ell=1}^{k-1} \big(n^k r_n^{d(k-1)}\big)^{-1} n^{2k-\ell} r_n^{d(2k-\ell-1)} \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{2k-\ell-1}} \mathbb{1}\{\|H(0, y_1, \ldots, y_{k-1})\| > \zeta,$$
$$\|H(0, y_1, \ldots, y_{\ell-1}, y_k, \ldots, y_{2k-\ell-1})\| > \zeta\} f(x) \prod_{i=1}^{2k-\ell-1} f(x + r_n y_i)\, \mathrm{d}\mathbf{y}\, \mathrm{d}x$$

$$\le \sum_{\ell=1}^{k-1} (n r_n^d)^{k-\ell} (\|f\|_\infty)^{2k-\ell-1} \int_{(\mathbb{R}^d)^{2k-\ell-1}} \mathbb{1}\{\|H(0, y_1, \ldots, y_{k-1})\| > \zeta,$$
$$\|H(0, y_1, \ldots, y_{\ell-1}, y_k, \ldots, y_{2k-\ell-1})\| > \zeta\}\, \mathrm{d}\mathbf{y}.$$

By property **(H3)** of H, the integral in the last term is again finite. As $n r_n^d \to 0$, $n \to \infty$, we can obtain $D_n \to 0$, $n \to \infty$, which concludes that $B_n \to 0$, $n \to \infty$, as desired.

Returning to A_n in (4.8), we observe that

$$\big(\{\|G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t})\| > \zeta,\ \|G_n(\mathcal{X}_{\mathbf{j}}, \mathcal{X}_n; \mathbf{t})\| \le \zeta \text{ for all } \mathbf{j} \in \mathcal{I}_{k,n} \text{ with } \mathbf{j} \ne \mathbf{i}\},\ \mathbf{i} \in \mathcal{I}_{k,n}\big)$$

are disjoint. Hence we can see from (4.7) that

$$A_n := \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i} \in \mathcal{I}_{k,n}} \mathbb{E}\big[\Theta_n \mathbb{1}\{\|G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t})\| > \zeta,$$
$$\|G_n(\mathcal{X}_{\mathbf{j}}, \mathcal{X}_n; \mathbf{t})\| \le \zeta \text{ for all } \mathbf{j} \in \mathcal{I}_{k,n} \text{ with } \mathbf{j} \ne \mathbf{i}\}\big]$$

$$= \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i} \in \mathcal{I}_{k,n}} \mathbb{E}\Bigg[\prod_{\ell=1}^{2} (1 - \exp\{-(U_\ell(G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t})) - \epsilon_\ell)_+\})$$
$$\times \mathbb{1}\{\|G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t})\| > \zeta,\ \|G_n(\mathcal{X}_{\mathbf{j}}, \mathcal{X}_n; \mathbf{t})\| \le \zeta \text{ for all } \mathbf{j} \in \mathcal{I}_{k,n} \text{ with } \mathbf{j} \ne \mathbf{i}\} \Bigg]$$

$$
\begin{aligned}
&= \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\Bigg[\prod_{\ell=1}^{2} (1-\exp\{-(U_\ell(G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})) - \epsilon_\ell)_+\})\Bigg] \\
&\quad - \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\Bigg[\prod_{\ell=1}^{2} (1-\exp\{-(U_\ell(G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})) - \epsilon_\ell)_+\}) \\
&\quad \times \mathbb{1}\Bigg\{\bigcup_{\mathbf{j}\in\mathcal{I}_{k,n}, \mathbf{j}\neq\mathbf{i}} \{\|G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\| > \zeta,\ \|G_n(\mathcal{X}_\mathbf{j}, \mathcal{X}_n; \mathbf{t})\| > \zeta\}\Bigg\}\Bigg]. \\
&=: E_n - F_n,
\end{aligned}
$$

Repeating the same argument as that for proving $B_n \to 0$, $n\to\infty$, it is not hard to see that $F_n \to 0$ as $n\to\infty$. Assuming without loss of generality that $0 \le t_1 \le \cdots \le t_m < \infty$, we divide E_n into two terms:

$$
\begin{aligned}
E_n &= \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\Bigg[\prod_{\ell=1}^{2} (1-\exp\{-(U_\ell(G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})) - \epsilon_\ell)_+\})\, c_{n,m}(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\Bigg] \\
&\quad + \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\Bigg[\prod_{\ell=1}^{2} (1-\exp\{-(U_\ell(G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})) - \epsilon_\ell)_+\}) \\
&\qquad\qquad \times (1 - c_{n,m}(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t}))\Bigg] \\
&=: I_n + J_n,
\end{aligned}
\tag{4.9}
$$

where $c_{n,m}(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})$ denotes the mth element of $c_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})$ (see (2.4)). Of the last two terms, we show that J_n is negligible as $n\to\infty$. Indeed, by (4.7) and $\|G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\| \le \|H_n(\mathcal{X}_\mathbf{i})\|$, we see that

$$
J_n \le \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\big[\mathbb{1}\{\|H_n(\mathcal{X}_\mathbf{i})\| > \zeta\}(1 - c_{n,m}(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t}))\big]. \tag{4.10}
$$

Then the right-hand side of (4.10) is equal to

$$
\begin{aligned}
&\big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\big[\mathbb{1}\{\|H_n(\mathcal{X}_\mathbf{i})\| > \zeta\}\, \mathbb{E}[1 - c_{n,m}(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t}) \mid \mathcal{X}_\mathbf{i}]\big] \qquad (4.11) \\
&= \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\Bigg[\mathbb{1}\{\|H_n(\mathcal{X}_\mathbf{i})\| > \zeta\}\Bigg(1 - \Bigg(1 - \int_{\mathcal{B}(\mathcal{X}_\mathbf{i}; r_n t_m)} f(z)\,\mathrm{d}z\Bigg)^{n-k}\Bigg)\Bigg] \\
&= \big(n^k r_n^{d(k-1)}\big)^{-1} \binom{n}{k} \int_{(\mathbb{R}^d)^k} \mathbb{1}\{\|H_n(x_1, \ldots, x_k)\| > \zeta\} \\
&\quad \times \Bigg(1 - \Bigg(1 - \int_{\mathcal{B}(\{x_1,\ldots,x_k\}; r_n t_m)} f(z)\,\mathrm{d}z\Bigg)^{n-k}\Bigg) \prod_{i=1}^{k} f(x_i)\,\mathrm{d}\mathbf{x}.
\end{aligned}
$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

In the above, $\mathcal{B}(\mathcal{X}_{\mathbf{i}}; r_n t_m)$ represents the union of balls of radius $r_n t_m$ around the points in $\mathcal{X}_{\mathbf{i}} = \{X_{i_1}, \ldots, X_{i_k}\}$, that is,

$$\mathcal{B}(\mathcal{X}_{\mathbf{i}}; r_n t_m) := \bigcup_{\ell=1}^{k} B(X_{i_\ell}, r_n t_m).$$

