**Universität Osnabrück**

Dissertation zur Erlangung des Doktorgrades (Dr. rer. nat.)
des Fachbereichs Mathematik/Informatik
der Universität Osnabrück

**August 2020**

# Limit Theorems for Random Simplicial Complexes

vorgelegt von

**Grace Itunuoluwa Akinwande**

aus

Ibadan, Nigeria

*Acknowledgements*

First, with a thankful heart, I wish to acknowledge the guidance and support of my supervisor, Matthias Reitzner, during my PhD studies. I'm deeply grateful for the opportunity to work with him and for the numerous eye-opening discussions that have proved to be helpful in writing this thesis. I'm highly grateful to the *Deutsche Forschungsgemeinschaft* via *Research Training Group 1916 - Combinatorial Structures in Geometry* for the financial aid which made this research possible.
I also wish to appreciate Hanna Döring, Matthias Schulte and Christoph Thäle for helpful discussions during the course of this thesis.
To my current and former colleagues, Carina, Binh, Jens, to mention a few, I would like to say "thank you" for being there during mathematical and non-mathematical discussions, which always served as a morale booster.
I'm deeply grateful to my family and friends for always cheering me on to the finish line. Special appreciation goes to my children who gave me a reason to keep on, and to my husband, siblings and parents for their love and support.
My deepest appreciation goes to the One who holds the universe for His divine grace in holding me together during my PhD studies.

# Contents

# 1. Introduction

The study of random simplicial complexes and even more generally, simplicial complexes, continues to be an interesting and active area of research because of its important applications. The study of random 1-simplicial complexes (random graphs) was first introduced by Erdös, Rényi and Gilbert [ER59] [Gil61], and since then, it has been an intensively studied object of discrete mathematics and stochastic. The Erdös-Rényi graph consists of a fixed vertex set with a fixed number of equally likely edges. It is a purely combinatorial object. But with the Gilbert graph being geometric in nature, it has found more applications in the real world, for example, in the study of communication networks. This is because the relative position of points in space plays an important role in its construction. This type of random graph model is based on point processes and has been generalised in various ways, for example in [Pen03], where an extensive study on it can be found.

The idea in this thesis is to investigate a random structure, precisely a random simplicial complex, which is constructed on a stationary Poisson point process. A Poisson point process, $\eta$, is such that for any Borel set $A$, in the space in consideration (a subset of $\mathbb{R}^d$), the number of points in $A$, $\eta(A)$, is a random variable with a Poisson distribution. That is, $\mathbb{P}(\eta(A) = 0) = \exp(-\mu(A))$, where $\mu$ is known as the intensity measure of the process. The Poisson point process $\eta_t$ is said to be stationary, if the intensity measure is given by $\mu = t\lambda$, where $t$ is the intensity and $\lambda$ is the usual Lebesque measure in $\mathbb{R}^d$. The Poisson point process plays quite a significant role in probability theory and its applications, and in several problems in stochastic geometry.

A simple example of such random simplicial complex is the Gilbert graph [Gil61] on the Poisson point process, in which the vertices are points in the process and two points are connected by an edge if and only if they are close together enough according to a prescribed distance parameter.

We will consider random variables depending on the Poisson point process described above, generally known as Poisson functionals. One such functional on the Gilbert graph is the length-power functional described in [RST17], which is related to the edge lengths of the graph.

In this thesis, we took it a step further, by considering more general simplicial complexes in which a $k$-simplex exists when its $k+1$ vertices are 'well connected'. Examples of such simplices are the Vietoris-Rips complex $\mathcal{R}(\eta_t, \delta_t)$ and *Čech complex* $\mathcal{C}(\eta_t, \delta_t)$, where $\delta_t$ is the prescribed distance parameter. We then considered the volume-power functional which is the sum of the $\alpha$-power of the volume of these simplices. We also examined the asymptotic expectation and asymptotic covariance of these types of Poisson functionals and established both univariate and multivariate central limit theorems. We will see that the univariate limit theorem exists only if the expectation tends to infinity.

We did observe these asymptotics in the following different regimes as introduced by Penrose [Pen03].

- The *sparse* or *subcritical* regime, where $\lim_{t\to\infty} t\delta_t^d = 0$.
- The *thermodynamic* or *critical* regime, where $\lim_{t\to\infty} t\delta_t^d = c \in (0, \infty)$.
- The *dense* or *supercritical* regime, where $\lim_{t\to\infty} t\delta_t^d = \infty$.

In the special case where $\alpha = 0$, we have the Poisson functional only counting the number of simplices. This gives the components of the $\boldsymbol{f}$-vector. We went on to get a Poisson limit theorem while restricting the expectation to tend to a constant. In this case, we discovered there is no Poisson multivariate limit theorem as the expectation of $k$-simplices tending to a constant implies that for $i > k$ and $j < k$, the expectations of the $i$- and $j$-simplices tend to 0 and $\infty$ respectively.

This thesis is based on the following paper and therefore coincides with it to a very large extent.

[AR20] G. Akinwande and M. Reitzner, Multivariate central limit theorems for random simplicial complexes. *Advances in Applied Mathematics*, 121(102076), 2020.

## 1.1. Outline

We organise the work as follows:

In Chapter 2, we provide the relevant materials useful in this thesis. First, we give some basic notations in Section 2.1, and then recall facts from stochastic geometry in Section 2.2. Section 2.3 provides an overview on random simplicial complexes, and the cardinality of some special class of partitions is discussed in Section 2.4.
Chapters 3 to 6 explore the first and second moments, central limit theorem and Poisson limit theorem respectively, of the so-called volume-power functionals in consideration.
In Chapter 3, we discuss the expectation for functionals based on random graphs and

then more generally for functionals based on random simplicial complexes in different regimes. We are able to see that the asymptotic expectation does not necessarily have to behave according to its current regime.

In Chapter 4, we investigate the covariance structure of the volume-power functional. We are able to analyse the rank of the covariance matrix for the normalised version of the volume-power functional in the different regimes. We see that the covariance matrix is of full rank in the sparse regime, and in the thermodynamic regime except for finitely many values, and it is singular in the dense regime.

We discuss the central limit theorem in Chapter 5, first the univariate case and then the multivariate case. Special fourth moment integrals involving partitions are used. We are able to see that the central limit theorem holds in any case once the expectation goes to infinity. Finally, we see in Chapter 6, the Poisson limit theorem for the components of the $\boldsymbol{f}$-vector. This is most applicable in the sparse regime.

# 2. Preliminaries

In this chapter, notions needed throughout this thesis are introduced. First, we give some basic notations in the first section, and then recall facts from stochastic geometry in the second section. The third section will provide an overview on random simplicial complexes, and the cardinality of some partition will be discussed in Section 2.4.

## 2.1. Basic notations

We take $\mathbb{N}$ to be the set of natural numbers, $0 \notin \mathbb{N}$, and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. $\mathbb{Z}$, $\mathbb{Z}_+$ and $\mathbb{R}$ are the set of integers, strictly positive integers and real numbers respectively. For $d \geq 1$, $\mathbb{R}^d$ is the $d$-dimensional Euclidean space equipped with scalar product $\langle \cdot, \cdot \rangle$, Euclidean norm $\| \cdot \|$ and Lebesgue measure $\lambda = \lambda_d$. For $r > 0$,

$$B^d(x, r) = \{y \in \mathbb{R}^d : \|x - y\| \leq r\}$$

is a $d$-dimensional (closed) ball with center $x \in \mathbb{R}^d$ and radius $r$. We denote by

$$\kappa_j = \frac{\pi^{\frac{j}{2}}}{\Gamma\left(\frac{j}{2} + 1\right)}$$

the volume of the $j$-dimensional unit ball $B^j = B^j(0, 1)$. In the following we fix a convex compact set $W \subset \mathbb{R}^d$ of unit volume.

Let $\mu$ be a $\sigma$-finite measure on a measurable space $(W, \mathcal{W})$. For $n \in \mathbb{N}$, we denote by $\mu^n$ the product measure of $\mu$ on $W^n$ defined as usual. We write the integral of a measurable function $f : W^n \to \mathbb{R}$ with respect to $\mu^n$ as

$$\int_{W^n} f \mathrm{d}\mu^n = \int_{W^n} f(x_1, \ldots, x_n) \, \mathrm{d}\mu(x_1, \ldots, x_n).$$

If $\mu = \lambda$, we use the standard notation $\int_{W^n} f(x_1, \ldots, x_n) \, \mathrm{d}x_1, \ldots, \mathrm{d}x_n$.

We denote by $L^p(\mu^n)$, the set of all measurable functions $f : W^n \to \mathbb{R}$ such that

$$\int_{W^n} |f|^p \mathrm{d}\mu^n < \infty.$$

We also let $L^p_s(\mu^n)$ be the subspace of all symmetric functions in $L^p(\mu^n)$, that is, those functions that are invariant under permutations of their arguments.

For random variables $X$ and $Y$, $\mathbb{E}X$ and $\mathbb{V}X$ are the expectation and variance of $X$ respectively, while $\mathbb{C}\mathrm{ov}(X, Y)$ is the covariance of $X$ and $Y$.

Since we will be insterested in the asymptotic behaviour of functions, we will use the Bachmann-Landau big-$O$ and little-$O$ notations. Let $g$ and $h$ be two non-negative functions, then we have the following.

(i) $g(x) = O(h(x)) \iff \exists\, x_0, n > 0$ such that for $x > x_0$, $g(x) \le n \cdot h(x)$.
This means $g$ is asymptotically bounded above by $h$, up to a constant factor.

(ii) $g(x) = \Omega(h(x)) \iff \exists\, x_0, n > 0$ such that for $x > x_0$, $g(x) \ge n \cdot h(x)$.
This means $g$ is asymptotically bounded below by $h$, up to a constant factor.

(iii) $g(x) = \Theta(h(n)) \iff g(x) = O(h(x))$ and $g(x) = \Omega(h(x))$.
This means $g$ is asymptotically bounded above and below by $h$, up to constant factors.

(iv) $g(x) = o(h(n)) \iff \forall\, \epsilon > 0,\ \exists\, x_0$ such that for $x > x_0$, $g(x) \le \epsilon \cdot h(x)$.
This means $g$ is asymptotically dominated by $h$.

(v) $g(x) = \omega(h(n)) \iff \forall\, n > 0,\ \exists\, x_0$ such that for $x > x_0$, $g(x) \le n \cdot h(x)$.
This means $g$ asymptotically dominates $h$.

A partition $\sigma$ of a set $X$ is a set of non-empty subsets of $X$ such that $X$ is a disjoint union of those subsets, that is, each element in $X$ is in exactly one of those subsets. The elements of $\sigma$ are called the blocks of $\sigma$.

For example, for $X = \{1, 2, 3, 4\}$, $\sigma_1 = \{\{1, 2\}, \{3\}, \{4\}\}$ and $\sigma_2 = \{\{1, 2\}, \{3, 4\}\}$ are both partitions of $X$ with 3 blocks and 2 blocks respectively.

Let $\mathcal{P}(A)$ stand for the set of partitions of an arbitrary set $A$. Then $|\sigma|$ represents the number of blocks in a partition, $\sigma \in \mathcal{P}(A)$. A partial order is defined on $\mathcal{P}(A)$ such that $\sigma \le \tau$ if each block of $\sigma$ is contained in a block of $\tau$, for $\sigma, \tau \in \mathcal{P}(A)$. The minimal partition $\hat{0}$ is the partition whose blocks are singletons, and the maximal partition $\hat{1}$ is the partition with a single block. For two partitions $\sigma, \tau \in \mathcal{P}(A)$, $\sigma \wedge \tau$ is the maximal partition in $\mathcal{P}(A)$ such that $\sigma \wedge \tau \le \sigma$ and $\sigma \wedge \tau \le \tau$, and $\sigma \vee \tau$ is the minimal partition in $\mathcal{P}(A)$ such that $\sigma \le \sigma \vee \tau$ and $\tau \le \sigma \vee \tau$.

# 2.2. Background on stochastic geometry

## 2.2.1. Poisson point processes

Let $N(\mathbb{X})$ be the set of all $\sigma$-finite integer-valued measures on a measurable space $(\mathbb{X}, \mathcal{X})$. Let $\mathcal{N}(\mathbb{X})$ be the $\sigma$-algebra generated by $N(\mathbb{X})$ such that the map, $g_A : N(\mathbb{X}) \to \bar{\mathbb{R}}$, $\eta \mapsto \eta(A)$ is measurable for $A \in \mathcal{X}$. Given the underlying probability space $(\Omega, \mathcal{F}, P)$, a measurable map $\eta : (\Omega, \mathcal{F}) \to (N(\mathbb{X}), \mathcal{N}(\mathbb{X}))$ is called an integer-valued random measure.

An integer-valued $\sigma$-finite random measure, $\eta$, is called a Poisson point process with intensity measure $\mu$, if

- $\eta(A)$ is Poisson distributed with parameter $\mu(A)$ for each $A \in \mathcal{X}$.
- $\eta(A_1), \dots, \eta(A_n)$ are independent for disjoint sets $A_1, \dots, A_n \in \mathcal{X}$ and $n \in \mathbb{N}$.

A stationary Poisson point process is one whose intensity measure is a constant multiplied by the Lebesgue measure, $\lambda$. In this thesis, we consider a stationary Poisson point process since the intensity measure is given by $\mu = t\lambda$, where $t > 0$ is the intensity of the Poisson point process, $\eta_t$. A random variable depending on a Poisson point process is called a *Poisson functional*.

We represent by $L^p(\mathbb{P}_\eta)$ the set of all measurable funcions $F : N(\mathbb{X}) \to \bar{\mathbb{R}}$ with $\mathbb{E}|F|^p < \infty$, where $\mathbb{P}_\eta = P \circ \eta$ is the image measure.

For a Poisson point process, $\eta$ with intensity measure $\mu$, the expected value of a Poisson functional, $F = F(\eta)$, is denoted by $\mathbb{E}F(\eta)$ and it's behaviour depends on $\mu$. Also, we can consider $\eta$ as a set since we can give distinct multiple points a number as a mark. So for $k \in \mathbb{N}$, $\eta^k_{t,\neq}$ is the set of all $k$-tuples of distinct points in the Poisson point process.

An important tool that will be needed is the multivariate Mecke formula for Poisson point processes given as follows.

$$\mathbb{E} \sum_{(x_1,\dots,x_k)\in\eta^k_{t,\neq}} f(\eta, x_1, \dots, x_k) = \int_{W^k} \mathbb{E} f(\eta + \sum_{i=1}^{k} \delta_{x_i}, x_1, \dots, x_k)\, d\mu(x_1, \dots, x_k)\,,$$

where $k \geq 1$ is a fixed integer, and $f : W^k \times \mathbf{N}(W) \to \mathbb{R}$ is a non-negative measurable function, cf. [SW08, Corollary 3.2.3].

Since the Poisson functionals we will be considering in this thesis do not depend on the Poisson process and we have a stationary Poisson point process, we only need the following special case. We note that the expectation inside the integral can be removed in this case.

$$\mathbb{E} \sum_{(x_1,\dots,x_k)\in\eta^k_{t,\neq}} f(x_1,\dots,x_k) = t^k \int\limits_{W^k} f(x_1,\dots,x_k)\, dx_1 \dots dx_k\,, \tag{2.1}$$

where $f \in L^1(\mu^k)$.

### 2.2.2. Malliavin operators

We will need the multiple Wiener-Itô integrals which play a role in the sequel. The definition was first given for integrable functions and can be extended to square integrable functions via limit on simple functions. We refer to [LP11, Chapter 3] for more on this.

**Definition 2.1.** Let $n \in \mathbb{N}$ and $f \in L^1_s(\mu^n)$. The $n$-th multiple Wiener-Itô integral, $I_n(f)$, of $f$ is given by

$$I_n(f) = \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} \sum_{(x_1,\dots,x_i)\in\eta^i_{\neq}} \int_{X^{n-i}} f(x_1,\dots,x_i,y_1,\dots,y_{n-i}) \mathrm{d}\mu(y_1,\dots,y_{n-i}).$$

The following isometry relation holds for multiple Wiener-Itô integrals with square integrable integrands.

**Lemma 2.2.** *Let $f \in L^2_s(\mu^n)$ and $g \in L^2_s(\mu^m)$, $m,n \geq 1$, then*

*(i)* $\mathbb{E} I_n(f) = 0$

*(ii)* $\mathbb{E} I_n(f) I_m(g) = \mathbb{1}_{\{n=m\}} n! \langle f, g\rangle$.