By the change of variables $x_i = x + r_n y_{i-1}$, $i = 1, \ldots, k$ (with $y_0 \equiv 0$) and the translation invariance of H, the last expression in (4.11) becomes

$$\begin{aligned}
&\big(n^k r_n^{d(k-1)}\big)^{-1} \binom{n}{k} r_n^{d(k-1)} \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{k-1}} \mathbb{1}\{\|H(0, \mathbf{y})\| > \zeta\} \\
&\quad \times \bigg(1 - \bigg(1 - \int_{\mathcal{B}(\{x, x+r_n y_1, \ldots, x+r_n y_{k-1}\}; r_n t_m)} f(z)\,\mathrm{d}z\bigg)^{n-k}\bigg) f(x) \prod_{i=1}^{k-1} f(x + r_n y_i)\,\mathrm{d}\mathbf{y}\,\mathrm{d}x \\
&\le \frac{(\|f\|_\infty)^{k-1}}{k!} \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{k-1}} \mathbb{1}\{\|H(0, \mathbf{y})\| > \zeta\} \\
&\quad \times \bigg(1 - \bigg(1 - \int_{\mathcal{B}(\{x, x+r_n y_1, \ldots, x+r_n y_{k-1}\}; r_n t_m)} f(z)\,\mathrm{d}z\bigg)^{n-k}\bigg) f(x)\,\mathrm{d}\mathbf{y}\,\mathrm{d}x, \qquad (4.12)
\end{aligned}$$

where $\mathbf{y} = (y_1, \ldots, y_{k-1}) \in (\mathbb{R}^d)^{k-1}$. For every $x \in \mathbb{R}^d$ and $\mathbf{y} = (y_1, \ldots, y_{k-1}) \in (\mathbb{R}^d)^{k-1}$,

$$\begin{aligned}
&\lim_{n\to\infty} \bigg(1 - \int_{\mathcal{B}(\{x, x+r_n y_1, \ldots, x+r_n y_{k-1}\}; r_n t_m)} f(z)\,\mathrm{d}z\bigg)^{n-k} \\
&= \lim_{n\to\infty} \bigg(1 - r_n^d \int_{\mathcal{B}(\{0, \mathbf{y}\}; t_m)} f(x + r_n v)\,\mathrm{d}v\bigg)^{n-k} \\
&= \exp\bigg\{- \lim_{n\to\infty} (n-k) r_n^d \int_{\mathcal{B}(\{0, \mathbf{y}\}; t_m)} f(x + r_n v)\,\mathrm{d}v\bigg\},
\end{aligned}$$

so that

$$n r_n^d \int_{\mathcal{B}(\{0, \mathbf{y}\}; t_m)} f(x + r_n v)\,\mathrm{d}v \le n r_n^d \|f\|_\infty \int_{\mathcal{B}(\{0, \mathbf{y}\}; t_m)} \mathrm{d}v \to 0, \quad n \to \infty.$$

Hence we have obtained

$$\bigg(1 - \int_{\mathcal{B}(\{x, x+r_n y_1, \ldots, x+r_n y_{k-1}\}; r_n t_m)} f(z)\,\mathrm{d}z\bigg)^{n-k} \to 1, \quad n \to \infty. \qquad (4.13)$$

Now the dominated convergence theorem, as well as property **(H3)** of H, ensures that the last expression in (4.12) goes to 0 as $n \to \infty$. Thus $J_n \to 0$ as $n \to \infty$.

From all of the convergence results derived thus far, we have $\eta_n(F_{U_1, U_2, \epsilon_1, \epsilon_2}) = I_n + \mathrm{o}(1)$ as $n \to \infty$. We note that if $c_{n,m}(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t}) = 1$, then all the other elements in $c_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t})$ are equal to 1, and therefore $G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t}) = H_n(\mathcal{X}_{\mathbf{i}})$. By the conditioning on $\mathcal{X}_{\mathbf{i}}$ as in (4.11),

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

as well as the same change of variables as in (4.12), we can see that

$$I_n = \left(n^k r_n^{d(k-1)}\right)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\left[\prod_{\ell=1}^{2} (1-\exp\{-(U_\ell(H_n(\mathcal{X}_\mathbf{i}))-\epsilon_\ell)_+\})c_{n,m}(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\right]$$

$$= \left(n^k r_n^{d(k-1)}\right)^{-1} \binom{n}{k} r_n^{d(k-1)} \int_{\mathbb{R}^d}\int_{(\mathbb{R}^d)^{k-1}} \prod_{\ell=1}^{2} (1-\exp\{-(U_\ell(H(0,\mathbf{y}))-\epsilon_\ell)_+\})$$

$$\times \left(1 - \int_{\mathcal{B}(\{x, x+r_ny_1, \dots, x+r_ny_{k-1}\}; r_nt_m)} f(z)\,\mathrm{d}z\right)^{n-k} f(x) \prod_{i=1}^{k-1} f(x+r_ny_i)\,\mathrm{d}\mathbf{y}\,\mathrm{d}x. \tag{4.14}$$

By the continuity of f, it holds that

$$\prod_{i=1}^{k-1} f(x+r_ny_i) \to f(x)^{k-1} \quad \text{as } n\to\infty \text{ a.e.}$$

Furthermore, the integrand in (4.14) is bounded above by $f(x)\|f\|_\infty^{k-1}\mathbb{1}\{\|H(0,\mathbf{y})\| > \zeta\}$, which is clearly integrable in $(x, \mathbf{y}) \in (\mathbb{R}^d)^k$. Therefore the dominated convergence theorem gives that as $n\to\infty$,

$$I_n \to C_k \int_{(\mathbb{R}^d)^{k-1}} \prod_{\ell=1}^{2} (1-\exp\{-(U_\ell(H(0,\mathbf{y}))-\epsilon_\ell)_+\})\,\mathrm{d}\mathbf{y} = \eta(F_{U_1,U_2,\epsilon_1,\epsilon_2}). \qquad \square$$

Proof of Theorem 2.1. As in the case of Proposition 4.1, the proofs of the two statements are similar in nature, so we again prove the case $i=2$ only. Let $0<\epsilon<1$, and define $V_\epsilon : E\to E$ by

$$V_\epsilon(z_1, \dots, z_m) = (z_1\mathbb{1}\{|z_1|\ge\epsilon\}, \dots, z_m\mathbb{1}\{|z_m|\ge\epsilon\}).$$

Next we define a map $T_{V_\epsilon} : M_p(E)\to E$ by

$$T_{V_\epsilon}\left(\sum_j \delta_{x_j}\right) = \sum_j V_\epsilon(x_j).$$

Below we only consider $\epsilon\in(0,1)$, so that

$$\lambda\left\{\mathbf{y}\in(\mathbb{R}^d)^{k-1} : \|H(0,\mathbf{y})\| = \epsilon\right\} = 0. \tag{4.15}$$

Note that (4.15) holds except at most countably many $\epsilon\in(0,1)$. Now we claim that

$$\eta_n\circ T_{V_\epsilon}^{-1} \xrightarrow{v} \eta\circ T_{V_\epsilon}^{-1} \quad \text{in } M_+(E) \quad \text{as } n\to\infty, \tag{4.16}$$

where η_n and η are defined at (4.5) and (4.6) respectively. Equivalently, we aim to show that

$$\int_{M_p(E)} F(T_{V_\epsilon}(\xi))\eta_n(\mathrm{d}\xi) \to \int_{M_p(E)} F(T_{V_\epsilon}(\xi))\eta(\mathrm{d}\xi)$$

for every $F\in C_K^+(E)$. To show this, by [32, Proposition 3.12] it suffices to verify that

$$\eta_n\circ T_{V_\epsilon}^{-1}(A) \to \eta\circ T_{V_\epsilon}^{-1}(A)$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

for all relatively compact sets $A \subset E$ with $\eta \circ T_{V_\epsilon}^{-1}(\partial A) = 0$, where ∂A denotes the boundary of A. According to [13, Theorem 2.4] along with (4.2), we must show

$$\eta\big(\partial T_{V_\epsilon}^{-1}(A)\big) = 0 \quad \text{and} \quad \emptyset \notin \overline{T_{V_\epsilon}^{-1}(A)}, \tag{4.17}$$

where $\emptyset \in M_p(E)$ is the null measure and $\overline{B}$ denotes the closure of B. For the proof of the first requirement in (4.17), it is elementary to check that

$$\partial T_{V_\epsilon}^{-1}(A) \subset T_{V_\epsilon}^{-1}(\partial A) \cup \mathcal{D}_{T_{V_\epsilon}}, \tag{4.18}$$

where $\mathcal{D}_{T_{V_\epsilon}}$ is the collection of $\xi \in M_p(E)$ such that T_{V_ϵ} is discontinuous at ξ. It then follows from (4.15) and (4.18) that

$$\eta\big(\partial T_{V_\epsilon}^{-1}(A)\big) \le \eta(\mathcal{D}_{T_{V_\epsilon}}) \le C_k \lambda\big\{\mathbf{y} \in (\mathbb{R}^d)^{k-1} : \|H(0, \mathbf{y})\| = \epsilon\big\} = 0.$$

Next, suppose for contradiction that $\emptyset \in \overline{T_{V_\epsilon}^{-1}(A)}$. Then there exists a sequence $(\xi_n) \subset T_{V_\epsilon}^{-1}(A)$ such that $\xi_n \xrightarrow{v} \emptyset$ in $M_p(E)$. Since T_{V_ϵ} is continuous at $\emptyset$ with $T_{V_\epsilon}(\emptyset) = \mathbf{0} = (0, \dots, 0) \in \mathbb{R}^m$, we have $T_{V_\epsilon}(\xi_n) \to \mathbf{0}$ as $n \to \infty$. This implies $\mathbf{0} \in \overline{A}$, which however contradicts the relative compactness of A in E.

For ease of description, using the notations in (4.3) and (4.4) we denote

$$\widetilde{\eta}_n(\cdot) := \big(n^k r_n^{d(k-1)}\big)^{-1} \mathbb{P}\Bigg(\sum_{\mathbf{i} \in \mathcal{I}_{k,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{t}) \in \cdot\Bigg),$$

$$\widetilde{\eta}(\cdot) := C_k \lambda\big\{\mathbf{y} \in (\mathbb{R}^d)^{k-1} : H(0, \mathbf{y}) \in \cdot\big\}.$$

Then the entire proof will be completed if we can verify that

$$\widetilde{\eta}_n(F) \to \widetilde{\eta}(F), \quad n \to \infty,$$

for every $F \in C_K^+(E)$. To begin, we bound $|\widetilde{\eta}_n(F) - \widetilde{\eta}(F)|$ as follows:

$$|\widetilde{\eta}_n(F) - \widetilde{\eta}(F)| \le |\widetilde{\eta}_n(F) - \eta_n \circ T_{V_\epsilon}^{-1}(F)| + |\eta_n \circ T_{V_\epsilon}^{-1}(F) - \eta \circ T_{V_\epsilon}^{-1}(F)| + |\eta \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}(F)|.$$

Because of (4.16), we have

$$\limsup_{n\to\infty} |\widetilde{\eta}_n(F) - \widetilde{\eta}(F)| \le \limsup_{n\to\infty} |\widetilde{\eta}_n(F) - \eta_n \circ T_{V_\epsilon}^{-1}(F)| + |\eta \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}(F)|.$$

It thus remains to demonstrate that

$$\lim_{\epsilon\to 0} |\eta \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}(F)| = 0, \tag{4.19}$$

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} |\eta_n \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}_n(F)| = 0. \tag{4.20}$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

For the proof of (4.19), note that there exists $\delta_0 > 0$ so that $\mathrm{supp}(F) \cap \mathbb{R}^m \subset \{x \in \mathbb{R}^m : \|x\| > \delta_0\}$. We also fix a constant $\delta' \in (0, \delta_0/2)$. Then, for every $0 < \delta < \delta_0/2$, we have

$$\begin{aligned}
|\eta \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}(F)| = C_k \left| \int_{(\mathbb{R}^d)^{k-1}} F(V_\epsilon(H(0, \mathbf{y}))) - F(H(0, \mathbf{y}))\, \mathrm{d}\mathbf{y} \right| \\
\le C_k \int_{\|V_\epsilon(H(0,\mathbf{y})) - H(0,\mathbf{y})\| > \delta} |F(V_\epsilon(H(0, \mathbf{y}))) - F(H(0, \mathbf{y}))|\, \mathrm{d}\mathbf{y} \\
+ C_k \int_{\|V_\epsilon(H(0,\mathbf{y})) - H(0,\mathbf{y})\| \le \delta} |F(V_\epsilon(H(0, \mathbf{y}))) - F(H(0, \mathbf{y}))|\, \mathrm{d}\mathbf{y} \\
=: A_n + B_n.
\end{aligned}$$

Since F is bounded, it follows from the dominated convergence theorem and property **(H3)** of H that

$$A_n \le 2\|F\|_\infty C_k \int_{(\mathbb{R}^d)^{k-1}} \mathbb{1}\{\|V_\epsilon(H(0, \mathbf{y})) - H(0, \mathbf{y})\| > \delta\}\, \mathrm{d}\mathbf{y} \to 0 \quad \text{as } \epsilon \to 0.$$

Next, turning our attention to B_n, we can see that

$$B_n = C_k \int_{\|V_\epsilon(H(0,\mathbf{y})) - H(0,\mathbf{y})\| \le \delta,\, \|H(0,\mathbf{y})\| > \delta'} |F(V_\epsilon(H(0, \mathbf{y}))) - F(H(0, \mathbf{y}))|\, \mathrm{d}\mathbf{y}. \tag{4.21}$$

To show this, suppose that $\|H(0, \mathbf{y})\| \le \delta' < \delta_0/2$. Then $F(H(0, \mathbf{y})) = 0$, and further,

$$\|V_\epsilon(H(0, \mathbf{y}))\| \le \|V_\epsilon(H(0, \mathbf{y})) - H(0, \mathbf{y})\| + \|H(0, \mathbf{y})\| \le \delta + \delta' < \frac{\delta_0}{2} + \frac{\delta_0}{2} = \delta_0,$$

which implies that $F(V_\epsilon(H(0, \mathbf{y}))) = 0$. From (4.21) we have

$$B_n \le C_k \omega_F(\delta) \lambda \big\{ \mathbf{y} \in (\mathbb{R}^d)^{k-1} : \|H(0, \mathbf{y})\| > \delta' \big\},$$

where

$$\omega_F(\delta) := \sup_{x, y \in \mathbb{R}^m,\, \|x - y\| \le \delta} |F(x) - F(y)|$$

is the modulus of continuity of F.