For a Poisson functional $F$ and $z \in W$, the *difference operator* or *add-one-cost operator*, $D_z F$, is given by

$$D_z F = F(\eta + \delta_z) - F(\eta),$$

where $\delta_z$ represents the Dirac measure concentrated at the point $z \in W$. Higher order differential operators are defined recursively.

$$D^n_{z_1,\dots,z_n} F = D_{z_1} D^{n-1}_{z_2,\dots,z_n} F.$$

It is well known that every square integrable Poisson functional can be written as a sum of multiple Wiener-Itô integrals, that is, they admit a chaos expansion. We have the following from [LP11].

**Theorem 2.3.** *Let $F \in L^2(\mathbb{P}_\eta)$ be a Poisson functional. Then $f_i \in L^2_s(\mu^n)$ for $i \in \mathbb{N}$ and*

$$F = \mathbb{E}F + \sum_{i=1}^{\infty} I_i(f_i). \tag{2.2}$$

Here, the $f_i$'s are called the kernels of the chaos expansion and are defined by

$$f_i(x_1, \dots, x_i) = \int_{X^{n-i}} f(x_1, \dots, x_i, y_1, \dots, y_{n-i}) \mathrm{d}\mu(y_1, \dots, y_{n-i}), \tag{2.3}$$

and also given by

$$f_i(x_1, \dots, x_i) = \frac{1}{i!} \mathbb{E} D^n_{x_1, \dots, x_n} F.$$

Furthermore,

$$\mathbb{V}F = \sum_{i=1}^{\infty} i! \|f_i\|^2. \tag{2.4}$$

Since our proof for Poisson limit theorem is based on results on the Poisson approximation of Poisson functionals from [LP11] which make use of operators from Malliavian calculus, we give a short overview of two of the operators.

**Definition 2.4.** Let dom $D$ be the set of all Poisson functionals, $F \in L^2(\mathbb{P}_\eta)$ such that

$$\sum_{i=1}^{\infty} i i! \|f_i\|^2 < \infty.$$

Then for $F \in$ dom $D$,

$$D_z F = \sum_{i=1}^{\infty} i I_{i-1}(f_i(z, .)).$$

We note the above definition coincides with the previously stated defintion for $F \in$ dom $D$.

**Definition 2.5.** Let dom $L$ be the set of all Poisson functionals, $F \in L^2(\mathbb{P})$ such that

$$\sum_{i=1}^{\infty} i^2 i! \|f_i\|^2 < \infty.$$

Then for $F \in$ dom $L$, the *Ornstein-Uhlenbeck generator*, $LF$, is given by

$$LF = -\sum_{i=1}^{\infty} i I_i(f_i).$$

It has a (pseudo) inverse operator given by

$$L^{-1}F = -\sum_{i=1}^{\infty} \frac{1}{i} I_i(f_i),$$

which is defined for all $F \in L^2(\mathbb{P}\eta)$, although the identity $LL^{-1}$ only holds if $\mathbb{E}F = 0$.

In probability theory, it has always been of interest to decide if a family of random variables converge in distribution to a random variable with a known target distribution. Although the Stein's method was first developed in [Ste72] for the approximation of sums of dependent random variables by the Gaussian distribution, it has been applied to other target distributions like the Poisson, Gamma and Binomial distributions. In this thesis, the target distribution will usually be the Gaussian or the Poisson distributions.

A class $\mathcal{H}$ of real-valued functions $h : \mathbb{R}^d \to \mathbb{R}$ is called *separating* if the following holds: If $F$ and $G$ are two random elements such that $\mathbb{E}|h(F)| < \infty$ and $\mathbb{E}|h(G)| < \infty$ for all $h \in \mathcal{H}$, then $\mathbb{E}h(F) = \mathbb{E}h(F)$ for all $h \in \mathcal{H}$ implies that $F$ and $G$ has the same distribution. Such $h \in \mathcal{H}$ are known as *test functions*.

**Definition 2.6.** Let $\mathcal{H}$ be a class of separating real-valued functions on $\mathbb{R}^d$, then the distance between the laws of two random elements $F$ and $G$, such that $\mathbb{E}|h(F)| < \infty$ and $\mathbb{E}|h(G)| < \infty$ for all $h \in \mathcal{H}$, is given by

$$d_{\mathcal{H}}(F, G) = \sup_{h \in \mathcal{H}} \{|\mathbb{E}[h(F)] - \mathbb{E}[h(G)]|\}. \tag{2.5}$$

We note that $d_{\mathcal{H}}(\cdot, \cdot)$ satisfies the usual axioms of a metric on probability distributions. The choice of the class of test functions present various distances in probability. We present the ones needed in this thesis.

Given two random variables $X$ and $Y$, if we take $\mathcal{H}$ to be the set of indicator functions of intervals $(-\infty, t]$, $t \in \mathbb{R}$, we get the *Kolmogorov distance* given by

$$d_K(X, Y) = \sup_{t \in \mathbb{R}} \{|\mathbb{P}(X \leq t) - \mathbb{P}(Y \leq t)|\}.$$

Alternatively, for two random variables $X$ and $Y$, let Lip(1) be the set of all functions $h : \mathbb{R} \to \mathbb{R}$ with a Lipschitz constant less than or equal to one. Taking $\mathcal{H} = \mathrm{Lip}(1)$, we obtain the *Wasserstein distance*

$$d_W(X, Y) = \sup_{h \in \mathrm{Lip}(1)} \{|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|\}.$$

The Wasserstein distance given above is defined for random variables which will be useful for the univariate central limit theorem, it is also defined for the $m$-dimensional random vectors by considering the functions $h : \mathbb{R}^m \to \mathbb{R}$ with a Lipschitz constant less than or equal to one. But this distance will be too strong for the multivariate case, so we consider another class of test functions instead.

Let $X$ and $Y$ be two $m$-dimensional random vectors such that $\mathbb{E}[\|X\|]^2, \mathbb{E}[\|Y\|]^2 < \infty$, and let $\mathcal{H}_m$ be the set of all thrice continuously differentiable functions $g : \mathbb{R}^m \to \mathbb{R}$

such that the second and third partial derivatives are bounded by one, that is,

$$\max_{1\le i_1\le i_2\le m}\sup_{x\in\mathbb{R}^m}\left|\frac{\partial^2 g}{\partial x_{i_1}\partial x_{i_2}}(x)\right|\le 1 \text{ and } \max_{1\le i_1\le i_2\le i_3\le m}\sup_{x\in\mathbb{R}^m}\left|\frac{\partial^3 g}{\partial x_{i_1}\partial x_{i_2}\partial x_{i_3}}(x)\right|\le 1,$$

then the $d_3$-distance is given by

$$d_3(X,Y)=\sup_{h\in\mathcal{H}_m}\{|\mathbb{E}[h(X)]-\mathbb{E}[h(Y)]|\}.$$

Let $X, Y$ be two $\mathbb{N}$-valued random variables. The *total variation distance* between the laws of $X$ and $Y$, written as $\mathrm{d}_{TV}$, is given by

$$\mathrm{d}_{TV} := \sup_{B\subseteq\mathbb{N}}|\mathbb{P}(X\in B)-\mathbb{P}(Y\in B)|.$$

We note that that the topologies induced by these distances as with other distances in the literature (see [Pe16, Chapter 6]) are strictly stronger than the one induced by convergence in distribution, so that convergence in these distances implies convergence in distribution.

# 2.3. Random structures

## 2.3.1. The Gilbert graph

Let $\eta_t$ represent a Poisson process with intensity $t$. We choose some parameter $\delta_t > 0$ which may depend on $t$. The Gilbert graph, $G(\eta_t, \delta_t)$, is defined as the random graph with vertex set $\eta_t$, and an edge, $e = (x, y)$, exist between two points $\{x, y\} \subset \eta_t$ if $\|x - y\| \le \delta_t$. Gilbert [Gil61] has studied the planar case of these graphs as applicable to communication networks and disease control. One can examine the behaviour of these graphs by counting the number of edges or subgraphs or components [Pen03] or the length of the edges as we see in [RST17]. Here the *length-power functional* was considered. It is given as follows. Let $E(G(\eta_t, \delta_t))$ be the edge set of the Gilbert graph.

$$\begin{aligned}\mathcal{V}^{(\alpha)} &:= \frac{1}{2}\sum_{e=(x,y)\in E(G(\eta_t,\delta_t))}\|x-y\|^\alpha\\ &= \frac{1}{2}\sum_{(x,y)\in\eta^2_{t,\neq}}\mathbb{1}(\|x-y\|\le\delta_t)\|x-y\|^\alpha. \end{aligned}\qquad (2.6)$$

$\eta^2_{t,\neq}$ represents the set of all pairs of distinct points of $\eta_t$ and $\alpha \in \mathbb{R}$. For $\alpha = 0$, the functional counts the number of edges and for $\alpha = 1$, it counts the total edge length of the graph.

Next we observe that the degree $\deg(x)$ of a typical vertex $x$ in the graph is given by $\kappa_d t \delta_t^d$. We have by using the concept of Palm distributions and the Mecke formula 2.1 that

$$\begin{aligned}
\mathbb{E}\deg(x) &= \frac{1}{t}\mathbb{E}\sum_{x\in\eta_t} \mathbb{1}(x \in W) \\
&= \frac{1}{t}\mathbb{E}\sum_{x\in\eta_t} \eta_t(B(x,\delta_t)-1) \\
&= \int_W \eta_t(B(x,\delta_t))\,\mathrm{d}x \\
&= t\delta_t^d \kappa_d.
\end{aligned}$$

The quantity, $t\delta_t^d$, is essential to studying the behaviour of functionals and it naturally leads to three different asymptotic regimes as introduced by Penrose [Pen03].

- The *sparse* or *subcritical* regime, where $\lim\limits_{t\to\infty} t\delta_t^d = 0$.
- The *thermodynamic* or *critical* regime, where $\lim\limits_{t\to\infty} t\delta_t^d = c \in (0,\infty)$.
- The *dense* or *supercritical* regime, where $\lim\limits_{t\to\infty} t\delta_t^d = \infty$.

In this thesis in fact, we shall consider the thermodynamic regime in two parts, namely, for $c < 1$ and $c > 1$, as this is important for our results as we shall see later.

### 2.3.2. Random simplicial complexes

Let $\Delta$ be a collection of subsets of a set $V$. Then $\Delta$ is called an *abstract simplicial complex* if for every set $X \in \Delta$ and every non-empty $Y \subseteq X$, $Y \in \Delta$. The elements of $\Delta$ and $V$ are called *faces* (or *simplices*) and *vertices* of the complex respectively. It is assumed that $\{v\} \in \Delta$ for every $v \in V$. If $Y \subseteq X$, then $Y$ is said to belong to face $X$.

To define a geometric simplicial complex, we start from the building blocks of the complex, that is, the simplices. Let $k \geq 0$ and let $u_0, u_1, \ldots u_k$ be points in some Euclidean space $\mathbb{R}^n$. The point $x = \sum\limits_{i=0}^{k} c_i u_i$ is an affine combination of the $u_i$'s if $\sum\limits_{i=1}^{k} c_i = 1$, for non-negative $c_i \in \mathbb{R}$. The $k+1$ points are said to be affinely independent points in $\mathbb{R}^n$ if for any two affine combinations $x = \sum\limits_{i=1}^{k} c_i u_i$ and $y = \sum\limits_{i=1}^{k} c'_i u_i$, $x$ and $y$ are the same if and only if $c_i = c_i'$ for all $i$. In other words, the $k+1$ points are affinely independent if and only if the vectors $u_i - u_0$, $1 \leq i \leq k$, are linearly independent.

**Definition 2.7.** A $k$-simplex, $S$ in $\mathbb{R}^n$ is defined to be the convex hull of the $k+1$ affinely independent points, that is, a set of the form

$$S = \Big\{ \sum_{i=0}^{k} t_i u_i \mid t_i \geq 0, 0 \leq i \leq n, \sum_{i=0}^{n} t_i = 1 \Big\}.$$

The dimension of a $k$-simplex, $S$ is $k$.

**Definition 2.8.** A simplicial complex, $\Delta$, is a finite set of simplices in $\mathbb{R}^n$ such that

- every face of a simplex belonging to $\Delta$ is also in $\Delta$, that is, $S \in \Delta$ and $T \subseteq S$ implies $T \in \Delta$,
- the intersection of two simplices is either empty or a common face of both simplices.

The dimension of $\Delta$ is given by $\dim \Delta = \max\{\dim S : S \in \Delta\}$. A 1-dimensional simplicial complex is simply called a *graph*.

The collection of $i$-dimensional faces of $\Delta$ is denoted by $\mathcal{F}_i(\Delta)$, and $f_i(\Delta) = |\mathcal{F}_i(\Delta)|$. The $\boldsymbol{f}$-vector of $\Delta$ is given by

$$\boldsymbol{f}(\Delta) = (f_{-1}(\Delta), f_0(\Delta), f_1(\Delta), \ldots, f_{dim(\Delta)}(\Delta)).$$

Here, $f_{-1}(\Delta)$ counts the empty set and it is usually 1.

The notion of randomness in simplicial complexes can be observed in various ways. The Erdös-Rényi graph for example has a deterministic vertex set and random edges. The approach in this thesis will be the Gilbert model where the vertex set is a random set. A random simplicial complex is in fact a generalized version of the Gilbert graph. We shall consider two different simplicial complexes which both have the Gilbert graph as their 1-skeleton.

The *Vietoris-Rips complex* is a random simplicial complex whose $k$-dimensional faces are the abstract simplices $\{x_0, \ldots, x_k\} \subset \eta_t$ iff $\|x_i - x_j\| \leq \delta_t$. I.e. the Vietoris-Rips complex is the clique complex of the Gilbert graph. The Gilbert graph is in fact the one-skeleton of the Vietoris-Rips complex. We denote the Vietoris-Rips complex by $\mathcal{R}(\eta_t, \delta_t)$ and its set of $k$ faces by $\mathcal{F}_k(\mathcal{R}(\eta_t, \delta_t))$. For $k \leq d$ the faces have a geometric realization which is just the convex hull $[x_0, \ldots, x_k] \subset W$ of $x_0, \ldots, x_k$.

We denote by $\Delta_s[x_0, \ldots x_k]$ the $k$-dimensional volume of the convex hull of the points $x_0, \ldots, x_k$ if all edges have length at most $s$, and set $\Delta_s[x_0, \ldots, x_k] = 0$ otherwise. For $k > d$ there is no $k$-dimensional realization and thus in this case we just define

$\Delta_s[x_0, \dots x_k]^0 = 1$ if and only if all pairwise distances are bounded by $s$. Thus for all $k \geq 0$,

$$F \in \mathcal{F}_k(\mathcal{R}(\eta_t, \delta_t)) \Leftrightarrow \Delta_{\delta_t}(F)^0 = 1.$$

Note that for $k \leq d$ we identify the abstract simplex $F = \{x_0, \dots, x_k\} \in \mathcal{F}_k(\mathcal{V}(\eta_t, \delta_t))$ with its geometric realization $[x_0, \dots, x_k]$.

The notation $\Delta_s$ here should not be confused with previous notation $\Delta$ of a simplicial complex.

Closely connected to the Vietoris-Rips complex is the *Čech complex* $\mathcal{C}(\eta_t, \delta_t)$. This is the random simplicial complex where an abstract $k$-simplex $\{x_0, \dots, x_k\}$ is in $\mathcal{C}(\eta_t, \delta_t)$ if $\cap_{i=0}^k B(x_i, \delta_t/2) \neq \emptyset$. All results on the Vietoris-Rips complex in this thesis can verbatim be formulated for the Čech complex $\mathcal{C}(\eta_t, \delta_t)$ instead of the Vietoris-Rips complex. This just changes some of the constants, see Chapter 5.

The quantity we are mostly interested in this thesis is the volume-power functional of the Vietoris-Rips complex, defined as follows.

$$\begin{aligned}\mathcal{V}_k^{(\alpha)} &:= \frac{1}{(k+1)!} \sum_{F \in \mathcal{F}_k(\mathcal{R}(\eta_t, \delta_t))} \lambda_k(F)^\alpha \\ &= \frac{1}{(k+1)!} \sum_{(x_0, \dots, x_k) \in \eta^k_{t,\neq}} \Delta_{\delta_t}[x_0, \dots, x_k]^\alpha. \end{aligned} \tag{2.7}$$

If $k \leq d$, this functional is in $L^1(\mathbb{P})$ for $\alpha > -d + k - 1$, as will be observed later, and for $k > d$ we just consider the case $\alpha = 0$. We note that when $\alpha = 0$, $\mathcal{V}_k^{(\alpha)}$ counts the number of $k$-simplices in the simplicial complex, and thus, $\mathcal{V}_k^{(0)} = f_k$, a component of the $\boldsymbol{f}$-vector of the simplicial complex.