Combining all of these results, we conclude that

$$\lim_{\epsilon \to 0} |\eta \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}(F)| \le C_k \omega_F(\delta) \lambda \big\{ \mathbf{y} \in (\mathbb{R}^d)^{k-1} : \|H(0, \mathbf{y})\| > \delta' \big\}$$

for all $0 < \delta < \delta_0/2$. Finally, letting $\delta \downarrow 0$, we find that $\lim_{\epsilon \to 0} |\eta \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}(F)| = 0$ as F is uniformly continuous on $\mathbb{R}^m$.

Next, let us proceed to the proof of (4.20). We fix δ_0 and δ' in the same way as above. Then, for $0 < \delta < \delta_0/2$,

$$\begin{aligned}
&|\eta_n \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}_n(F)| \\
&= \big(n^k r_n^{d(k-1)}\big)^{-1} \left| \mathbb{E}\left[F\left(\sum_{\mathbf{i} \in \mathcal{I}_{k,n}} V_\epsilon(G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})) \right) - F\left(\sum_{\mathbf{i} \in \mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t}) \right) \right] \right| \\
&\le \big(n^k r_n^{d(k-1)}\big)^{-1} \mathbb{E}\left[\left| F\left(\sum_{\mathbf{i} \in \mathcal{I}_{k,n}} V_\epsilon(G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})) \right) - F\left(\sum_{\mathbf{i} \in \mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t}) \right) \right| \right.
\end{aligned}$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

$$\times\, \mathbb{1}\left\{\left\| \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} V_\epsilon(G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})) - \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t}) \right\| > \delta \right\}\Bigg]$$

$$+\, \big(n^k r_n^{d(k-1)}\big)^{-1}\mathbb{E}\Bigg[\Bigg| F\Bigg(\sum_{\mathbf{i}\in\mathcal{I}_{k,n}} V_\epsilon(G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t}))\Bigg) - F\Bigg(\sum_{\mathbf{i}\in\mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})\Bigg)\Bigg|$$

$$\times\, \mathbb{1}\left\{\left\| \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} V_\epsilon(G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})) - \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t}) \right\| \le \delta \right\}\Bigg]$$

$$=: C_n + D_n.$$

Noting that F is bounded, while assuming without loss of generality that $0 \le t_1 \le \cdots \le t_m < \infty$, we can bound C_n as follows:

$$C_n \le 2\|F\|_\infty \big(n^k r_n^{d(k-1)}\big)^{-1}\mathbb{P}\left(\left\| \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} V_\epsilon(G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})) - \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t}) \right\| > \delta\right)$$

$$\le \frac{2}{\delta}\, \|F\|_\infty \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\big[\| V_\epsilon(G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})) - G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})\|\big]$$

$$= \frac{2}{\delta}\, \|F\|_\infty \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\big[\| V_\epsilon(G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})) - G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})\|\, c_{n,m}(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})\big]$$

$$+ \frac{2}{\delta}\, \|F\|_\infty \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}\big[\| V_\epsilon(G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})) - G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})\|$$

$$\times\, (1 - c_{n,m}(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t}))\big]$$

$$=: E_n + F_n,$$

where $c_{n,m}(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t})$ was defined in (4.9). Of the last two terms, we have that as $n \to \infty$,

$$F_n \le \frac{4}{\delta}\, \|F\|_\infty \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}[\|H_n(\mathcal{X}_\mathbf{i})\|(1 - c_{n,m}(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t}))] \to 0, \tag{4.22}$$

where the last convergence is obtained by repeating the same argument as that for proving that (4.10) converges to 0 as $n \to \infty$. For the asymptotics of E_n, recall that $G_n(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t}) = H_n(\mathcal{X}_\mathbf{i})$ whenever $c_{n,m}(\mathcal{X}_\mathbf{i},\, \mathcal{X}_n;\mathbf{t}) = 1$. Therefore

$$E_n \le \frac{2}{\delta}\, \|F\|_\infty \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k,n}} \mathbb{E}[\| V_\epsilon(H_n(\mathcal{X}_\mathbf{i})) - H_n(\mathcal{X}_\mathbf{i})\|]$$

$$= \frac{2}{\delta}\, \|F\|_\infty \big(n^k r_n^{d(k-1)}\big)^{-1} \binom{n}{k}$$

$$\times \int_{(\mathbb{R}^d)^k} \| V_\epsilon(H_n(x_1, \ldots, x_k)) - H_n(x_1, \ldots, x_k)\| \prod_{i=1}^k f(x_i)\, \mathrm{d}\mathbf{x}.$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

Making the change of variables by $x_i = x + r_n y_{i-1}$, $i = 1, \ldots, k$ (with $y_0 \equiv 0$) and using the translation invariance of H,

$$E_n \le \frac{2\|F\|_\infty (\|f\|_\infty)^{k-1}}{\delta k!} \int_{(\mathbb{R}^d)^{k-1}} \|V_\epsilon(H(0, \mathbf{y})) - H(0, \mathbf{y})\| \, d\mathbf{y} \to 0 \quad \text{as } \epsilon \to 0, \tag{4.23}$$

where the last convergence is obtained as a consequence of the dominated convergence theorem and condition **(H4)** of H. Combining (4.22) and (4.23), we conclude that $\lim_{\epsilon \to 0} \limsup_{n \to \infty} C_n = 0$. Next, the same reasoning as in (4.21) yields that, for every $0 < \delta < \delta_0/2$,

$$\begin{aligned}
D_n &= \big(n^k r_n^{d(k-1)}\big)^{-1} \mathbb{E}\Bigg[\Bigg| F\Bigg(\sum_{\mathbf{i} \in \mathcal{I}_{k,n}} V_\epsilon(G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t}))\Bigg) - F\Bigg(\sum_{\mathbf{i} \in \mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\Bigg)\Bigg| \\
&\quad \times \mathbb{1}\Bigg\{\Bigg\| \sum_{\mathbf{i} \in \mathcal{I}_{k,n}} V_\epsilon(G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})) - \sum_{\mathbf{i} \in \mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\Bigg\| \le \delta, \\
&\qquad\qquad \Bigg\| \sum_{\mathbf{i} \in \mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\Bigg\| > \delta' \Bigg\}\Bigg] \\
&\le \omega_F(\delta) \big(n^k r_n^{d(k-1)}\big)^{-1} \mathbb{P}\Bigg(\Bigg\| \sum_{\mathbf{i} \in \mathcal{I}_{k,n}} G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{t})\Bigg\| > \delta'\Bigg) \\
&\le \frac{\omega_F(\delta)}{\delta'} \big(n^k r_n^{d(k-1)}\big)^{-1} \sum_{\mathbf{i} \in \mathcal{I}_{k,n}} \mathbb{E}[\|H_n(\mathcal{X}_\mathbf{i})\|] \\
&\le \frac{\omega_F(\delta)(\|f\|_\infty)^{k-1}}{\delta' k!} \int_{(\mathbb{R}^d)^{k-1}} \|H(0, \mathbf{y})\| \, d\mathbf{y}.
\end{aligned} \tag{4.24}$$

Therefore, for every $0 < \delta < \delta_0/2$,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} |\eta_n \circ T_{V_\epsilon}^{-1}(F) - \widetilde{\eta}_n(F)| \le \frac{\omega_F(\delta)(\|f\|_\infty)^{k-1}}{\delta' k!} \int_{(\mathbb{R}^d)^{k-1}} \|H(0, \mathbf{y})\| \, d\mathbf{y}.$$

The rightmost term above tends to 0 as $\delta \downarrow 0$ because F is uniformly continuous. Now the entire proof has been completed. □

4.2. Proof of Theorem 3.1

Before starting the proof of Theorem 3.1, we will introduce a certain lemma that gives the upper and lower bounds of the kth persistent Betti number in (3.3). Before stating the lemma, we need to recall the notations (4.3) and (4.4).

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

Lemma 4.1. (Lemma 4.1 in [28].) *Under the assumptions of Theorem 3.1, we have, for all* $0 \le s \le t \le \infty$*,*

$$\Bigg|\beta_{k,n}(s,t) - \sum_{\mathbf{i}\in\mathcal{I}_{k+2,n}} h_{r_n s}(\mathcal{X}_\mathbf{i})\, h_{r_n t}(\mathcal{X}_\mathbf{i}) \times \mathbb{1}\{\|y-z\| \ge r_n t \text{ for all } y \in \mathcal{X}_\mathbf{i} \text{ and } z \in \mathcal{X}_n \setminus \mathcal{X}_\mathbf{i}\}\Bigg| \le \binom{k+3}{k+1} L_{r_n t},$$

where

$$L_{r_n t} := \sum_{\mathbf{i}\in\mathcal{I}_{k+3,n}} \mathbb{1}\{\check{C}(\mathcal{X}_\mathbf{i}, r_n t) \text{ is connected}\}.$$

Moreover, for all $0 < t < \infty$*,*

$$\big(n^{k+2} r_n^{d(k+1)}\big)^{-1} \mathbb{E}[L_{r_n t}] \to 0, \quad n \to \infty. \tag{4.25}$$

Proof of Theorem 3.1. We first define a scaled version of H in (3.5) by

$$\begin{aligned} H_n(\{x_1, \dots, x_{k+2}\}; \mathbf{s}, \mathbf{t}) &= H\big(\{r_n^{-1}x_1, \dots, r_n^{-1}x_{k+2}\}; \mathbf{s}, \mathbf{t}\big) \\ &= \big(h_{r_n s_i}(x_1, \dots, x_{k+2})\, h_{r_n t_i}(x_1, \dots, x_{k+2})\big)_{i=1}^m, \quad x_i \in \mathbb{R}^d. \end{aligned}$$

For a subset $\mathcal{Y}$ of $k+2$ points in $\mathbb{R}^d$ and a finite point set $\mathcal{Z} \supset \mathcal{Y}$ in $\mathbb{R}^d$, define $c(\mathcal{Y}, \mathcal{Z}; \mathbf{t})$ and $c_n(\mathcal{Y}, \mathcal{Z}; \mathbf{t})$ as in (2.2) and (2.4) respectively. Analogously to (2.3) and (2.5), we also define

$$\begin{aligned} G(\mathcal{Y}, \mathcal{Z}; \mathbf{s}, \mathbf{t}) &:= H(\mathcal{Y}; \mathbf{s}, \mathbf{t}) \circ c(\mathcal{Y}, \mathcal{Z}; \mathbf{t}), \\ G_n(\mathcal{Y}, \mathcal{Z}; \mathbf{s}, \mathbf{t}) &:= G(r_n^{-1}\mathcal{Y}, r_n^{-1}\mathcal{Z}; \mathbf{s}, \mathbf{t}) = H_n(\mathcal{Y}; \mathbf{s}, \mathbf{t}) \circ c_n(\mathcal{Y}, \mathcal{Z}; \mathbf{t}). \end{aligned}$$

Since Theorem 2.1 yields

$$\begin{aligned} &\big(n^{k+2} r_n^{d(k+1)}\big)^{-1} \mathbb{P}\Bigg(\sum_{\mathbf{i}\in\mathcal{I}_{k+2,n}} G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{s}, \mathbf{t}) \in \cdot\Bigg) \\ &\quad \xrightarrow{v} C_{k+2}\lambda\big\{\mathbf{y} \in (\mathbb{R}^d)^{k+1} : H(\{0, \mathbf{y}\}; \mathbf{s}, \mathbf{t}) \in \cdot\big\} \quad \text{in } M_+([0,\infty]^m \setminus \{\mathbf{0}\}), \end{aligned}$$

(3.6) will follow, provided that for every $F \in C_K^+([0,\infty]^m \setminus \{\mathbf{0}\})$,

$$\big(n^{k+2} r_n^{d(k+1)}\big)^{-1} \mathbb{E}\Bigg[F(\beta_{k,n}(\mathbf{s}, \mathbf{t})) - F\Bigg(\sum_{\mathbf{i}\in\mathcal{I}_{k+2,n}} G_n(\mathcal{X}_\mathbf{i}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\Bigg)\Bigg] \to 0, \quad n \to \infty, \tag{4.26}$$

where

$$\beta_{k,n}(\mathbf{s}, \mathbf{t}) := \big(\beta_{k,n}(s_i, t_i)\big)_{i=1}^m.$$