# 2.4. Counting partitions

It will interest us to know the number of partitions a set has, and the recurring function below gives this number for a typical partition. The *Bell number*, $B_n$ is the total number of patitions of a set with $n$ elements. It is given by

$$B_n = \sum_{i=0}^{n-1} \binom{n-1}{i} B_i.$$

In the sequel, we consider special types of partitions.

Let $f^{(l)}(x_1^{(l)}, \dots, x_{k_l}^{(l)}) \in L^2(\mu^{k_l})$ with $k_l \in \mathbb{N}$, for $l = 1, \dots, m$. The set of variables of $f^{(l)} : W^{k_l} \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, $l = 1, \dots, m$, will be used to create a partition as done in [Sch13].

The tensor product, $\otimes_{l=1}^{m} f^{(l)} : W^{\sum_{l=1}^{m} k_l} \to \bar{\mathbb{R}}$, of the functions $f^{(l)}$ is given by

$$\otimes_{l=1}^{m} f^{(l)}(x_1^{(1)}, \ldots, x_{k_m}^{(m)}) = \prod_{l=1}^{m} f^{(l)}(x_1^{(l)}, \ldots, x_{k_l}^{(l)}).$$

Let $V(k_1, \ldots, k_m) = \left\{x_1^{(1)}, \ldots, x_{k_1}^{(1)}, x_1^{(2)}, \ldots, x_{k_{m-1}}^{(m-1)}, x_1^{(m)}, \ldots, x_{k_m}^{(m)}\right\}$. Let $\bar{\pi} \in \mathcal{P}(V(k_1, \ldots, k_m))$ be the partition whose blocks are $\{x_1^{(l)}, \ldots, x_{k_l}^{(l)}\}$, $l = 1, \ldots, m$. Define

$$\begin{aligned} \Pi(k_1, \ldots, k_m) &= \left\{\sigma \in \mathcal{P}(V(k_1, \ldots, k_m)) : \sigma \wedge \bar{\pi} = \hat{0}\right\} \\ \tilde{\Pi}(k_1, \ldots, k_m) &= \left\{\sigma \in \mathcal{P}(V(k_1, \ldots, k_m)) : \sigma \wedge \bar{\pi} = \hat{0},\, \sigma \vee \bar{\pi} = \hat{1}\right\} \\ \Pi_{\geq 2}(k_1, \ldots, k_m) &= \left\{\sigma \in \mathcal{P}(V(k_1, \ldots, k_m)) : \sigma \wedge \bar{\pi} = \hat{0},\, |B| \geq 2 \,\forall\, B \in \sigma\right\} \end{aligned}$$

Clearly, both $\tilde{\Pi}(k_1, \ldots, k_m)$ and $\Pi_{\geq 2}(k_1, \ldots, k_m)$ are contained in $\Pi(k_1, \ldots, k_m)$.

For a partition, $\sigma \in \Pi(k_1, \ldots, k_m)$, we note from the definition that the number of blocks in the partition is bounded thus:

$$\max_{1 \leq i \leq m} k_i \leq |\alpha| \leq \sum_{i=1}^{m} k_i, \tag{2.8}$$

the upper bound is the case where all blocks are singletons and the lower bound is so because no block contains elements with the same index.

Also, we construct a new function, $(\otimes_{l=1}^{m} f^{(l)})_\sigma : W^{|\sigma|} \to \bar{\mathbb{R}}$, by replacing all variables that belong to the same block of $\sigma$ by a new common variable. We refer to Section 5.1 for an example. We shall be interested in finding the cardinality of $\Pi(k_1, \ldots, k_m)$ for $m \geq 2$.

In the literature [PT11], we found the MeetSolve function used in the Mathematica software, and it proves useful to get $|\Pi(k_1, \ldots, k_m)|$ numerically, but in this thesis, we present an analytical approach.

Let $m = 2$. Then $V(k_1, k_2) = \left\{x_1^{(1)}, \ldots, x_{k_1}^{(1)}, x_1^{(2)}, \ldots, x_{k_2}^{(2)}\right\}$. By definition, each partition in $\Pi(k_1, k_2)$ are such that no block contains identical upper indices. Thus, the blocks in a partition can be viewed as coincidences between the elements with upper index 1 and the elements with upper index 2. Let $h_1$ be the number of such coincidences for a partition $\sigma_{\{h_1\}} \in \Pi(k_1, k_2)$. Then $h_1 = 0, 1, \ldots, h_1^{max}$, where $h_1^{max} = \min\{k_1, k_2\}$ and the number of blocks in $\sigma_{\{h_1\}}$ is given by $|\sigma_{\{h_1\}}| = k_1 + k_2 - h_1$.

We need to know how many of such $\sigma_{\{h_1\}}$ exists in $\Pi(k_1, k_2)$ for a particular $h_1$. Taking the $k_1$ elements as the underlying set, it is a case of choosing $h_1$ out of $k_1$ elements and arranging $k_2$ elements $h_1$ at a time, that is, $\binom{k_1}{h_1} \cdot {}^{k_2}P_{h_1}$. The number of such $\sigma_{\{h_1\}}$ is thus given by

$$\alpha_{h_1} = h_1!\binom{k_1}{h_1}\binom{k_2}{h_1},$$

and

$$\left|\Pi(k_1,k_2)\right| = \sum_{h_1=0}^{h_1^{max}} \alpha_{h_1}.$$

We now take it a step further. Let $m = 3$. Then $V(k_1,k_2,k_3) = V(k_1,k_2) \cup \left\{x_1^{(3)},\ldots,x_{k_3}^{(3)}\right\}$. We take the partitions, $\sigma_{\{h_1\}}$, $h_1 = 0,1,\ldots,h_1^{max}$, in $\Pi(k_1,k_2)$ as the underlying set, and let $h_2$ be the number of coincidences between the underlying set and $\left\{x_1^{(3)},\ldots,x_{k_3}^{(3)}\right\}$ then $h_2 = 0,1,\ldots,h_2^{max}$, where $h_2^{max} = \min\{k_1+k_2-h_1, k_3\}$. Let the resulting partitions be given by $\sigma_{\{h_1+h_2\}} \in \Pi(k_1,k_2,k_3)$, then the number of blocks in $\sigma_{\{h_1+h_2\}}$ is given by $|\sigma_{\{h_1+h_2\}}| = k_1+k_2+k_3-(h_1+h_2)$. We need to choose $h_2$ out of $k_1+k_2-h_1$ elements and arrange $k_3$ elements $h_2$ at a time. Again, the number of such $\sigma_{\{h_1+h_2\}}$ is given by $\alpha_{\{h_1+h_2\}} = \alpha_{h_1}\cdot\alpha_{h_2}$ where

$$\alpha_{h_2} = h_2!\binom{k_1+k_2-h_1}{h_2}\binom{k_3}{h_2},$$

and

$$\left|\Pi(k_1,k_2,k_3)\right| = \sum_{h_1+h_2=0}^{h_1^{max}+h_2^{max}} \alpha_{h_1}\cdot\alpha_{h_2}.$$

We note that the above sum includes all possible values of of $h_1$ and $h_2$ for the sum $h_1+h_2$.

In general for $2 \le l \le m$, we have the following. First, we define

$$K_j = \sum_{i=1}^{j} k_i \quad \text{and} \quad H_j = \sum_{i=0}^{j} h_i \quad \text{with} \quad h_0 = 0.$$

We have $V(k_1,k_2,\ldots,k_{l+1}) = V(k_1,\ldots,k_l) \cup \left\{x_1^{(l+1)},\ldots,x_{k_{l+1}}^{(l+1)}\right\}$. We take the partitions $\sigma_{\{H_{l-1}\}} \in \Pi(k_1,\ldots,k_l)$ as the underlying set and $h_l$ as the number of coincidences, with $h_l = 0,1,\ldots,h_l^{max}$, where $h_l^{max} = \min\{K_l - H_{l-1}, k_{l+1}\}$. The number of blocks of the resulting partitions $\sigma_{\{H_l\}} \in \Pi(k_1,\ldots,k_{l+1})$ is given by $|\sigma_{\{H_l\}}| = K_{l+1} - H_l$. The number of such $\sigma_{\{H_l\}}$ is given by $\alpha_{\{H_l\}} = \prod_{j=1}^{l} \alpha_{h_j}$ with

$$\alpha_{h_j} = h_j!\binom{K_j - H_{j-1}}{h_j}\binom{k_{j+1}}{h_j},$$

and

$$\left|\Pi(k_1,\ldots,k_{l+1})\right| = \sum_{H_l=0}^{H_l^{max}} \alpha_{\{H_l\}}$$

where $H_l^{max} = \sum_{i=1}^{l} h_i^{max}$.

The above can be summarized in the following theorem.

**Theorem 2.9.** *Let $m \geq 2$. For $i = 1, \ldots, m$, let $k_i \in \mathbb{N}$ and $h_i = 0, \ldots, h_i^{max}$, where $h_i^{max} = \min\{K_i - H_{i-1}, k_{i+1}\}$.*
*Define*

$$K_j = \sum_{i=1}^{j} k_i \quad \text{and} \quad H_j = \sum_{i=0}^{j} h_i \quad \text{with} \quad h_0 = 0.$$

*Furthermore, define*

$$\alpha_{h_j} = h_j! \binom{K_j - H_{j-1}}{h_j} \binom{k_{j+1}}{h_j}.$$

*Then for $z = H_{m-1}$, there exists $\sigma_{\{z\}} \in \Pi(k_1, \ldots, k_m)$ such that $|\sigma_{\{z\}}| = K_m - z$ and*

$$\Big|\Pi(k_1, \ldots, k_m)\Big| = \sum_{z=0}^{H_{m-1}^{max}} \alpha_{\{z\}},$$

*where $\alpha_{\{z\}} = \prod_{j=1}^{m-1} \alpha_{h_j}$ and $H_{m-1}^{max} = \sum_{i=1}^{m-1} h_i^{max}$.*

*Remark* 2.10.

(i) Let $k_1 \leq \cdots \leq k_m$. Then $H_{m-1}^{max} = \sum_{i=1}^{m-1} h_i^{max} = K_{m-1}$, since $h_i^{max} = k_i$ for each $i$.

(ii) Let $k_1 = \cdots = k_m = k$. Then $h_i^{max} = k \;\; \forall\, i$ and

$$\alpha_{h_j} = h_j! \binom{jk - H_{j-1}}{h_j} \binom{k}{h_j}.$$

Also, $K_i = ik$, $H_i^{max} = ik \;\; \forall\, i$ and with $\Pi(\underbrace{k, \ldots, k}_{m \text{ times}}) = \Pi(\{k\}_m)$,

$$\Big|\Pi(\{k\}_m)\Big| = \sum_{z=0}^{k(m-1)} \alpha_{\{z\}}.$$

(iii) For $k_1 = \cdots = k_m = k$, we will be interested in finding $\alpha_{\{z\}}$ for which $|\sigma_{\{z\}}| = k$. That is, $|\sigma_{\{k(m-1)\}}| = mk - k(m-1) = k$. The only possibility to have $z = k(m-1)$ is when $h_i = k$ for $i = 1, \ldots, m-1$. This gives $\alpha_{h_j = k} = k!$ for $j = 1, \ldots, m-1$, and so, $\alpha_{\{z\}} = \prod_{j=1}^{m-1} \alpha_{h_j = k} = k!^{(m-1)}$.

(iv) Furthermore, still for $k_1 = \cdots = k_m = k$, to find $\alpha_{\{z\}}$ for which $|\sigma_{\{z\}}| = k+1$, it has to be that $z = k(m-1) - 1$. We observe that this is possible when only one of the $h_i$'s is $k-1$ and all others are $k$. Going through all these possibilities gives $\alpha_{\{z\}} = kk!^{m-1} \sum_{i=0}^{m-2} (k+1)^i$.

# 3. Expectation

Moments are used in describing the behaviour of random variables. The $p$-th moment of a random varaible, $X$ is given by $\mathbb{E}X^p$. In this chapter, we consider the first moment of Poisson functionals which is precisely the expectation in three asymptotic regimes as mentioned earlier.

**Definition 3.1.** A *Poisson U-statistic*, $F$, of order $k$, is a Poisson functional of the form

$$F = \sum_{(x_1,\ldots,x_k)\in\eta^k_{\neq}} f(x_1,\ldots,x_k)$$

where $k \in \mathbb{N}$ and $f \in L^1_s(\mu^k)$.

The expectation of a Poisson U-statistic is given by

$$\mathbb{E}F = \int_{W^k} f(x_1,\ldots,x_k)\ \mathrm{d}x_1,\ldots,\mathrm{d}x_k.$$

This is obtainable by applying the Mecke formula (2.1), $f$ being integrable.

By the above definition, we observe that the length-power functional, $\mathcal{V}^{(\alpha)}$, and the volume-power functional, $\mathcal{V}^{(\alpha)}_k$, are Poisson U-statistics of orders 2 and $k+1$ respectively.

We take note of the following.

*Remark* 3.2.

(i) The Gamma function is given by

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\ \mathrm{d}x.$$

Furthermore, by simply applying integration by parts in the above, we have the following recurring formula.

$$\Gamma(z+1) = z\Gamma(z).$$

(ii) In terms of the Gamma function, the volumes of the unit $d$-ball, $B^d$, and its boundary, the unit $(d-1)$-sphere, $S^{d-1}$, are respectively given by

$$\kappa_d = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)} \quad \text{and} \quad \omega_d = \mathrm{Vol}(S^{d-1}) = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} = d\kappa_d$$

(iii) Let $W_{-\delta_t} = \{w\colon B^d(w,\delta_t) \subset W\}$, be the inner parallel set of $W$. Then for a point $x \in W_{-\delta_t}$, we have

$$\begin{aligned}
\int_W \mathbb{1}(\|x-y\| \le \delta_t)\|x-y\|^\alpha \,\mathrm{d}y &= \int_{W-x} \mathbb{1}(\|y\| \le \delta_t)\|y\|^\alpha \,\mathrm{d}y \\
&= \delta_t^{\alpha+d} \int_{\delta_t^{-1}(W-x)} \mathbb{1}(\|y\| \le 1)\|y\|^\alpha \,\mathrm{d}y \\
&= \delta_t^{\alpha+d} \int_{\delta_t^{-1}(W-x)\cap B^d} \|y\|^\alpha \,\mathrm{d}y \\
&= \delta_t^{\alpha+d} \int_{B^d} \|y\|^\alpha \,\mathrm{d}y \\
&= \delta_t^{\alpha+d} \int_0^1 r^{\alpha+d-1} \int_{S^{d-1}} \,\mathrm{d}u\mathrm{d}r \\
&= \delta_t^{\alpha+d} \frac{\omega_d}{\alpha+d} \\
&= \frac{d\kappa_d}{\alpha+d}\delta_t^{\alpha+d}.
\end{aligned}$$

In the above, we translated by $x$ and scaled by $\delta_t$, and used the fact that $\delta_t^{-1}(W - x) \cap B^d = B^d$ since the ball around the point $x \in W_{-\delta_t}$ is completely contained in $W$. We also transformed to sperical coordinates where $\mathrm{d}u$ stands for the infinitesimal element of the spherical Lebesgue measure.

In the same vein, for $x \in W \backslash W_{-\delta_t}$, we have that

$$\int_W \mathbb{1}(\|x-y\| \le \delta_t)\|x-y\|^\alpha \,\mathrm{d}y \le \frac{d\kappa_d}{\alpha+d}\delta_t^{\alpha+d}.$$

# 3.1. Functionals of random graphs

A fundamental quantity in random graphs is counting the number of edges. Penrose in [Pen03] considered it in a more general context by counting the number of subgraphs isomorphic to a given and fixed connected graph, and also component counts. In [RST17], the number of edges of the Gilbert graph was not only considered, but more generally the sum of real powers of the edge lengths. That is,

$$
\begin{aligned}
\mathcal{V}^{(\alpha)} &:= \frac{1}{2} \sum_{e=(x,y)\in E(G(\eta_t,\delta_t))} \|x-y\|^{\alpha} \\
&= \frac{1}{2} \sum_{(x,y)\in\eta^2_{t,\neq}} \mathbb{1}(\|x-y\| \le \delta_t)\|x-y\|^{\alpha}.
\end{aligned} \tag{3.1}
$$

The expectation is given by the theorem below.

**Theorem 3.3.** *Let $\alpha > -d$, and $c_{\alpha,d} = \dfrac{d\kappa_d}{2(\alpha+d)}$. Then*

$$
\mathbb{E}\mathcal{V}^{(\alpha)} = c_{\alpha,d}\, t^2 \delta_t^{\alpha+d}(1 + O(\delta_t)) \tag{3.2}
$$

*where $c_{\alpha,d}$ is a constant that depends on the space dimension $d$, and $\alpha \in \mathbb{R}$.*

*Proof.* Applying the Mecke formula to (3.1) gives

$$
\begin{aligned}
\mathbb{E}\mathcal{V}^{(\alpha)} &= \frac{t^2}{2} \int_{W^2} \mathbb{1}(\|x-y\| \le \delta_t)\|x-y\|^{\alpha} \,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{t^2}{2} \int_{W}\int_{W} \mathbb{1}(\|x-y\| \le \delta_t)\|x-y\|^{\alpha} \,\mathrm{d}x\,\mathrm{d}y \\
&\le c_{\alpha,d}\, t^2 \delta_t^{\alpha+d} \qquad \text{(by Remark 3.2, since } W \text{ is of unit volume).}
\end{aligned}
$$

Also,

$$
\mathbb{E}\mathcal{V}^{(\alpha)} \ge c_{\alpha,d}\, t^2 \delta_t^{\alpha+d} V(W_{-\delta_t})
$$

which together with the above and the fact that

$$
V(W_{-\delta_t}) \ge 1 - S(W)\delta_t
$$

gives (3.2). □

*Remark* 3.4. We observe the asymptotic expectation by considering the term $t(t\delta_t^d)\delta_t^{\alpha}$ in (3.2) for the different regimes. We note that the fact that $t\delta_t^d$ tends to 0, a constant $c \in (0,\infty)$ or $\infty$ as $t \to \infty$, in the sparse, thermodynamic and dense regimes respectively, does not imply that the expectation will behave accordingly, and that, because of $\delta_t^{\alpha}$. The value of $\alpha$ can change a lot of things as $\delta_t \to 0$.

For example, in the dense regime, for $\alpha > d$, if we take $\delta_t = t^{-\frac{1}{d}} f(t)$, where $f(t) \to \infty$ slowly, say $f(t) = \ln t$, then $\delta_t \to 0$ and $t\delta_t^d \to \infty$ slowly. Also, $t\delta_t^{\alpha} = t^{1-\frac{\alpha}{d}} f(t)^{\alpha} \to 0$ since the first factor goes to 0 faster than the second goes to infinity. Thus, the expectation tends to 0 in this case.

# 3.2. Functionals of random simplicial complexes

In this section, we count the simplices in the Vietoris-Rips complex, which could be seen as complete subgraph counts, and more generally the sum of real powers of the volumes of these simplices.

It was important to know the values of $\alpha$ in the computation of the expectation of the length-power functional, for which the constant $c_{\alpha,d}$ is finite, which obtains in the condition of the Theorem 3.3. In the sequel, we will consider the values of $\alpha$ for which the constant we will encounter in the expectation of the volume-power functional, $\mathbb{E}\mathbb{V}_k^\alpha$, will be finite.

## 3.2.1. Moments of random simplices

Let $X_1, \ldots, X_k$ be $k$ independently and uniformly distributed points in the unit ball. To shorten our notation, we write $\{x_l\}_{l=j}^k$ for the point set $\{x_j, \ldots, x_k\}$. The random points $\{X_l\}_{l=1}^k$ and the origin form a random $k$-simplex. For $k \le d$ we denote by $\mu_k^{(\alpha)}$ the moment of order $\alpha$ of its volume if all edges are bounded by one.

$$\mu_k^{(\alpha)} = \int\limits_{(B^d)^k} \Delta_1[0, \{x_l\}_{l=1}^k]^\alpha \, \mathrm{d}x_1 \cdots \mathrm{d}x_k \tag{3.3}$$

with $\mu_0^{(\alpha)} = 1$. In the case $k > d$ the definition only applies to $\alpha = 0$ which then just tests whether all pairwise distances are bounded by 1. It is a key quantity in the expectation of $\mathbb{E}\mathbb{V}_k^\alpha$, and the essential question is for which $\alpha \in \mathbb{R}$ this quantity is finite.

To derive moments of random simplices obtained by the convex hull of uniform points in the unit ball is a classical question in geometric probabilities. There is an elegant way of computing arbitrary moments of the volume of the convex hull $\Delta[0, \{x_l\}_{l=1}^k]^\alpha$ which is shorthand for $\Delta_\infty[0, \{x_l\}_{l=1}^k]^\alpha = \lambda_d([0, \{x_l\}_{l=1}^k])^\alpha$,

$$\nu_k^{(\alpha)} = \int\limits_{(B^d)^k} (\Delta[0, \{x_l\}_{l=1}^k])^\alpha \mathrm{d}x_1 \ldots \mathrm{d}x_k.$$

We follow the computations in Schneider and Weil [SW08, Section 8.2.2]. We note that the mentioned reference is interested only in the cases $\alpha \in \mathbb{N}$ where the result can be rewritten as a product of certain $\kappa_j$ using its defintion and the representation of the Beta function by Gamma functions.

Assume $L$ is an arbitrary $q$-dimensional subspace in $\mathbb{R}^d$. Denote by $d(x, L)$ the distance of a point $x$ to $L$. We apply the proof of [SW08, Theorem 8.2.2] for $(q-1)$-dimensional $L$ where Fubini's theorem was applied. In our case, $L$ is of $q$-dimension. We take $\alpha \in \mathbb{R}$, and using the definitions in Remark 3.2(ii), we have

$$\begin{aligned}
\int_{B^d} d(x,L)^\alpha \mathrm{d}x &= \kappa_{d+\alpha} \frac{\omega_{d-q}}{\omega_{d+\alpha-q}} \\
&= \frac{\pi^{\frac{d+\alpha}{2}}}{\Gamma\left(\frac{d+\alpha}{2}+1\right)} \cdot \frac{2\pi^{\frac{d-q}{2}}}{\Gamma\left(\frac{d-q}{2}\right)} \cdot \frac{\Gamma\left(\frac{d+\alpha-q}{2}\right)}{2\pi^{\frac{d+\alpha-q}{2}}} \\
&= \frac{\pi^{\frac{d}{2}}\Gamma(\frac{d+\alpha-q}{2})}{\Gamma(\frac{d-q}{2})\Gamma(\frac{d+\alpha}{2}+1)} \\
&= \beta_q^{(\alpha)}.
\end{aligned} \tag{3.4}$$

Note that the Gamma function, $\Gamma(x)$, is finite for $x > 0$. Thus, $\beta_q^{(\alpha)}$ is finite for $d+\alpha-q > 0$. Since the volume of the simplex $[0, x_1, \dots, x_k]$ is given by

$$\Delta[0, \{x_l\}_{l=1}^k] = \frac{1}{k}\Delta[0, x_1, \dots, x_{k-1}]d(x_k, L) \tag{3.5}$$

where $L$ is the linear hull of $x_1, \dots, x_k$, having dimension $k-1$, then we have by applying (3.5) that

$$\nu_k^{(\alpha)} = \int_{(B^d)} \int_{(B^d)^{k-1}} \frac{1}{k^\alpha}(\Delta[0, \{x_l\}_{l=1}^{k-1}])^\alpha \, d(x_k, L)^\alpha \mathrm{d}x_1 \dots \mathrm{d}x_k = \frac{1}{k^\alpha}\beta_{k-1}^{(\alpha)}\nu_{k-1}^{(\alpha)} \tag{3.6}$$

so that we obtain as in [SW08] the recursion

$$\nu_k^{(\alpha)} = \frac{1}{k^\alpha}\beta_{k-1}^{(\alpha)}\nu_{k-1}^{(\alpha)} = \frac{1}{k^\alpha(k-1)^\alpha}\beta_{k-1}^{(\alpha)}\beta_{k-2}^{(\alpha)}\nu_{k-2}^{(\alpha)} = \dots = \frac{1}{(k!)^\alpha}\prod_{j=0}^{k-1}\beta_j^{(\alpha)} \tag{3.7}$$

(where $\nu_0^{(\alpha)} = 1$). This implies that $\nu_k^{(\alpha)} < \infty$ if $\beta_j^{(\alpha)} < \infty$ for $j = 0, \dots, k-1$, that is, when $d + \alpha - (k-1) > 0$ implying $\alpha > -d + k - 1$. Because $\Delta_1[\cdot]^\alpha = \Delta[\cdot]^\alpha$ except if $\Delta_1$ vanishes, we obtain

$$\mu_k^{(\alpha)} \leq \nu_k^{(\alpha)}$$

and thus we see that $\mu_k^{(\alpha)}$ exists as long as $\alpha > -d + k - 1$. On the other hand, for $x_i \in \frac{1}{2}B^d$, all distances are trivially bounded by 1, which shows

$$\mu_k^{(\alpha)} \geq \int_{(\frac{1}{2}B^d)^k} \Delta[0, x_1, \dots, x_k])^\alpha \mathrm{d}x_1 \dots \mathrm{d}x_k = \left(\frac{1}{2}\right)^{k(\alpha+d)} \nu_k^{(\alpha)}.$$

The above proves the following lemma.

**Lemma 3.5.** *$\mu_k^{(\alpha)}$ exists if and only if $\alpha > -d + k - 1$.*

### 3.2.2. First moment for the volume-power functional

Having obtained the threshold for $\alpha$, we consider the volume-power functional. Recall from (2.7) that

$$\mathcal{V}_k^{(\alpha)} := \frac{1}{(k+1)!} \sum_{(x_0,\dots,x_k)\in\eta_{t,\neq}^k} \Delta_{\delta_t}[x_0,\dots,x_k]^\alpha. \tag{3.8}$$

**Theorem 3.6.** *Assume $\alpha > -d + k - 1$ for $k \leq d$ and $\alpha = 0$ for $k > d$. Then we have*

$$\mathbb{E}\mathcal{V}_k^{(\alpha)} = \frac{\mu_k^{(\alpha)}}{(k+1)!} t^{k+1} \delta_t^{k(\alpha+d)} (1 + O(\delta_t))$$

*where the implicit constant in $O(\delta_t)$ only depends on $W$.*

*Proof.* We apply the multivariate Mecke formula (2.1) to (3.8), and substitute $\delta_t(x_i - x_k)$ for $x_i$, $i \neq k$, to get

$$\begin{aligned}
\mathbb{E}\mathcal{V}_k^{(\alpha)} &= \frac{t^{k+1}}{(k+1)!} \int\limits_{W^{k+1}} \Delta_{\delta_t}[\{x_l\}_{l=0}^k]^\alpha \, \mathrm{d}x_0 \cdots \mathrm{d}x_k \\
&= \frac{\delta_t^{k(\alpha+d)} t^{k+1}}{(k+1)!} \int\limits_W \int\limits_{(\delta_t^{-1}(W-x_k)\cap B^d)^k} \Delta_1[0, \{x_l\}_{l=0}^{k-1}]^\alpha \, \mathrm{d}x_0 \cdots \mathrm{d}x_k
\end{aligned}$$

where the condition $\|x_i\| \leq 1$ has been taken into account in the range of integration $x_i \in B^d$. As an upper bound we have

$$\mathbb{E}\mathcal{V}_k^{(\alpha)} \leq \frac{\delta_t^{k(\alpha+d)} t^{k+1}}{(k+1)!} \int\limits_W \int\limits_{(B^d)^k} \Delta_1[0, \{x_l\}_{l=0}^{k-1}]^\alpha \, \mathrm{d}x_0 \cdots \mathrm{d}x_k.$$

This proves that the expectation is finite for $\alpha > -d$. For an estimate from below we consider the inner parallel set of $W$, $W_{-\delta_t} = \{x \colon B^d(x, \delta_t) \subset W\}$. Observe that for $x_k \in W_{-\delta_t}$ we have $\delta_t^{-1}(W - x_k) \cap B^d = B^d$.

$$\begin{aligned}
\mathbb{E}\mathcal{V}_k^{(\alpha)} &\geq \frac{\delta_t^{k(\alpha+d)} t^{k+1}}{(k+1)!} \int\limits_{W_{-\delta_t}} \int\limits_{(B^d)^k} \Delta_1[0, \{x_l\}_{l=0}^{k-1}]^\alpha \mathrm{d}x_0 \cdots \mathrm{d}x_k \\
&= \frac{\delta_t^{k(\alpha+d)} t^{k+1}}{(k+1)!} \mu_k^{(\alpha)} V(W_{-\delta_t})
\end{aligned}$$

We thus obtain

$$\frac{\delta_t^{k(\alpha+d)} t^{k+1}}{(k+1)!} \mu_k^{(\alpha)} V(W_{-\delta_t}) \le \mathbb{E}\mathcal{V}_k^{(\alpha)} \le \frac{\delta_t^{k(\alpha+d)} t^{k+1}}{(k+1)!} \mu_k^{(\alpha)}.$$

The well known inequality

$$V(W_{-\delta_t}) \ge 1 - S(W)\delta_t \tag{3.9}$$

for convex sets of volume one gives the desired result . □

*Remark* 3.7. Once again, we observe the expectation in the various asymptotic regimes as in remark 3.4. We consider the term $t(t\delta_t^d)^k \delta_t^{\alpha k}$.

Consider the sparse regime, if we take $\delta_t = t^{-\frac{1}{d}} f(t)^{\frac{1}{dk}}$, where $f(t) \to 0$ slowly, say $f(t) = e^{-t}$, then $\delta_t \to 0$ and $t\delta_t^d \to 0$ slowly and $(t\delta_t^d)^k = f(t)$. Also, $t\delta_t^{\alpha k} = t^{1-\frac{\alpha k}{d}} f(t)^{\frac{\alpha}{d}} \to 0$ for $\alpha < \frac{d}{k}$, since the first factor goes to infinity faster than the second goes to zero.

Thus, once again, the fact that $t\delta_t^d$ tends to 0, a constant $c \in (0, \infty)$ or $\infty$ as $t \to \infty$, in the sparse, thermodynamic and dense regimes respectively, does not imply that the expectation will behave accordingly, and that, because of $\delta_t^{\alpha k}$ in this case. So the asymptotic expectation depends on the value of $\alpha$ and how fast $\delta_t \to 0$.

# 4. Variance

In this chapter, we consider the covariance structure, in particular, the second moment of the volume-power functional. Recall that the covariance of two random variables $X$ and $Y$ is given by

$$\mathbb{C}\text{ov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y.$$

Once again, we take a look at the moments of random simplices similar to the one discussed in Secction 3.2.

## 4.1. Mixed moments of random simplices

In investigating the covariance structure of the volume-power functionals, the mixed moments of the volume of two random simplices in $B^d$ with edge lengths bounded by one occurs. For this, we put one vertex in the origin, choose in $B^d$ for a $k_1$-dimensional simplex first $k_1$ independent vertices $\{X_1, \dots, X_{k_1}\}$. For $1 \leq m \leq \min\{k_1, k_2\}+1$, we use the last $m-1$ vertices also for a second simplex and in addition independent random vertices $\{X_{k_1+1} \dots X_{k_1+k_2-m+1}\}$ to define a $k_2$-dimensional simplex $\Delta_1[0, \{X_l\}_{l=k_1-m+2}^{k_1+k_2-m+1}]$. And so we have the following as the mixed moment of the volume of the simplices.

$$\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)} = \int\limits_{(B^d)^{k_1+k_2+1-m}} \Delta_1[0, \{x_l\}_{l=1}^{k_1}]^{\alpha_1} \Delta_1[0, \{x_l\}_{l=k_1-m+2}^{k_1+k_2-m+1}]^{\alpha_2} \mathrm{d}x_1 \cdots \mathrm{d}x_{k_1+k_2-m+1} \tag{4.1}$$

It can be seen that for $m = 1$,

$$\mu_{k_1,k_2:1}^{(\alpha_1,\alpha_2)} = \mu_{k_1}^{(\alpha_1)} \mu_{k_2}^{(\alpha_2)}. \tag{4.2}$$

And for $k_1 = 0$ we have $\mu_{0,k_2:m}^{(\alpha_1,\alpha_2)} = \mu_{k_2}^{(\alpha_2)}$, also for $k_1 = k_2 = k$ at $m = k+1$, we have

$$\mu_{k,k:k+1}^{(\alpha_1,\alpha_2)} = \mu_{k}^{(\alpha_1+\alpha_2)}. \tag{4.3}$$

This is the instance where all points in the identical simplices coincide. For $k_i > d$ we only allow $\alpha_i = 0$. Again, we need to determine for which $\alpha_i$ this moment exists. We let $k_1 \leq k_2$ without loss of generality, and we follow the strategy outlined in the previous chapter. We set

$$\nu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)} = \int\limits_{(B^d)^{k_1+k_2+1-m}} \Delta[0,\{x_l\}_{l=1}^{k_1}]^{\alpha_1}\Delta[0,\{x_l\}_{l=k_1-m+2}^{k_1+k_2-m+1}]^{\alpha_2}\mathrm{d}x_1\cdots\mathrm{d}x_{k_1+k_2-m+1}$$

where $\Delta[0,\{x_l\}_{l=1}^{k}]^{\alpha}$ represents $\Delta_\infty[0,\{x_l\}_{l=1}^{k}]^{\alpha} = \lambda_d([0,\{x_l\}_{l=1}^{k}])^{\alpha}$.