As in the proof of Theorem 2.1, we fix $\delta_0, \delta' > 0$ so that

$$\mathrm{supp}(F) \cap [0,\infty)^m \subset \{x \in [0,\infty)^m : \|x\| > \delta_0\}$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

and $\delta' \in (0, \delta_0/2)$. Then, for every $0 < \delta < \delta_0/2$, the absolute value of (4.26) is bounded by

$$\begin{aligned}
&\big(n^{k+2} r_n^{d(k+1)}\big)^{-1} \mathbb{E}\Bigg[\Bigg| F(\beta_{k,n}(\mathbf{s}, \mathbf{t})) - F\Bigg(\sum_{\mathbf{i} \in \mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\Bigg)\Bigg| \\
&\qquad \times \mathbb{1}\Bigg\{\Bigg\| \beta_{k,n}(\mathbf{s}, \mathbf{t}) - \sum_{\mathbf{i} \in \mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\Bigg\| > \delta\Bigg\}\Bigg] \\
&\qquad + \big(n^{k+2} r_n^{d(k+1)}\big)^{-1} \mathbb{E}\Bigg[\Bigg| F(\beta_{k,n}(\mathbf{s}, \mathbf{t})) - F\Bigg(\sum_{\mathbf{i} \in \mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\Bigg)\Bigg| \\
&\qquad \times \mathbb{1}\Bigg\{\Bigg\| \beta_{k,n}(\mathbf{s}, \mathbf{t}) - \sum_{\mathbf{i} \in \mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\Bigg\| \le \delta\Bigg\}\Bigg] \\
&=: A_n + B_n.
\end{aligned}$$

By Markov's inequality and $\|F\|_\infty < \infty$,

$$A_n \le \frac{2\|F\|_\infty}{\delta} \big(n^{k+2} r_n^{d(k+1)}\big)^{-1} \mathbb{E}\Bigg[\Bigg\| \beta_{k,n}(\mathbf{s}, \mathbf{t}) - \sum_{\mathbf{i} \in \mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\Bigg\|\Bigg].$$

For ease of description, we assume that $0 \le t_1 \le \cdots \le t_m < \infty$. Then Lemma 4.1 gives that

$$\begin{aligned}
&\Bigg\| \beta_{k,n}(\mathbf{s}, \mathbf{t}) - \sum_{\mathbf{i} \in \mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\Bigg\| \\
&\quad = \Bigg\{\sum_{j=1}^m \Bigg(\beta_{k,n}(s_j, t_j) - \sum_{\mathbf{i} \in \mathcal{I}_{k+2,n}} h_{r_n s_j}(\mathcal{X}_{\mathbf{i}})\, h_{r_n t_j}(\mathcal{X}_{\mathbf{i}}) \\
&\qquad\quad \times \mathbb{1}\{\|y - z\| \ge r_n t_j \text{ for all } y \in \mathcal{X}_{\mathbf{i}} \text{ and } z \in \mathcal{X}_n \setminus \mathcal{X}_{\mathbf{i}}\}\Bigg)^2\Bigg\}^{1/2} \\
&\quad \le \binom{k+3}{k+1} \Bigg\{\sum_{j=1}^m L_{r_n t_j}^2\Bigg\}^{1/2} \le \sqrt{m} \binom{k+3}{k+1} L_{r_n t_m}.
\end{aligned}$$

The last inequality is due to the fact that L_r is non-decreasing in r. By virtue of this bound and (4.25) in Lemma 4.1,

$$A_n \le \frac{2\sqrt{m}\|F\|_\infty}{\delta} \binom{k+3}{k+1} \big(n^{k+2} r_n^{d(k+1)}\big)^{-1} \mathbb{E}[L_{r_n t_m}] \to 0 \quad \text{as } n \to \infty.$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

Repeating the same argument as in (4.21) and (4.24), one can obtain that

$$B_n = \left(n^{k+2} r_n^{d(k+1)}\right)^{-1} \mathbb{E}\left[\left| F(\beta_{k,n}(\mathbf{s},\mathbf{t})) - F\left(\sum_{\mathbf{i}\in\mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\right)\right|\right.$$

$$\left. \times \mathbb{1}\left\{\left\| \beta_{k,n}(\mathbf{s},\mathbf{t}) - \sum_{\mathbf{i}\in\mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\right\| \le \delta,\ \left\| \sum_{\mathbf{i}\in\mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\right\| > \delta' \right\}\right]$$

$$\le \omega_F(\delta)\left(n^{k+2} r_n^{d(k+1)}\right)^{-1} \mathbb{P}\left(\left\| \sum_{\mathbf{i}\in\mathcal{I}_{k+2,n}} G_n(\mathcal{X}_{\mathbf{i}}, \mathcal{X}_n; \mathbf{s}, \mathbf{t})\right\| > \delta'\right)$$

$$\le \frac{\omega_F(\delta)}{\delta'}\left(n^{k+2} r_n^{d(k+1)}\right)^{-1} \sum_{\mathbf{i}\in\mathcal{I}_{k+2,n}} \mathbb{E}[\|H_n(\mathcal{X}_{\mathbf{i}}; \mathbf{s}, \mathbf{t})\|]$$

$$\le \frac{\omega_F(\delta)(\|f\|_\infty)^{k+1}}{\delta'(k+2)!} \int_{(\mathbb{R}^d)^{k+1}} \|H(\{0, y_1, \dots, y_{k+1}\}; \mathbf{s}, \mathbf{t})\| \, \mathrm{d}\mathbf{y}.$$

Therefore, for every $0 < \delta < \delta_0/2$,

$$\limsup_{n\to\infty} (A_n + B_n) \le \frac{\omega_F(\delta)(\|f\|_\infty)^{k+1}}{\delta'(k+2)!} \int_{(\mathbb{R}^d)^{k+1}} \|H(\{0, y_1, \dots, y_{k+1}\}; \mathbf{s}, \mathbf{t})\| \, \mathrm{d}\mathbf{y}.$$

Finally, the right-hand side converges to 0 as $\delta \downarrow 0$. Hence $\limsup_{n\to\infty} (A_n + B_n) = 0$, and we have established (3.6).

Finally, applying Portmanteau's theorem for vague convergence (see [32, Proposition 3.12]) to (3.6), we can see that as $n \to \infty$,

$$\begin{aligned}&\left(n^{k+2} r_n^{d(k+1)}\right)^{-1} \mathbb{P}(\beta_{k,n}(s_i, t_i) \ge u_i) \\ &\quad \to C_{k+2}\lambda\left\{\mathbf{y} \in (\mathbb{R}^d)^{k+1} : h_{s_i}(0, \mathbf{y}) h_{t_i}(0, \mathbf{y}) \ge u_i,\ i = 1, \dots, m\right\}\end{aligned} \tag{4.27}$$

for all $u_i \ge 0$, $u_i \ne 1$, $i = 1, \dots, m$, with $\max_{1\le i\le m} u_i > 0$. By the customary change of variables, it is elementary to show that as $n \to \infty$,

$$\binom{n}{k+2} \mathbb{P}(h_{r_n}(X_1, \dots, X_{k+2}) = 1) \sim n^{k+2} r_n^{d(k+1)} C_{k+2} \int_{(\mathbb{R}^d)^{k+1}} h_1(0, \mathbf{y}) \, \mathrm{d}\mathbf{y}. \tag{4.28}$$