We relabel to let the first $m-1$ vertices of the first simplex coincide with the first $m-1$ vertices of the second simplex. That is, in the first simplex each $x_i$ becomes $x_{k_i-i+1}$ and in the second simplex, each $x_i$ becomes $\hat{x}_{i-(k_i-m+1)}$, while $x_i = \hat{x}_i$ for $i = 1,\ldots,m-1$. So we obtain

$$\nu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)} = \int\limits_{(B^d)^{k_1+k_2+1-m}} \Delta[0,\{x_l\}_{l=1}^{k_1}]^{\alpha_1}\Delta[0,\{\hat{x}_l\}_{l=1}^{k_2}]^{\alpha_2}\mathrm{d}x_1\cdots\mathrm{d}x_{m-1}\mathrm{d}x_m\cdots\mathrm{d}x_{k_1}\mathrm{d}\hat{x}_m\cdots\mathrm{d}\hat{x}_{k_2}$$

We recall from (3.5) that

$$\Delta[0,\{x_l\}_{l=1}^{k_i}] = \frac{1}{k_i}\Delta[0,\{x_l\}_{l=1}^{k_i-1}]\,d(x_{k_i},L_i), \quad i=1,2$$

where $L_i$ of dimension $k_i-1$, so that using (3.4) in the above gives

$$\begin{aligned}
\nu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)} &= \int\limits_{B^d}\frac{1}{k_1^{\alpha_1}}\int\limits_{B^d}\frac{1}{k_2^{\alpha_2}}\int\limits_{(B^d)^{k_1+k_2-m-1}} \Delta[0,\{x_l\}_{l=1}^{k_1-1}]^{\alpha_1}\Delta[0,\{x_l\}_{l=1}^{k_2-1}]^{\alpha_2}d(x_{k_1},L_1)^{\alpha_1} \\
&\qquad d(x_{k_2},L_2)^{\alpha_2}\ \mathrm{d}x_1\cdots\mathrm{d}x_{m-1}\mathrm{d}x_m\cdots\mathrm{d}x_{k_1-1}\mathrm{d}x_m\cdots\mathrm{d}x_{k_2-1} \\
&= \frac{1}{k_1^{\alpha_1}}\frac{1}{k_2^{\alpha_2}}\beta_{k_1-1}^{(\alpha_1)}\beta_{k_2-1}^{(\alpha_2)}\nu_{k_1-1,k_2-1:m}^{(\alpha_1,\alpha_2)} \\
&= \frac{1}{k_1^{\alpha_1}(k_1-1)^{\alpha_1}}\frac{1}{k_2^{\alpha_2}(k_2-1)^{\alpha_2}}\beta_{k_1-1}^{(\alpha_1)}\beta_{k_1-2}^{(\alpha_1)}\beta_{k_2-1}^{(\alpha_2)}\beta_{k_2-2}^{(\alpha_2)}\nu_{k_1-2,k_2-2:m}^{(\alpha_1,\alpha_2)} \\
&= \frac{1}{k_1^{\alpha_1}(k_1-1)^{\alpha_1}\cdots m^{\alpha_1}}\frac{1}{k_2^{\alpha_2}(k_2-1)^{\alpha_2}\cdots m^{\alpha_2}}\prod_{j=m-1}^{k_1-1}\beta_i^{(\alpha_1)}\prod_{j=m-1}^{k_2-1}\beta_i^{(\alpha_2)}\nu_{m-1,m-1:m}^{(\alpha_1,\alpha_2)} \\
&= \frac{((m-1)!)^{\alpha_1}}{(k_1!)^{\alpha_1}}\frac{((m-1)!)^{\alpha_2}}{(k_2!)^{\alpha_2}}\prod_{j=m-1}^{k_1-1}\beta_j^{(\alpha_1)}\prod_{j=m-1}^{k_2-1}\beta_j^{(\alpha_2)}\ \nu_{m-1,m-1:m}^{(\alpha_1,\alpha_2)} \\
&= \frac{((m-1)!)^{\alpha_1+\alpha_2}}{(k_1!)^{\alpha_1}(k_2!)^{\alpha_2}}\prod_{j=m-1}^{k_1-1}\beta_j^{(\alpha_1)}\prod_{j=m-1}^{k_2-1}\beta_j^{(\alpha_2)}\ \nu_{m-1}^{(\alpha_1+\alpha_2)} \qquad \text{(by (4.3))} \\
&= \frac{1}{(k_1!)^{\alpha_1}(k_2!)^{\alpha_2}}\prod_{j=m-1}^{k_1-1}\beta_j^{(\alpha_1)}\prod_{j=m-1}^{k_2-1}\beta_j^{(\alpha_2)}\prod_{j=0}^{m-2}\beta_j^{(\alpha_1+\alpha_2)}
\end{aligned}$$

where in the last expression we used (3.7).

We note that $\nu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}$ is finite if all the $\beta_j$'s are finite, that is,

$$d + \alpha_1 - (k_1 - 1) > 0, \quad d + \alpha_2 - (k_2 - 1) > 0, \quad d + \alpha_1 + \alpha_2 - (m - 2) > 0. \tag{4.4}$$

Since $\Delta_1[\cdot]^\alpha = \Delta[\cdot]^\alpha$ except if $\Delta_1$ vanishes, we have

$$\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)} \leq \nu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)},$$

so that $\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)} < \infty$ once (4.4) is satisfied.

To estimate $\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}$ from below, we consider $x_i \in \frac{1}{2}B^d$, and observe that all distances are trivially bounded by 1, so that

$$\begin{aligned} \mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)} &\geq \int\limits_{(\frac{1}{2}B^d)^{k_1+k_2+1-m}} \Delta[0,\{x_l\}_{l=1}^{k_1}]^{\alpha_1} \Delta[0,\{x_l\}_{l=k_1-m+2}^{k_1+k_2-m+1}]^{\alpha_2} \, \mathrm{d}x_1 \cdots \mathrm{d}x_{k_1+k_2-m+1} \\ &= \left(\frac{1}{2}\right)^{d(k_1+k_2-m+1)+\alpha_1 k_1+\alpha_2 k_2} \nu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}. \end{aligned}$$

This proves the following lemma.

**Lemma 4.1.** $\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}$ *exists if and only if* $\alpha_i > -d + k_i - 1$ *for* $i = 1, 2$, *and* $\alpha_1 + \alpha_2 > -d + m - 2$.

# 4.2. Moment matrices

To be able to discuss the features of the covariance matrix in the covariance structure of $\mathcal{V}_k^{(\alpha)}$, we give a brief overview of moment matrices.

Let $X$ be a random variable, and let $\boldsymbol{c} = (c_1, \ldots, c_n) \in \mathbb{R}^n$ be chosen such that $m_{c_i+c_j} = \mathbb{E}X^{c_i+c_j}$ exists for $i, j = 1, \ldots, n$. Denote by $M_X(\boldsymbol{c})$ the generalized moment matrix

$$M_X(\boldsymbol{c}) = (m_{c_i+c_j})_{i,j=1,\ldots,n}.$$

The following theorem gives a criterion whether the generalized moment matrix is of full rank.

**Theorem 4.2.** *The generalized moment matrix* $M_X(\boldsymbol{c})$ *is positive semidefinite. Moreover* $M_X(\boldsymbol{c})\, p = 0$ *for some* $p \in \mathbb{R}^n$ *implies*

$$X \in \left\{ x \in \mathbb{R} \colon \sum_{i=1}^n p_i x^{c_i} = 0 \right\} \quad a.s.$$

*If, in particular, $c_i \neq c_j$ for all $i \neq j$, and* $\operatorname{supp}(X)$ *contains an interval, then $M_X(\boldsymbol{c})$ is of full rank.*

*Proof.* The proof is a modification of similar results on moment matrices, see for example [Lau09]. For $p \in \mathbb{R}^n$ we have

$$p^T M_X(\boldsymbol{c})p = \sum_{ij} p_i m_{c_i+c_j} p_j = \mathbb{E} \sum_{ij} p_i X^{c_i} X^{c_j} p_j = \mathbb{E}(\sum_i p_i X^{c_i})^2 \geq 0$$

and thus $M_X(\boldsymbol{c})$ is positive semidefinite. Further, if $M_X(\boldsymbol{c})p = 0$, then also

$$p^T M_X(\boldsymbol{c})p = \mathbb{E}(\sum_i p_i X^{c_i})^2 = 0.$$

Hence with probability one, $X$ takes values in the root of the function $\sum p_i x^{c_i}$. If $\operatorname{supp}(X)$ contains an interval $I$ then $\sum p_i x^{c_i} = 0$ for all $x \in I$. This is only possible if $p_i = 0$ for $i = 1, \ldots n$, because the functions $x^{c_1}, \ldots, x^{c_n}$ are independent for $c_i \neq c_j$. □

## 4.3. The Covariance Structure of $\mathcal{V}_k^{(\alpha)}$

The next step is to investigate the variances and covariances of the volume-power functionals. In the following, we shall see that $1 \leq m \leq \min k_i + 1$. From Lemma 4.1, $\alpha_1 + \alpha_2 > -d + m - 2$ holds true for all values of $m$ once it holds for $m = \min k_i + 1$. And so the assumptions of the following theorem implies that the occuring moments exist.

**Theorem 4.3.** *Let $\alpha_i > -d + k_i - 1$ for $i = 1, 2$, and $\alpha_1 + \alpha_2 > -d + \min k_i - 1$. The covariance is given by*

$$\operatorname{Cov}\big(\mathcal{V}_{k_1}^{(\alpha_1)}, \mathcal{V}_{k_2}^{(\alpha_2)}\big) = \sum_{m=1}^{\min k_i+1} \frac{\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}}{m! \prod (k_i - m + 1)!} t^{\sum k_i - m + 2} \delta_t^{\sum (d+\alpha_i)k_i - d(m-1)} (1 + o(1)).$$

*In particular, for $\alpha > \frac{1}{2}(-d + k - 1)$ we have*

$$\mathbb{V}\, \mathcal{V}_k^{(\alpha)} = \sum_{m=1}^{k+1} \frac{\mu_{k,k:m}^{(\alpha,\alpha)}}{m!((k-m+1)!)^2} t^{2k - m + 2} \delta_t^{2(d+\alpha)k - d(m-1)} (1 + o(1)).$$

*Proof.* Without loss of generality, we assume $k_1 \leq k_2$. By definition,

$$\mathcal{V}_{k_1}^{(\alpha_1)} \mathcal{V}_{k_2}^{(\alpha_2)} = \frac{1}{\prod\limits_{i=1,2} (k_i + 1)!} \sum_{\substack{(x_0,\ldots,x_{k_1}) \in \eta_{t,\neq}^{k_1+1} \\ (x'_0,\ldots,x'_{k_2}) \in \eta_{t,\neq}^{k_2+1}}} \Delta_{\delta_t}[\{x_l\}_{l=0}^{k_1}]^{\alpha_1} \Delta_{\delta_t}[\{x'_l\}_{l=0}^{k_2}]^{\alpha_2}$$

Here, $m$ points of the $k_1$-tuple and $k_2$-tuple coincide, $m = 0, \dots k_1 + 1$. We assume that $x_{k_1-m+1} = x'_{k_1-m+1}, \dots, x_{k_1} = x'_{k_1}$, multiply by $\binom{k_1+1}{m}\frac{(k_2+1)!}{(k_2-m+1)!}$, and rename the variables $(x'_0, \dots, x'_{k_1-m})$ by $(x_{k_2+1}, \dots, x_{k_1+k_2-m+1})$. We note here that $\binom{k_1+1}{m}\frac{(k_2+1)!}{(k_2-m+1)!}$ represents the number of copies of a particular $m$-coincidence. This yields

$$\mathcal{V}_{k_1}^{(\alpha_1)}\mathcal{V}_{k_2}^{(\alpha_2)} = \sum_{m=0}^{k_1+1} \frac{1}{m! \prod\limits_{i=1,2}(k_i - m + 1)!} \sum_{(x_0,\dots,x_{k_1+k_2-m+1}) \in \eta_{t,\neq}^{k_1+k_2-m+2}} \Delta_{\delta_t}[\{x_l\}_{l=0}^{k_1}]^{\alpha_1} \Delta_{\delta_t}[\{x_l\}_{l=k_1-m+1}^{k_1+k_2-m+1}]^{\alpha_2}$$

and applying the Mecke formula gives

$$\mathbb{E}\mathcal{V}_{k_1}^{(\alpha_1)}\mathcal{V}_{k_2}^{(\alpha_2)} = \sum_{m=0}^{k_1+1} \frac{t^{k_1+k_2-m+2}}{m! \prod\limits_{i=1,2}(k_i - m + 1)!} \int\limits_{W^{k_1+k_2-m+2}} \Delta_{\delta_t}[\{x_l\}_{l=0}^{k_1}]^{\alpha_1} \Delta_{\delta_t}[\{x_l\}_{l=k_1-m+1}^{k_1+k_2-m+1}]^{\alpha_2} \, \mathrm{d}x_0 \cdots \mathrm{d}x_{k_1+k_2-m+1}.$$

The first term of this sum with $m = 0$ is

$$\frac{t^{k_1+k_2+2}}{(k_1+1)!(k_1+1)!} \int\limits_{W^{k_1+k_2+2}} \Delta_{\delta_t}[\{x_l\}_{l=0}^{k_1}]^{\alpha_1} \Delta_{\delta_t}[\{x_l\}_{l=k_1+1}^{k_1+k_2+1}]^{\alpha_2} \mathrm{d}x_0 \cdots \mathrm{d}x_{k_1+k_2+1}$$

and it is precisely equal to $\mathbb{E}\mathcal{V}_{k_1}^{(\alpha)}\mathbb{E}\mathcal{V}_{k_2}^{(\alpha)}$, and thus the covariance is given by the summands from $m = 1$ to $m = k_1 + 1$. To obtain the asymptotic behavior of the covariance we follow the same approach as in the proof of Theorem 3.6. We substitute $x_i = \delta_t \tilde{x}_i + x_{k_1}$ for $x_i$, $i \neq k_1$ to get

$$\mathbb{C}\mathrm{ov}\big(\mathcal{V}_{k_1}^{(\alpha_1)}, \mathcal{V}_{k_2}^{(\alpha_2)}\big) = \sum_{m=1}^{k_1+1} \frac{t^{k_1+k_2-m+2}\delta_t^{d(k_1+k_2-m+1)+\alpha_1 k_1+\alpha_2 k_2}}{m! \prod\limits_{i=1,2}(k_i - m + 1)!} \int\limits_{W} \int\limits_{(\delta_t^{-1}(W - x_{k_1}) \cap B^d)^{k_1+k_2+1-m}} \Delta_1[0, \{\tilde{x}_l\}_{l=0}^{k_1-1}]^{\alpha_1} \Delta_1[0, \{\tilde{x}_l\}_{\substack{l=k_1-m+1 \\ l \neq k_1}}^{k_1+k_2-m+1}]^{\alpha_2} \, \mathrm{d}\tilde{x}_0 \cdots \mathrm{d}\tilde{x}_{k_1-1} \, \mathrm{d}\tilde{x}_{k_1+1} \cdots \mathrm{d}\tilde{x}_{k_1+k_2-m+1} \, dx_{k_1}.$$

For an upper bound we obtain

$$\mathbb{C}\mathrm{ov}\big(\mathcal{V}_{k_1}^{(\alpha_1)}, \mathcal{V}_{k_2}^{(\alpha_2)}\big) \leq \sum_{m=1}^{k_1+1} \frac{t^{k_1+k_2-m+2}\delta_t^{d(k_1+k_2-m+1)+\alpha_1 k_1+\alpha_2 k_2}}{m! \prod\limits_{i=1,2}(k_i - m + 1)!} \mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}.$$

Once again, for the lower bound we consider $x_{k_1} \in W_{-\delta_t}$ and use $B^d \subset \delta_t^{-1}(W - x_{k-1})$, which yields

$$\mathbb{C}\mathrm{ov}\big(\mathcal{V}_{k_1}^{(\alpha_1)}, \mathcal{V}_{k_2}^{(\alpha_2)}\big) \geq \sum_{m=1}^{k_1+1} \frac{t^{k_1+k_2-m+2}\delta_t^{d(k_1+k_2-m+1)+\alpha_1 k_1+\alpha_2 k_2}}{m! \prod\limits_{i=1,2}(k_i - m + 1)!} V(W_{-\delta_t})\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}$$

and using the estimate

$$V(W_{-\delta_t}) \geq 1 - S(W)\delta_t$$

proves Theorem 4.3.

The variance is implied with $k_1 = k_2 = k$ and $\alpha_1 = \alpha_2 = \alpha$. □

# 4.4. The Normalised $\mathcal{V}_k^{(\alpha)}$

To be able to distinguish the behavior of the covariance in the different asymptotic regimes already introduced, we consider the normalised volume-power functional.

Define

$$Q_i = t^{\frac{1}{2}}\delta_t^{\alpha_i k_i} \max_{1\leq m\leq k_i+1}\{(t\delta_t^d)^{k_i - \frac{m-1}{2}}\}. \tag{4.5}$$

We note that

$$\{(t\delta_t^d)^{k_i-\frac{m-1}{2}}\}_{1\leq m\leq k_i+1} = \{(t\delta_t^d)^{\frac{2k_i}{2}}, (t\delta_t^d)^{\frac{2k_i-1}{2}}, \dots, (t\delta_t^d)^{\frac{k_i}{2}}\}$$

and thus the maximum is always attained either for the first or last term, that is, for $m = 1$ or for $m = k_i + 1$. So that we can write

$$Q_i = t^{\frac{1}{2}}\delta_t^{\alpha_i k_i} \max\{(t\delta_t^d)^{k_i}, (t\delta_t^d)^{\frac{1}{2}k_i}\}. \tag{4.6}$$

Also, once $t\delta_t^d > 1$, the maximum is $(t\delta_t^d)^{k_i}$. In the same vein, once $t\delta_t^d < 1$, the maximum is $(t\delta_t^d)^{\frac{1}{2}k_i}$.

The product $Q_1 \times Q_2$ as seen in Theorem 4.3 indeed determines the asymptotic behavior of the covariance functionals.

**Definition 4.4.** For $i \in \mathbb{N}$, let $\alpha_i \in \mathbb{R}$ and $k_i \in \mathbb{N}$. The normalized volume-power functionals are given by

$$\widehat{\mathcal{V}}_{k_i}^{(\alpha_i)} = \frac{\mathcal{V}_{k_i}^{(\alpha_i)}}{Q_i}$$

*Remark* 4.5. We give the expectation of the normalized volume-power functional and discuss asymptotics in the various regimes.

$$\mathbb{E}\widehat{\mathcal{V}}_{k_i}^{(\alpha_i)} = \frac{\mathbb{E}\mathcal{V}_{k_i}^{(\alpha_i)}}{Q_i}$$

(i) In the sparse regime, $\lim_{t\to\infty} t\delta_t^d = 0$ and $Q_i = t^{\frac{1}{2}}\delta_t^{\alpha_i k_i}(t\delta_t^d)^{\frac{1}{2}k_i}$, so applying Theorem 3.6, we have

$$\mathbb{E}\widehat{\mathcal{V}}_{k_i}^{(\alpha_i)} = \frac{\frac{\mu_{k_i}^{(\alpha_i)}}{(k_i+1)!}t^{k_i+1}\delta_t^{k_i(\alpha_i+d)}(1+O(\delta_t))}{t^{\frac{1}{2}}\delta_t^{\alpha_i k_i}(t\delta_t^d)^{\frac{1}{2}k_i}} = \frac{\mu_{k_i}^{(\alpha_i)}}{(k_i+1)!}\left[t(t\delta_t^d)^{k_i}\right]^{\frac{1}{2}}(1+O(\delta_t)).$$

The asymptotic expectation here depends on the value of $\delta_t$. For example, if $\delta_t = t^{-\frac{1}{d}}e^{-\frac{t}{dk_i}}$, then $(t\delta_t^d)^{k_i} = e^{-t} \to 0$ slower than $t\to\infty$.

(ii) In the dense regime, $\lim_{t\to\infty} t\delta_t^d = \infty$ and $Q_i = t^{\frac{1}{2}}\delta_t^{\alpha_i k_i}(t\delta_t^d)^{k_i}$, so we obtain

$$\mathbb{E}\widehat{\mathcal{V}}_{k_i}^{(\alpha_i)} = \frac{\frac{\mu_{k_i}^{(\alpha_i)}}{(k_i+1)!}t^{k_i+1}\delta_t^{k_i(\alpha_i+d)}(1+O(\delta_t))}{t^{\frac{1}{2}}\delta_t^{\alpha_i k_i}(t\delta_t^d)^{k_i}} = \frac{\mu_{k_i}^{(\alpha_i)}}{(k_i+1)!}t^{\frac{1}{2}}(1+O(\delta_t)).$$

So in this case,

$$\mathbb{E}\widehat{\mathcal{V}}_{k_i}^{(\alpha_i)} \to \infty$$

in the limit.

**Theorem 4.6.** *Let $\alpha_i > -d+k_i-1$ for $i=1,2$, and $\alpha_1+\alpha_2 > -d+\min k_i - 1$.*

*(i) In the sparse regime, where $\lim_{t\to\infty} t\delta_t^d = 0$, we have $\lim_{t\to\infty}\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}_{k_1}^{(\alpha_1)},\widehat{\mathcal{V}}_{k_2}^{(\alpha_2)}\big) = 0$ for $k_1<k_2$, and for $k_1=k_2=k$*

$$\lim_{t\to\infty}\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}_{k}^{(\alpha_1)},\widehat{\mathcal{V}}_{k}^{(\alpha_2)}\big) = \frac{\mu_k^{(\alpha_1+\alpha_2)}}{(k+1)!}.$$

*(ii) In the dense regime, where $\lim_{t\to\infty} t\delta_t^d = \infty$, we have*

$$\lim_{t\to\infty}\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}_{k_1}^{(\alpha_1)},\widehat{\mathcal{V}}_{k_2}^{(\alpha_2)}\big) = \frac{\mu_{k_1}^{(\alpha_1)}}{k_1!}\frac{\mu_{k_2}^{(\alpha_2)}}{k_2!}. \tag{4.7}$$

*(iii) In the thermodynamic regime, where $\lim_{t\to\infty} t\delta_t^d = c\in(0,\infty)$, we have for $k_1\le k_2$*

$$\lim_{t\to\infty}\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}_{k_1}^{(\alpha_1)},\widehat{\mathcal{V}}_{k_2}^{(\alpha_2)}\big) = \begin{cases}\displaystyle\sum_{m=0}^{k_1}\frac{\mu_{k_1,k_2:k_1-m+1}^{(\alpha_1,\alpha_2)}}{(k_1-m+1)!m!(k_2-k_1+m)!}c^{\frac{k_2-k_1}{2}+m}, & c\le 1\\ \displaystyle\sum_{m=0}^{k_1}\frac{\mu_{k_1,k_2:m+1}^{(\alpha_1,\alpha_2)}}{(m+1)!(k_1-m)!(k_2-m)!}c^{-m}, & c\ge 1.\end{cases}$$

*Proof.* Recall that $k_1 \leq k_2$. Because

$$Q_i = t^{\frac{1}{2}} \delta_t^{\alpha_i k_i} \max\{(t\delta_t^d)^{k_i}, (t\delta_t^d)^{\frac{1}{2}k_i}\} \tag{4.8}$$

the behavior in the sparse and dense regimes are immediate.

In the sparse regime $t\delta_t^d \to 0$, and hence as soon as $t\delta_t^d < 1$, the maximum is attained for $m = k_1 + 1$, resp. $m = k_2 + 1$. Thus

$$Q_1 Q_2 = t\, \delta_t^{\alpha_1 k_1 + \alpha_2 k_2} (t\delta_t^d)^{\frac{k_1+k_2}{2}}$$

and

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}_{k_1}^{(\alpha_1)}, \widehat{\mathcal{V}}_{k_2}^{(\alpha_2)}\big) &= \frac{1}{Q_1 Q_2} \mathbb{C}\mathrm{ov}\big(\mathcal{V}_{k_1}^{(\alpha_1)}, \mathcal{V}_{k_2}^{(\alpha_2)}\big) \\
&= \frac{\sum_{m=1}^{k_1+1} \frac{t^{k_1+k_2-m+2} \delta_t^{d(k_1+k_2-m+1)+\alpha_1 k_1 + \alpha_2 k_2}}{m!(k_1-m+1)!(k_2-m+1)!} \mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}}{t\, \delta_t^{\alpha_1 k_1 + \alpha_2 k_2} (t\delta_t^d)^{\frac{k_1+k_2}{2}}} (1 + O(\delta_t)) \\
&= \sum_{m=1}^{k_1+1} \frac{\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}}{m!(k_1 - m + 1)!(k_2 - m + 1)!} (t\delta_t^d)^{\frac{k_1+k_2}{2} - m + 1} (1 + O(\delta_t)) \qquad (4.9) \\
&= \frac{\mu_{k_1,k_2:k_1+1}^{(\alpha_1,\alpha_2)}}{(k_1+1)!(k_2 - (k_1+1) + 1)!} (t\delta_t^d)^{\frac{k_1+k_2}{2} - (k_1+1) + 1} (1 + O(t\delta_t^d))(1 + O(\delta_t)) \\
&= \frac{\mu_{k_1,k_2:k_1+1}^{(\alpha_1,\alpha_2)}}{(k_1+1)!(k_2 - k_1)!} (t\delta_t^d)^{\frac{k_2-k_1}{2}} (1 + O(\delta_t) + O(t\delta_t^d)).
\end{aligned}$$

Hence for $k_1 < k_2$, asymptotically the covariance vanishes, so that

$$\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}_{k_1}^{(\alpha_1)}, \widehat{\mathcal{V}}_{k_2}^{(\alpha_2)}\big) = O((t\delta_t^d)^{\frac{1}{2}}). \tag{4.10}$$

In the case $k_1 = k_2 = k$, the covariance equals asymptotically a moment of the volume. Using (4.3), we obtain

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}_{k}^{(\alpha_1)}, \widehat{\mathcal{V}}_{k}^{(\alpha_2)}\big) &= \frac{\mu_k^{(\alpha_1+\alpha_2)}}{(k+1)!} (1 + O(\delta_t + t\delta_t^d)) \\
&= \mathbb{E}\widehat{\mathcal{V}}_k^{(\alpha_1+\alpha_2)} \Big(t(t\delta_t^d)^k\Big)^{-\frac{1}{2}} (1 + O(\delta_t + t\delta_t^d)). \qquad (4.11) \\
&= \mathbb{E}\widehat{\mathcal{V}}_k^{(\alpha_1+\alpha_2)} (1 + O(\delta_t + t\delta_t^d)). \qquad (4.12)
\end{aligned}$$

In the dense regime $t\delta_t^d \to \infty$ and hence (4.8) shows that the maximum is attained for $m = 1$ as soon as $t\delta_t^d > 1$. Thus $Q_1 Q_2 = t\, \delta_t^{\alpha_1 k_1 + \alpha_2 k_2} (t\delta_t^d)^{k_1+k_2}$ and

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}_{k_1}^{(\alpha_1)}, \widehat{\mathcal{V}}_{k_2}^{(\alpha_2)}\big) &= \frac{1}{Q_1 Q_2} \mathbb{C}\mathrm{ov}\big(\mathcal{V}_{k_1}^{(\alpha_1)}, \mathcal{V}_{k_2}^{(\alpha_2)}\big) \\
&= \frac{\sum_{m=1}^{k_1+1} \frac{t^{k_1+k_2-m+2} \delta_t^{d(k_1+k_2-m+1)+\alpha_1 k_1 + \alpha_2 k_2}}{m!(k_1-m+1)!(k_2-m+1)!} \mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}}{t\, \delta_t^{\alpha_1 k_1 + \alpha_2 k_2} (t\delta_t^d)^{k_1+k_2}} (1 + O(\delta_t))
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{m=1}^{k_1+1} \frac{\mu^{(\alpha_1,\alpha_2)}_{k_1,k_2:m} (t\delta_t^d)^{-m+1}}{m!(k_1-m+1)!(k_2-m+1)!}(1+O(\delta_t)) \qquad (4.13)\\
&= \frac{\mu^{(\alpha_1,\alpha_2)}_{k_1,k_2:1}}{k_1!k_2!}(t\delta_t^d)^{-(1)+1}(1+O((t\delta_t^d)^{-1}))(1+O(\delta_t))\\
&= \frac{\mu^{(\alpha_1,\alpha_2)}_{k_1,k_2:1}}{k_1!k_2!}(1+O(\delta_t+(t\delta_t^d)^{-1}))\\
&= \frac{\mu^{(\alpha_1)}_{k_1}\mu^{(\alpha_2)}_{k_2}}{k_1!k_2!}(1+O(\delta_t+(t\delta_t^d)^{-1})).
\end{aligned}
$$

Note that in this case the limiting covariance is the product of the suitable normalized expectations,

$$
\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}^{(\alpha_1)}_{k_1}, \widehat{\mathcal{V}}^{(\alpha_2)}_{k_2}\big) = (k_1+1)(k_2+1)t^{-1}\mathbb{E}\widehat{\mathcal{V}}^{(\alpha_1)}_{k_1}\mathbb{E}\widehat{\mathcal{V}}^{(\alpha_2)}_{k_2}(1+o(1)).
$$

In the thermodynamic regime, $t\delta_t^d$ tends to a constant $c \in \mathbb{R}$, hence all terms in the sum occurring in the covariance contribute in the same way. We differentiate between when $c < 1$ and when $c > 1$. Accordingly to the sparse regime we obtain for $c < 1$

$$
Q_1Q_2 = t\,\delta_t^{\sum \alpha_i k_i}(t\delta_t^d)^{\sum \frac{k_i}{2}}
$$

for $t$ sufficiently large, and by (4.9)

$$
\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}^{(\alpha_1)}_{k_1}, \widehat{\mathcal{V}}^{(\alpha_2)}_{k_2}\big) = \sum_{m=1}^{k_1+1} \frac{\mu^{(\alpha_1,\alpha_2)}_{k_1,k_2:m}}{m!(k_1-m+1)!(k_2-m+1)!}c^{\frac{k_1+k_2}{2}-m+1}(1+o(1)).
$$

And rewriting $m$ as $k_1 - m + 1$ gives the statement in the theorem.
Analogously, for $c \geq 1$ we obtain by (4.13)

$$
\mathbb{C}\mathrm{ov}\big(\widehat{\mathcal{V}}^{(\alpha_1)}_{k_1}, \widehat{\mathcal{V}}^{(\alpha_2)}_{k_2}\big) = \sum_{m=1}^{k_1+1} \frac{\mu^{(\alpha_1,\alpha_2)}_{k_1,k_2:m}}{m!(k_1-m+1)!(k_2-m+1)!}c^{-m+1}(1+o(1)).
$$

And rewriting $m$ as $m+1$ gives the statement in the theorem.
In both cases we see that the error term $o(1)$ is given by

$$
O(\delta_t) + O(c - t\delta_t^d). \tag{4.14}
$$

□

Putting things together we obtain the limiting covariance matrix of the random vector $(\widehat{\mathcal{V}}^{(\alpha_1)}_{k_1}, \ldots, \widehat{\mathcal{V}}^{(\alpha_n)}_{k_n})$. For this we call $(k_1,\alpha_1), \ldots, (k_n,\alpha_n)$ an *admissible sequence* if

(i) $0 \leq k_1 \leq \cdots \leq k_n$,

(ii) the pairs $(k_1,\alpha_1)$, …, $(k_n,\alpha_n)$ are distinct,

(iii) $\alpha_i > -d + k_i - 1$ and $\alpha_i + \alpha_j > -d + \min_{ij} k_l - 1$ for all $i, j \in \{1, \dots, n\}$,

(iv) $\alpha_i = 0$ if $k_i > d$.