Now (3.7) is obtained as a result of (4.27) and (4.28). □

4.3. Proof of Corollary 3.1

Proof. We need to rewrite $F_{k,n}$ in the notation of Theorem 2.1. First it is easy to show that H in (3.9) fulfills conditions **(H1)–(H4)**. Letting H_n be defined as in (2.1), one can write

$$(r_n^{-kb_i})_{i=1}^m \circ F_{k,n} = \sum_{\mathcal{Y}\subset\mathcal{X}_n, |\mathcal{Y}|=k+1} H_n(\mathcal{Y}).$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

Then (3.8) is an easy consequence of Theorem 2.1. Note also that

$$
\begin{aligned}
&\binom{n}{k+1}\mathbb{P}\bigg(\bigcap_{j=1}^{k+1} B(X_j, r_n/2) \neq \emptyset\bigg) \\
&= \binom{n}{k+1}\int_{(\mathbb{R}^d)^{k+1}} \mathbb{1}\bigg\{\bigcap_{j=1}^{k+1} B(x_j, r_n/2) \neq \emptyset\bigg\}\prod_{i=1}^{k+1} f(x_i)\,\mathrm{d}\mathbf{x} \\
&= \binom{n}{k+1} r_n^{dk}\int_{\mathbb{R}^d}\int_{(\mathbb{R}^d)^k} \mathbb{1}\bigg\{B(0, 1/2)\cap\bigcap_{j=1}^{k} B(y_j, 1/2) \neq \emptyset\bigg\} f(x)\prod_{i=1}^{k} f(x + r_n y_i)\,\mathrm{d}\mathbf{y}\,\mathrm{d}x \\
&\sim n^{k+1} r_n^{dk} C_{k+1}\lambda\bigg\{\mathbf{y} \in (\mathbb{R}^d)^k : B(0, 1/2)\cap\bigcap_{j=1}^{k} B(y_j, 1/2) \neq \emptyset\bigg\}, \quad n \to \infty. \qquad (4.29)
\end{aligned}
$$

Finally, (3.10) can be obtained from (3.8) and (4.29) as well as Portmanteau's theorem for vague convergence. □

4.4. Proof of Corollary 3.2

Proof. As in the proof of Corollary 3.1, one needs to reformulate $S_{k,n}$ in the notation of Theorem 2.1. First we notice that H in (3.13) satisfies conditions **(H1)–(H4)**. Define H_n by (2.1), and for a subset $\mathcal{Y} \subset \mathbb{R}^d$ with $|\mathcal{Y}| = k + 1$ and a finite $\mathcal{Z} \supset \mathcal{Y}$ in $\mathbb{R}^d$,

$$c(\mathcal{Y}, \mathcal{Z}) := \big(\mathbb{1}\{\mathcal{U}(\mathcal{Y}) \cap \mathcal{Z} = \emptyset\}\big)_{i=1}^m. \qquad (4.30)$$

Note that we have defined (4.30) in a way different from the original definition in (2.2). Then, unlike the definition in (2.4),

$$c_n(\mathcal{Y}, \mathcal{Z}) := c(r_n^{-1}\mathcal{Y}, r_n^{-1}\mathcal{Z}) = c(\mathcal{Y}, \mathcal{Z})$$

does not depend on $n \geq 1$. Defining $G_n(\mathcal{Y}, \mathcal{Z}) := H_n(\mathcal{Y}) \circ c_n(\mathcal{Y}, \mathcal{Z}) = H_n(\mathcal{Y}) \circ c(\mathcal{Y}, \mathcal{Z})$ as in (2.5), we have

$$\big(r_n^{-b_i}\big)_{i=1}^m \circ S_{k,n} = \sum_{\mathcal{Y} \subset \mathcal{X}_n, |\mathcal{Y}|=k+1} G_n(\mathcal{Y}, \mathcal{X}_n). \qquad (4.31)$$

For our purposes we need to apply Theorem 2.1 to (4.31). Before doing so, however, one must slightly modify the proof of Theorem 2.1 as we have changed the definition of c as in (4.30). Below, we show that for every $x \in \mathbb{R}^d$ and $\mathbf{y} = (y_1, \ldots, y_k) \in (\mathbb{R}^d)^k$,

$$\lim_{n\to\infty}\bigg(1 - \int_{\mathcal{U}(x, x+r_n y_1, \ldots, x+r_n y_k)} f(z)\,\mathrm{d}z\bigg)^{n-k-1} = 1. \qquad (4.32)$$

In fact this replaces the argument in (4.13). If (4.32) is established, the remainder of the argument in the proof of Theorem 2.1 can be altered in a very obvious manner. To show (4.32), we note that

$$
\begin{aligned}
&\lim_{n\to\infty}\bigg(1 - \int_{\mathcal{U}(x, x+r_n y_1, \ldots, x+r_n y_k)} f(z)\,\mathrm{d}z\bigg)^{n-k-1} \\
&= \exp\bigg\{-\lim_{n\to\infty}(n-k-1)\int_{\mathcal{U}(x, x+r_n y_1, \ldots, x+r_n y_k)} f(z)\,\mathrm{d}z\bigg\},
\end{aligned}
$$

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

so that

$$n\int_{\mathcal{U}(x,x+r_ny_1,\dots,x+r_ny_k)} f(z)\,\mathrm{d}z = n(r_nR(0,\mathbf{y}))^d\theta_d\frac{\int_{\mathcal{U}(x,x+r_ny_1,\dots,x+r_ny_k)} f(z)\,\mathrm{d}z}{\lambda(\mathcal{U}(x,x+r_ny_1,\dots,x+r_ny_k))},$$

where θ_d is a volume of the unit ball in $\mathbb{R}^d$. By the Lebesgue differentiation theorem,

$$\frac{\int_{\mathcal{U}(x,x+r_ny_1,\dots,x+r_ny_k)} f(z)\,\mathrm{d}z}{\lambda(\mathcal{U}(x,x+r_ny_1,\dots,x+r_ny_k))} \to f(x), \quad n\to\infty.$$

Since $nr_n^d\to 0$ as $n\to\infty$, we have obtained (4.32).

Now we can apply Theorem 2.1 to get (3.12). Furthermore, by Portmanteau's theorem for vague convergence, we obtain that

$$\begin{aligned}&\big(n^{k+1}r_n^{dk}\big)^{-1}\mathbb{P}\left(\big(r_n^{-b_i}\big)_{i=1}^m\circ S_{k,n}\in\prod_{i=1}^m[u_i,\infty)\right)\\ &\quad\to C_{k+1}\lambda\left\{\mathbf{y}\in(\mathbb{R}^d)^k : H(0,\mathbf{y})\in\prod_{i=1}^m[u_i,\infty)\right\}\end{aligned} \tag{4.33}$$

for all $0<u_i\le T_i^{b_i}$, $i=1,\dots,m$. By the same calculation as in (4.29), we can show that as $n\to\infty$,

$$\binom{n}{k+1}\mathbb{P}(R(X_1,\dots,X_{k+1})\le r_n)\sim n^{k+1}r_n^{dk}C_{k+1}\lambda\big\{\mathbf{y}\in(\mathbb{R}^d)^k : R(0,\mathbf{y})\le 1\big\}. \tag{4.34}$$

Now (3.14) is a direct consequence of (4.33) and (4.34). □

Acknowledgement

The author is very grateful for useful comments received from an anonymous referee and an anonymous Associate Editor. These comments helped the author to introduce a number of improvements to the paper.