First we rewrite the sum occurring in Theorem 4.6 in the case $c < 1$. The reason for this is to leave the power of $c$ without any $k_i$'s. So we rewrite $m$ as $\frac{m-k_2+k_1}{2}$ such that $\frac{m-k_2+k_1}{2} \in \{0, 1, 2, \dots, k_1\}$ or in other words $m - k_2 + k_1 \in \{0, 2, 4, \dots, 2k_1\}$. We obtain the following.

$$\sum_{m=0}^{k_1} \frac{\mu^{(\alpha_1,\alpha_2)}_{k_1,k_2:k_1-m+1}}{(k_1 - m + 1)!m!(k_2 - k_1 + m)!} c^{\frac{k_2-k_1}{2}+m}$$
$$= \sum_{m=0}^{\infty} \mu^{(\alpha_1,\alpha_2)}_{k_1,k_2:\frac{(k_2+k_1-m+2)}{2}} \frac{\mathbb{1}(m - (k_2 - k_1) \in \{0, 2, 4, \dots, 2k_1\})}{(\frac{k_2+k_1-m+2}{2})!(\frac{m-k_2+k_1}{2})!(\frac{m+k_2-k_1}{2})!} c^{\frac{m}{2}} .$$

We define for $m = 0, \dots, 2k_n$ the $(n \times n)$-matrices

$$A^{<1}_m = \left( \mu^{(\alpha_1,\alpha_2)}_{k_l,k_j:\frac{(k_l+k_j-m+2)}{2}} \frac{\mathbb{1}(m - |k_l - k_j| \in \{0, 2, 4, \dots, 2 \min k_i\})}{(\frac{k_l+k_j-m+2}{2})!(\frac{m-k_l+k_j}{2})!(\frac{m+k_l-k_j}{2})!} \right)_{l,j=1,\dots,n} \tag{4.15}$$

with $\min k_i$ short for $\min_{i\in\{j,l\}} k_i$, and for $m = 0, \dots, k_n$

$$A^{>1}_m = \left( \mu^{(\alpha_l,\alpha_j)}_{k_l,k_j:m+1} \frac{\mathbb{1}(m \le \min k_i)}{(m + 1)!(k_l - m)!(k_j - m)!} \right)_{l,j=1,\dots,n} . \tag{4.16}$$

Note that for large $m$ these matrices contain a large number of zeros. E.g., for $k_{n-1} < m \le k_n$ the matrix $A^{>1}_m$ contains only one nonzero entry,

$$(A^{>1}_m)_{nn} = \frac{\mu^{(\alpha_n,\alpha_n)}_{k_n,k_n:m+1}}{(m + 1)!((k_n - m)!)^2}.$$

Recall the normalized volume-power functionals $\widehat{\mathcal{V}}^{(\alpha_i)}_{k_i} = \mathcal{V}^{(\alpha_i)}_{k_i}/Q_i$ defined in 4.4.

**Theorem 4.7.** *Assume that* $(k_1, \alpha_1)$, $\dots$, $(k_n, \alpha_n)$ *is an admissible sequence. The random vector* $(\widehat{\mathcal{V}}^{(\alpha_1)}_{k_1}, \dots, \widehat{\mathcal{V}}^{(\alpha_n)}_{k_n})$ *has the asymptotic covariance matrix*

$$\Sigma =: \begin{cases} A^{<1}_0 & : \lim_{t\to\infty} t\delta^d_t = 0 \\ \sum_{m=0}^{2k_n} A^{<1}_m c^{\frac{m}{2}} & : \lim_{t\to\infty} t\delta^d_t = c \in (0, 1] \\ \sum_{m=0}^{k_n} A^{>1}_m c^{-m} & : \lim_{t\to\infty} t\delta^d_t = c \in [1, \infty) \\ A^{>1}_0 & : \lim_{t\to\infty} t\delta^d_t = \infty \end{cases} \tag{4.17}$$

Clearly, in the case $c = 1$ the identity $\sum A^{<1}_m = \sum A^{>1}_m$ is satisfied which follows from the definitions (4.15) and (4.16).

By Theorem 4.6, (4.7), the matrix $A_0^{>1}$ takes the form of a tensor product.

$$A_0^{>1} = \begin{pmatrix} \vdots \\ \frac{\mu_{k_i}^{(\alpha_i)}}{k_i!} \\ \vdots \end{pmatrix} \otimes \begin{pmatrix} \vdots \\ \frac{\mu_{k_i}^{(\alpha_i)}}{k_i!} \\ \vdots \end{pmatrix}$$

Hence, in the dense case the covariance matrix $\Sigma$ is of rank 1, and thus is singular in this regime.

Also, the covariance matrix $A_0^{<1}$ takes a particular nice form. Using (4.11) we see that

$$A_0^{<1} = \left( \mu_{k_j}^{(\alpha_l+\alpha_j)} \frac{\mathbb{1}(k_l = k_j)}{(k_j+1)!} \right)_{l,j=1,\dots,n}$$

is a diagonal block matrix. A block is of size $i$ if $k_m = \cdots = k_{m+i-1}$, and then is a constant times the matrix

$$\left( \mathbb{E}\Delta_1^{(\alpha_l+\alpha_j)} \right)_{l,j=m,\dots,m+i-1}$$

with $\Delta_1 = \Delta_1[0, \{X_l\}_{l=1}^{k_m}]$. Thus each block is a generalized moment matrix, and we know by Theorem 4.2 that this is of full rank if $\alpha_l \neq \alpha_j$ for $l \neq j$. Since all $\alpha_i$ are distinct, $A_0^{<1}$ is of full rank.

Further, on $c \in [0,1]$ the determinant $|\Sigma|$ of the covariance matrix $\Sigma$ is a polynomial in $c$ with $\lim_{c\to 0} |\Sigma| \to |A_0^{<1}| > 0$ and thus this polynomial is not trivial. Hence it has at most finitely many zeros. Analogously, for $c \in [1,\infty)$ the determinant of $\Sigma$ is a polynomial in $c$. Because $\widehat{\mathcal{V}}_{k_i}^{(\alpha_i)}$ on $c \geq 1$ is just a renormalized version of $\widehat{\mathcal{V}}_{k_i}^{(\alpha_i)}$ on $c \leq 1$ the polynomial is not trivial. (In the limit the renormalizations are just multiplications by $c^{\frac{k_i}{2}}$.) Hence there are again only finitely many zeros of this polynomial. We summarize our findings.

**Corollary 4.8.** *Assume that $(k_1, \alpha_1)$, ..., $(k_n, \alpha_n)$ is an admissible sequence. Then the rank of $\Sigma$ equals $n$ in the sparse regime. In the thermodynamic regime $\Sigma$ is of rank $n$ except for finitely many values of $c$. In the dense regime $\Sigma$ is of rank one.*

# 5. Central Limit Theorems

In this chapter, the rate of convergence of the volume-power functionals to a Guassian random variable and a Guassian random vector in the univariate and the multivariate cases respectively is presented, with respect to the Vietoris-Rips complex. The existence of the central limit theorem is also investigated for the different asymptotic regimes. Finally, the Cech complex counter parts of previous results is stated.

A Poisson U-statistic is *absolutely convergent* if $F = \sum_{\eta^k_{t,\neq}} |f(x_1, \ldots, x_k)|$ is in $L^2(\mathbb{P})$. Note that $\mathcal{V}_k^{(\alpha)}$ is an absolutely convergent U-statistic since all occurring functions are bounded and vanish outside the compact convex set $W$.

Let $F^{(1)}, \ldots, F^{(n)}$ be absolutely convergent Poisson U-statistics of order $k_1, \ldots, k_n$ respectively,

$$F^{(l)} = \sum_{(x_1^{(l)}, \ldots, x_{k_l}^{(l)}) \in \eta^{k_l}_{t,\neq}} f^{(l)}(x_1^{(l)}, \ldots, x_{k_l}^{(l)})$$

for $l = 1, \ldots, n$. It will be essential to define suitable partitions on the set of variables $\{x_1^{(l)}, \ldots, x_{k_l}^{(l)}\}$, $l = 1, \ldots, n$, of $f^{(l)} : W^{k_l} \to \bar{\mathbb{R}}$.

## 5.1. The Fourth Moment Integrals

We recall from Section 2.4 the set of variables,

$$V(k_1, \ldots, k_n) = \left\{x_1^{(1)}, \ldots, x_{k_1}^{(1)}, x_1^{(2)}, \ldots, x_{k_{n-1}}^{(n-1)}, x_1^{(n)}, \ldots, x_{k_n}^{(n)}\right\},$$

and

$$\tilde{\Pi}(k_1, \ldots, k_n) = \left\{\sigma \in \mathcal{P}(V(k_1, \ldots, k_n)) : \sigma \wedge \bar{\pi} = \hat{0},\, \sigma \vee \bar{\pi} = \hat{1}\right\},$$

where $\mathcal{P}(V(k_1, \ldots, k_n))$ the set of partitions of $V(k_1, \ldots, k_n)$. Clearly, each block of $\sigma \in \tilde{\Pi}(k_1, \ldots, k_n)$ has at most $n$ variables with different upper index $l$. In this section, we shall be restricted to the case $n = 4$. So we have

$$V(k_1, \ldots, k_4) = \left\{x_1^{(1)}, \ldots, x_{k_1}^{(1)}, x_1^{(2)}, \ldots, x_{k_2}^{(2)}, x_1^{(3)}, \ldots, x_{k_3}^{(3)}, x_1^{(4)}, \ldots, x_{k_4}^{(4)}\right\},$$

which consists of four sets of variables, and with $\bar{\pi} \in \mathcal{P}(V(k_1, \ldots, k_4))$, the partition whose blocks are the fundamental building blocks $\{x_1^{(l)}, \ldots, x_{k_l}^{(l)}\}$, $l = 1, \ldots, 4$, we have

$$\tilde{\Pi}(k_1, \ldots, k_4) = \left\{\sigma \in \mathcal{P}(V(k_1, \ldots, k_4))\colon \sigma \wedge \bar{\pi} = \hat{0},\, \sigma \vee \bar{\pi} = \hat{1}\right\}$$

as the set of all partitions such that each block contains at most one element from each of the building blocks $\{x_1^{(l)}, \ldots, x_{k_l}^{(l)}\}$, $l = 1, \ldots, 4$, and all four fundamental blocks are connected. Clearly, for $\sigma \in \tilde{\Pi}(k_1, \ldots, k_4)$ it may happen that some variables are singletons, and we define $s(\sigma) = (s_1, \ldots, s_4)$ to be the vector consisting of the number of singletons in each of the building blocks.

As defined previously, we will need the notion of a 4-fold tensor product, $\otimes_{l=1}^4 f^{(l)} : W^{\sum_{l=1}^4 k_l} \to \mathbb{R}$, of functions $f^{(l)}$, given by

$$\left(\otimes_{l=1}^4 f^{(l)}\right)(x_1^{(1)}, \ldots, x_{k_4}^{(4)}) = \prod_{l=1}^4 f^{(l)}(x_1^{(l)}, \ldots, x_{k_l}^{(l)}).$$

For a partition, $\sigma \in \Pi(k_1, \ldots, k_4)$, we construct a new function, $(\otimes_{l=1}^4 f^{(l)})_\sigma : W^{|\sigma|} \to \bar{\mathbb{R}}$, by replacing all variables that belong to the same block of $\sigma$ by a new common variable. We give an example below.

Let $k_1 = k_2 = k_3 = k_4 = 2$. Let $F^{(l)} = \sum_{(x_1^{(l)}, x_2^{(l)}) \in \eta^2_{t,\neq}} f^{(l)}(x_1^{(l)}, x_2^{(l)})$ for $l = 1, 2, 3, 4$. Then

$$V(2,2,2,2) = \{x_1^{(1)}, x_2^{(1)}, x_1^{(2)}, x_2^{(2)}, x_1^{(3)}, x_2^{(3)}, x_1^{(4)}, x_2^{(4)}\}.$$

Let $\sigma = \left\{\{x_1^{(1)}, x_1^{(2)}, x_1^{(3)}, x_1^{(4)}\}, \{x_2^{(1)}, x_2^{(3)}\}, \{x_2^{(2)}, x_2^{(4)}\}\right\} \in \Pi(k_1, \ldots, k_4)$. Since there are no singletons in any of the variables, $s(\sigma) = (0, 0, 0, 0)$. Also, $|\sigma| = 3$ and $(\otimes_{l=1}^4 f^{(l)})_\sigma : W^3 \to \bar{\mathbb{R}}$ is given by

$$(\otimes_{l=1}^4 f^{(l)})_\sigma(y_1, y_2, y_3) = f^{(1)}(y_1, y_2) f^{(2)}(y_1, y_3) f^{(3)}(y_1, y_2) f^{(4)}(y_1, y_3).$$

Finally we are able to introduce the functions $M_{ij}$ defined in [RS13] for the univariate case and in [Sch13] for the multivariate case. The functions are given by

$$M_{ij}(f^{(l)}, f^{(m)}) = \sum_{\substack{\sigma \in \tilde{\Pi}(k_l, k_l, k_m, k_m) \\ s(\sigma) = (k_l - i, k_l - i, k_m - j, k_m - j)}} \int_{W^{|\sigma|}} |(f^{(l)} \otimes f^{(l)} \otimes f^{(m)} \otimes f^{(m)})_\sigma| \, \mathrm{d}\mu^{|\sigma|}$$

where in our case $d\mu$ is the intensity measure $tdx$. Apart from the precise definition given above, the main point is that the functions $M_{ij}$ is something like a mixed fourth moment of $f^{(l)}$ and $f^{(m)}$ where all functions are linked via the common use of some of the variables. These functions $M_{ij}$ are the main ingredients in the central limit theorems considered in the next sections.

*Remark* 5.1. By Fubinis theorem one can integrate first the functions $f^{(l)}$ over the $k_l - i$ free variables, i.e. singletons, which produces reduced functions $f_i^{(l)}$, and analogously the functions $f^{(m)}$ over the $k_m - j$ free variables producing $f_j^{(m)}$. In this form the result was stated in [Sch13].

# 5.2. Univariate Central Limit Theorem

The Wasserstein distance, $d_W$ between a Poisson functional and a standard Guassian random variable was given in [PSTU10] using Stein's method and Malliavin calculus. The following theorem states the result.

**Theorem 5.2.** *Let $F \in domD$ such that $\mathbb{E}F = 0$, and let $N$ be a standard Guassian random variable. Then*

$$\begin{aligned} d_W(F,N) &\leq \mathbb{E}|1 - \langle DF, -DL^{-1}F\rangle_{L^2(\mu)}| + \int_W \mathbb{E}\left[|D_zF|^2|D_zL^{-1}F|\right] \,\mathrm{d}\mu(z) \\ &\leq \sqrt{\mathbb{E}(1 - \langle DF, -DL^{-1}F\rangle_{L^2(\mu)})^2} + \int_W \mathbb{E}\left[|D_zF|^2|D_zL^{-1}F|\right] \,\mathrm{d}\mu(z) \end{aligned}$$

This together with Wiener-Itô chaos expansion of these Malliavin operators was used in [RS13] to approximate this distance in terms of the fourth moment integrals $M_{ij}$ discussed in the previous section.

**Theorem 5.3.** *Let $F = F(\eta_t) \in L^2(\mathbb{P})$ be an absolutely convergent U-statistic of order $k$ with $\mathbb{V}F > 0$, and $N$ be a standard Gaussian random variable. Then*

$$d_W\left(\frac{F - \mathbb{E}F}{\sqrt{\mathbb{V}\,F}}, N\right) \leq 2k^{\frac{7}{2}} \sum_{1 \leq i \leq j \leq k} \frac{\sqrt{M_{ij}(f,f)}}{\mathbb{V}F} .$$

The univariate central limit theorem also uses the Kolmogorov distance $d_K$ of random variables. Theorem 5.3 was extended to the Kolmogorov distance in [Sch16, Theorem 4.2].

**Theorem 5.4.** *Let $F = F(\eta_t) \in L^2(\mathbb{P})$ be an absolutely convergent U-statistic of order $k$ with $\mathbb{V}F > 0$, and $N$ be a standard Gaussian random variable. Then*

$$d_K\left(\frac{F - \mathbb{E}F}{\sqrt{\mathbb{V}\,F}}, N\right) \leq 19k^5 \sum_{i,j=1}^{k} \frac{\sqrt{M_{ij}(f,f)}}{\mathbb{V}F} .$$

To get a central limit theorem in these distances, we only need to show that the fourth moment integrals, $M_{ij}$ tend to 0. We now apply the above to $\mathcal{V}_k^{(\alpha)}$ as an absolutely convergent U-statistic.

In the following theorem we assume that $4\alpha > -d + k - 1$ for $0 \leq k \leq d$ and $\alpha = 0$ for $k > d$.

**Theorem 5.5.** *Let $N$ be a standard Gaussian random variable. Then for $d_\star = d_W$ or $d_\star = d_K$ there is a constant $c_{k,\alpha}$ such that*

$$d_\star\left(\frac{\mathcal{V}_k^\alpha - \mathbb{E}\mathcal{V}_k^\alpha}{\sqrt{\mathbb{V}\,\mathcal{V}_k^\alpha}}, N\right) \leq c_{k,\alpha} t^{-\frac{1}{2}} \max\{(t\delta_t^d)^{-\frac{k}{2}}, 1\}. \tag{5.1}$$

*Remark* 5.6. Note that it was to be expected that a central limit theorem only holds if $\mathbb{E}f_k \to \infty$ which happens if $t(t\delta_t^d)^k \to \infty$. It turns out that this is precisely the requirement in Theorem 5.5.