Funding information

This research was partially supported by the NSF grant DMS-1811428 and the AFOSR grant FA9550-22-0238.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Björner, A. (1995). Topological methods. In *Handbook of Combinatorics*, eds R. Graham *et al.*, pp. 1819–1872. Elsevier, Amsterdam.

[2] Blaszczyszyn, B., Yogeshwaran, D. and Yukich, J. E. (2019). Limit theory for geometric statistics of point processes having fast decay of correlations. *Ann. Prob.* **47**, 835–895.

[3] Bobrowski, O. and Adler, R. J. (2014). Distance functions, critical points, and the topology of random Čech complexes. *Homology Homotopy Appl.* **16**, 311–344.

[4] BOBROWSKI, O. AND MUKHERJEE, S. (2015). The topology of probability distributions on manifolds. *Prob. Theory Related Fields* **161**, 651–686.
[5] BOBROWSKI, O. AND WEINBERGER, S. (2017). On the vanishing of homology in random Čech complexes. *Random Structures Algorithms* **51**, 14–51.
[6] DECREUSEFOND, L., SCHULTE, M. AND THÄLE, C. (2016). Functional Poisson approximation in Kantorovich–Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Prob.* **44**, 2147–2197.
[7] EDELSBRUNNER, H., LETSCHER, D. AND ZOMORODIAN, A. (2002). Topological persistence and simplification. *Discrete Comput. Geom.* **28**, 511–533.
[8] FASEN, V. AND ROY, P. (2016). Stable random fields, point processes and large deviations. *Stoch. Process. Appl.* **126**, 832–856.
[9] FOWLER, C. F. (2019). Homology of multi-parameter random simplicial complexes. *Discrete Comput. Geom.* **62**, 87–127.
[10] GERSHKOVICH, V. AND RUBINSTEIN, H. (1997). Morse theory for min-type functions. *Asian J. Math.* **1**, 696–715.
[11] HATCHER, A. (2002). *Algebraic Topology*. Cambridge University Press, Cambridge.
[12] HIRAOKA, Y., SHIRAI, T. AND TRINH, K. D. (2018). Limit theorems for persistence diagrams. *Ann. Appl. Prob.* **28**, 2740–2780.
[13] HULT, H. AND LINDSKOG, F. (2006). Regular variation for measures on metric spaces. *Publ. de l'Institut Math.* **80**, 121–140.
[14] HULT, H. AND SAMORODNITSKY, G. (2010). Large deviations for point processes based on stationary sequences with heavy tails. *J. Appl. Prob.* **47**, 1–40.
[15] KAHLE, M. (2011). Random geometric complexes. *Discrete Comput. Geom.* **45**, 553–573.
[16] KAHLE, M. (2014). Sharp vanishing thresholds for cohomology of random flag complexes. *Ann. Math.* **179**, 1085–1107.
[17] KAHLE, M. AND MECKES, E. (2013). Limit theorems for Betti numbers of random simplicial complexes. *Homology Homotopy Appl.* **15**, 343–374.
[18] KAHLE, M. AND PITTEL, B. (2016). Inside the critical window for cohomology of random k-complexes. *Random Structures Algorithms* **48**, 102–124.
[19] KALLENBERG, O. (2017). *Random Measures, Theory and Applications*. Springer.
[20] KREBS, J. T. N. AND POLONIK, W. (2020). On the asymptotic normality of persistent Betti numbers. Available at arXiv:1903.03280.
[21] KREBS, J., ROYCRAFT, B. AND POLONIK, W. (2021). On approximation theorems for the Euler characteristic with applications to the bootstrap. *Electron. J. Statist.* **15**, 4462–4509.
[22] LACHIÈZE-REY, R. AND REITZNER, M. (2016). U-statistics in stochastic geometry. In *Stochastic Analysis for Poisson Point Processes: Malliavin Calculus, Wiener–Itô Chaos Expansions and Stochastic Geometry*, eds G. Peccati and M. Reitzner, pp. 229–254. Springer/Bocconi University Press, Milan.
[23] LAST, G. AND PENROSE, M. (2018). *Lectures on the Poisson Process*. Cambridge University Press.
[24] LAST, G., PECCATI, G. AND SCHULTE, M. (2016). Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Prob. Theory Related Fields* **165**, 667–723.
[25] LINDSKOG, F., RESNICK, S. I. AND ROY, J. (2014). Regularly varying measures on metric spaces: hidden regular variation and hidden jumps. *Prob. Surveys* **11**, 270–314.
[26] MUNKRES, J. R. (1996). *Elements of Algebraic Topology*, 1st edn. Westview Press.
[27] OWADA, T. (2019). Topological crackle of heavy-tailed moving average processes. *Stoch. Process. Appl.* **129**, 4965–4997.
[28] OWADA, T. (2022). Convergence of persistence diagram in the sparse regime. *Ann. Appl. Prob.*, to appear. Available at arXiv:2103.12943.
[29] OWADA, T. AND THOMAS, A. (2020). Limit theorems for process-level Betti numbers for sparse and critical regimes. *Adv. Appl. Prob.* **52**, 1–31.
[30] PENROSE, M. (2003). *Random Geometric Graphs* (Oxford Studies in Probability **5**). Oxford University Press, Oxford.
[31] REITZNER, M. AND SCHULTE, M. (2013). Central limit theorems for U-statistics of Poisson point processes. *Ann. Prob.* **41**, 3879–3909.
[32] RESNICK, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.
[33] SEGERS, J., ZHAO, Y. AND MEINGUET, T. (2017). Polar decomposition of regularly varying time series in star-shaped metric spaces. *Extremes* **20**, 539–566.
[34] SKRABA, P., THOPPE, G. AND YOGESHWARAN, D. (2020). Randomly weighted d-complexes: minimal spanning acycles and persistence diagrams. *Electron. J. Combinatorics* **27**.

https://doi.org/10.1017/jpr.2022.39 Published online by Cambridge University Press

[35] THOMAS, A. M. AND OWADA, T. (2021). Functional limit theorems for the Euler characteristic process in the critical regime. *Adv. Appl. Prob.* **53**, 57–80.
[36] YOGESHWARAN, D. AND ADLER, R. J. (2015). On the topology of random complexes built over stationary point processes. *Ann. Appl. Prob.* **25**, 3338–3380.
[37] YOGESHWARAN, D., SUBAG, E. AND ADLER, R. J. (2017). Random geometric complexes in the thermodynamic regime. *Prob. Theory Related Fields* **167**, 107–142.