In the case $\alpha = 0$, Theorem 5.5 just gives a univariate central limit theorem for the number of facets. For the Kolmogorov distance this is already well known due to work by Penrose [Pen03] although the central limit theorems there come without error term. In a recent paper by Lachiéze-Rey, Schulte and Yukich [LRSY19] the error terms for the thermodynamic regime and the dense regime have been obtained as a consequence of a much more general theorem for stabilizing functionals. For the Wasserstein distance a central limit theorem with error bounds is due to Decreusefond et al. [DFRV14].

*Proof of Theorem 5.5.* We apply Theorem 5.3. $\mathcal{V}_k^{(\alpha)}$ is an absolutely convergent Poisson U-statistic of order $k+1$ with

$$f(x_0, \ldots, x_k) = \frac{1}{(k+1)!}\Delta_{\delta_t}[\{x_l\}_{l=0}^k]^\alpha.$$

We have to show that the functionals $M_{ij}$ tend to zero. We take a closer look at the summands in $M_{ij}$ in this case, that is,

$$\int\limits_{W^{|\sigma|}} |(f \otimes f \otimes f \otimes f)_\sigma| \, \mathrm{d}\mu^{|\sigma|}$$

They are of the form

$$\frac{t^{|\sigma|}}{(k+1)!^4} \int\limits_{W^{|\sigma|}} (\Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha)_\sigma \, \mathrm{d}x_0 \ldots \mathrm{d}x_{|\sigma|-1}$$

where the functionals are positive and depend on simplices of dimension $k$. The essential feature in the definition of $\sigma$ is that all four functionals $\Delta_{\delta_t}[\cdot]$ are linked by common variables, and each of these functionals depends on $k+1$ variables. First, for the number $|\sigma|$ of variables, by using (2.8) this implies

$$k + 1 \leq |\sigma| \leq 4k + 1, \tag{5.2}$$

where $4k+1 = 4(k+1)-3$ is using the fact that if all are connected by at least one variable, then we have 3 possible blocks less.

Second, assuming without loss of generality, that $x_0$ occurs in the first functional, e.g. $\Delta_{\delta_t}[\cdot] = \Delta_{\delta_t}[\{\cdot\}_0^k]$, all other variables in this first function are at most at distance $\delta_t$, in the functional directly linked to the first one by at most $2\delta_t$, etc. Thus

$$\max \|x_i - x_0\| \leq 4\delta_t$$

if the integrand is not vanishing, and we write this in the form $\Delta_{4\delta_t}[\{x_l\}_{l=0}^{|\sigma|-1}]^0$. We use this and apply Hölder's inequality, which gives the bound

$$\begin{aligned}
&\int\limits_{W^{|\sigma|}} (\Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha)_\sigma \, \mathrm{d}x_0 \ldots \mathrm{d}x_{|\sigma|-1} \\
&\leq \int\limits_{W^{|\sigma|}} \Delta_{\delta_t}[\{x_l\}_{l=0}^k]^{4\alpha} \Delta_{4\delta_t}[\{x_l\}_{l=0}^{|\sigma|-1}]^0 \, \mathrm{d}x_0 \ldots \mathrm{d}x_{|\sigma|-1}
\end{aligned}$$

Now substituting $x_i = \delta_t \tilde{x}_i + x_0$ for $i \geq 1$, changing the order of integration and integrating over $x_0$ while noting that $\mathrm{Vol}(W) = 1$ gives

$$\begin{aligned}
&\int\limits_{W^{|\sigma|}} (\Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha \otimes \Delta_{\delta_t}[\cdot]^\alpha)_\sigma \, \mathrm{d}x_0 \ldots \mathrm{d}x_{|\sigma|-1} \\
&\leq \delta_t^{4k\alpha + d(|\sigma|-1)} \int\limits_W \int\limits_{(\delta_t^{-1}(W - x_0) \cap 4B^d)^{|\sigma|-1}} \Delta_1[0, \{\tilde{x}_l\}_{l=1}^k]^{4\alpha} \, \mathrm{d}\tilde{x}_1 \ldots \mathrm{d}\tilde{x}_{|\sigma|-1} \, \mathrm{d}x_0 \\
&\leq \delta_t^{4k\alpha + d(|\sigma|-1)} \int\limits_{(4B^d)^{|\sigma|-1}} \Delta_1[0, \{\tilde{x}_l\}_{l=1}^k]^{4\alpha} \, \mathrm{d}\tilde{x}_1 \ldots \mathrm{d}\tilde{x}_{|\sigma|-1} \\
&= \delta_t^{4k\alpha + d(|\sigma|-1)} \int\limits_{(4B^d)^{|\sigma|-k-1}} \int\limits_{(4B^d)^k} \Delta_1[0, \{\tilde{x}_l\}_{l=1}^k]^{4\alpha} \, \mathrm{d}\tilde{x}_1 \ldots \mathrm{d}\tilde{x}_{|\sigma|-1} \\
&= \delta_t^{4k\alpha + d(|\sigma|-1)} (4\kappa_d)^{|\sigma|-k-1} \mu_k^{(4\alpha)}.
\end{aligned}$$

By Lemma 3.5, this is finite for $4\alpha > -d + k - 1$. By (5.2) and the definition of $M_{ij}$, this implies

$$\begin{aligned}
M_{ij}(\Delta_{\delta_t}[\cdot]^\alpha, \Delta_{\delta_t}[\cdot]^\alpha) &= \sum_\sigma \frac{t^{|\sigma|}}{(k+1)!^4} \delta_t^{4k\alpha + d(|\sigma|-1)} (4\kappa_d)^{|\sigma|-k-1} \mu_k^{(4\alpha)} \\
&= \sum_\sigma \frac{t (t\delta_t^d)^{|\sigma|-1}}{(k+1)!^4} \delta_t^{4k\alpha} (4\kappa_d)^{|\sigma|-k-1} \mu_k^{(4\alpha)} \\
&\leq c_2 t \, \delta_t^{4k\alpha} \max\{(t\delta_t^d)^k, (t\delta_t^d)^{4k}\}
\end{aligned}$$

with some $c_2 > 0$ depending on $\alpha$ and $k$. Next we use the variance asymptotics from Theorem 4.3. Recall that

$$\mathbb{V}\,\mathcal{V}_k^{(\alpha)} = \sum_{m=1}^{k+1} \frac{\mu_{k,k:m}^{(\alpha,\alpha)}}{m!((k-m+1)!)^2} t^{2k-m+2} \delta_t^{d(2k-m+1)} \delta_t^{2\alpha k}(1+o(1)).$$

Since the maximum term is always attained either at $m = 1$ or $m = k+1$, they imply

$$\mathbb{V}\,\mathcal{V}_k^{(\alpha)} \quad \geq \quad c_3 t \delta_t^{2\alpha k} \max\{(t\delta_t^d)^k, (t\delta_t^d)^{2k}\}$$

with some $c_3 > 0$ for $\delta_t$ sufficiently small. This shows

$$\begin{aligned} \frac{\sqrt{M_{ij}(\Delta_{\delta_t}[\cdot]^\alpha, \Delta_{\delta_t}[\cdot]^\alpha)}}{\mathbb{V}\,\mathcal{V}_k^{(\alpha)}} &\leq \frac{\sqrt{c_2} t^{\frac{1}{2}}\, \delta_t^{2k\alpha} \max\{(t\delta_t^d)^{\frac{k}{2}}, (t\delta_t^d)^{2k}\}}{c_3 t \delta_t^{2\alpha k} \max\{(t\delta_t^d)^k, (t\delta_t^d)^{2k}\}} \\ &= c_4 t^{-\frac{1}{2}} \max\{(t\delta_t^d)^{-\frac{k}{2}}, 1\} \end{aligned}$$

where $c_4$ depends on $k$ and $\alpha$. Summing over all $M_{ij}$ in both distances $d_W$ and $d_K$ gives the desired result. $\square$

*Remark* 5.7. Next, we explore in which regimes we have a central limit theorem. This is obtained when the RHS of (5.1) goes to 0 in the limit. Clearly, this happens when $t\delta_t^d \in (1, \infty]$.

In the dense regime, since $t\delta_t^d \xrightarrow{t\to\infty} \infty$, the RHS of (5.1) gives $c_{k,\alpha} t^{-\frac{1}{2}}$ which goes to 0 in the limit. In this case, the central limit theorem holds with rate of convergence $t^{-\frac{1}{2}}$.

In the thermodynamic regime, $t\delta_t^d \xrightarrow{t\to\infty} c \in (0, \infty)$, we have to differentiate between two cases. For $c \in [1, \infty)$, we have the same results as in the dense regime. But for $c \in (0, 1)$, the RHS of (5.1) gives $c_{k,\alpha}(t(t\delta_t^d)^k)^{-\frac{1}{2}}$ and the central limit theorem does not hold if $c$ is close to 0 enough.

In the sparse regime, it's clear that the central limit theorem does not hold except $t(t\delta_t^d)^k = \mathbb{E} f_k \to \infty$, in the case for example where $\delta_t = t^{-\frac{1}{d}} e^{-\frac{t}{dk}}$, which makes $(t\delta_t^d)^k)$ go to 0 slower than $\delta_t$ goes to $\infty$.

## 5.3. Multivariate Central Limit Theorem

In this section, for a random vector of volume-power functionals, we seek a multivariate central limit theorem. The multivariate central limit theorem makes use of the $d_3$-distance, already introduced earlier, which is obtained by taking $\mathcal{C}_1^3$ to be the set of functions $g : \mathbb{R}^n \to \mathbb{R}$ that are three times differentiable and all partial derivatives of order 2 and 3 are bounded by 1.

It was first proven in [PZ10, Theorem 4.2] for a vector of Poisson functionals, by making use of the Malliavin calculus and an interpolation technique. It was then proven for a vector of absolutely convergent Poisson U-statistics in [Sch16, Theorem 6.3]. The later result is presented below.

**Theorem 5.8.** *Let $\boldsymbol{F} = (F^{(1)}, \dots, F^{(n)})$ be a vector of absolutely convergent Poisson U-statistics of orders $k_1, \dots, k_n$,*

$$F^{(l)} = \sum_{(x_1,\dots,x_{k_l}) \in \eta_{t,\neq}^{k_l}} f^{(l)}(x_1, \dots, x_{k_l}).$$

*And let $\boldsymbol{N}(\Sigma)$ be an $n$-dimensional centered Gaussian random vector with a positive semidefinite covariance matrix $\Sigma$. Then*

$$d_3\Big(\boldsymbol{F} - \mathbb{E}\boldsymbol{F}, \boldsymbol{N}(\Sigma)\Big) \leq \frac{1}{2} \sum_{l,m=1}^{n} |\sigma_{lm} - \mathbb{C}\mathrm{ov}(F^{(l)}, F^{(m)})|$$
$$+ \frac{n}{2}\Big(\sum_{l=1}^{n} \sqrt{\mathbb{V}F^{(l)}} + 1\Big) \sum_{l,m=1}^{n} \sum_{i=1}^{k_l} \sum_{j=1}^{k_m} k_l^{\frac{7}{2}} \sqrt{M_{ij}(f^{(l)}, f^{(m)})}.$$

We shall seek to bound the terms on the right hand side.

Recall the definition of the normalized volume-power functionals $\widehat{\mathcal{V}}_{k_i}^{(\alpha_i)} = \mathcal{V}_{k_i}^{(\alpha_i)}/Q_i$ with $Q_i$ defined in (4.5).

**Theorem 5.9.** *Assume that $(k_1, \alpha_1)$, $\dots$, $(k_n, \alpha_n)$ is an admissible sequence with $4\alpha_i > -d + k_i - 1$ for all $i \in \{1, \dots, n\}$. Let $\boldsymbol{V}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})} = (\widehat{\mathcal{V}}_{k_1}^{(\alpha_1)}, \dots, \widehat{\mathcal{V}}_{k_n}^{(\alpha_n)})$, and let $\boldsymbol{N}(\Sigma_t)$ be the centered $n$-dimensional Gaussian random vector with covariance matrix*

$$\Sigma_t = (\sigma_{lm})_{l,m} \quad \textit{with} \quad \sigma_{lm} = \mathbb{C}\mathrm{ov}(\widehat{\mathcal{V}}_{k_l}^{(\alpha_l)}, \widehat{\mathcal{V}}_{k_m}^{(\alpha_m)}).$$

*Then there is a constant $c_{\boldsymbol{k},\boldsymbol{\alpha}}$ such that*

$$d_3\Big(\boldsymbol{V}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})} - \mathbb{E}\boldsymbol{V}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})}, \boldsymbol{N}(\Sigma_t)\Big) \leq c_{\boldsymbol{k},\boldsymbol{\alpha}} t^{-\frac{1}{2}} \max\{1, (t\delta_t^d)^{-\frac{1}{2}k_n}\}.$$

Thus in the dense and thermodynamic regime a central limit theorem holds with rate of convergence $t^{-\frac{1}{2}}$ which most probably is optimal. In the sparse regime, where $t\delta_t^d \to 0$, the rate of convergence is

$$t^{-\frac{1}{2}} (t\delta_t^d)^{-\frac{1}{2}k_n} = \Theta((\mathbb{E}f_{k_n}(\mathcal{R}(\eta_t, \delta_t)))^{-\frac{1}{2}}),$$

and thus there is a multivariate central limit theorem as long as the expectation $\mathbb{E}f_k$ tend to infinity for all $k \in \boldsymbol{k}$. To the best of our knowledge, Theorem 5.9 is new even in the case $\alpha_i = 0$ where we obtain a central limit theorem for the **f**-vector of the Vietoris-Rips complex, and similar for the Čech complex in the next section.

In the view of the first term on the RHS in Theorem 5.8, it is of interest to state the difference between $\Sigma_t$ and the limiting covariance matrix $\Sigma$ given in (4.17). By equations (4.10), (4.11), (4.13), and (4.14) we see that in the sparse case

$$\frac{1}{2}\sum_{l,m=1}^{n} |\sigma_{lm} - \mathbb{C}\mathrm{ov}(\widehat{\mathcal{V}}_{k_l}^{(\alpha_l)}, \widehat{\mathcal{V}}_{k_m}^{(\alpha_m)})| \leq O(\delta_t + t\delta_t^d),$$

that in the thermodynamic regime, where $\lim_{t\to\infty} t\delta_t^d = c \in (0,\infty)$, this error term is of order

$$O(\delta_t + (c - t\delta_t^d)),$$

and in the dense regime of order

$$O(\delta_t + (t\delta_t^d)^{-1}).$$

Thus the $d_3$-distance $d_3(\boldsymbol{V}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})} - \mathbb{E}\boldsymbol{V}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})}, \boldsymbol{N}(\Sigma))$ would have this additional error terms.

*Proof of Theorem 5.9.* We apply Theorem 5.8 to $\boldsymbol{V}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})} = (\widehat{\mathcal{V}}_{k_1}^{(\alpha_1)}, \ldots, \widehat{\mathcal{V}}_{k_n}^{(\alpha_n)})$, a vector of absolutely convergent Poisson $U$-statistics of orders $k_1+1, \ldots, k_n+1$.

By definition the first term on the RHS in Theorem 5.8 vanishes. And by Theorem 4.6 the variance $\mathbb{V}\widehat{\mathcal{V}}_{k_l}^{(\alpha_l)}$ tends to a constant. Hence we just have to show that the functionals $M_{ij}$ tend to zero. In our case the summands in $M_{ij}$ take the form

$$\frac{t^{|\sigma|}}{(k_l+1)!^2(k_m+1)!^2}\int_{W^{|\sigma|}} \frac{(\Delta_{\delta_t}[\cdot]^{\alpha_l} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_l} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_m} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_m})_\sigma}{Q_l^2 Q_m^2}\, \mathrm{d}x_0 \ldots \mathrm{d}x_{|\sigma|-1}$$

where the first two functionals depend on simplices of volume $k_l$ and the other two on simplices of dimension $k_m$. Assume from now on that $k_l \leq k_m$. The essential feature in the definition of $\sigma$ is that all four functionals $\Delta_{\delta_t}[\cdot]$ are linked by common variables, and each of these functionals depends on $k_l+1$, resp. $k_m+1$ variables. First, for the number $|\sigma|$ of variables this implies

$$\max\{k_l, k_m\} + 1 = k_m + 1 \leq |\sigma| \leq 2(k_l + k_m) + 1 \tag{5.3}$$

since $k_l \leq k_m$. Second, assuming w.l.o.g. that $x_0$ occurs in the first functional, $\Delta_{\delta_t}[\cdot] = \Delta_{\delta_t}[\{\cdot\}_0^{k_l}]$, all other variables are at most at distance $4\delta_t$, and we write this again in the form $\Delta_{4\delta_t}[\{x_j\}_{j=0}^{|\sigma|-1}]^0$. We use this and apply Hölder's inequality, which gives the bound

$$\int_{W^{|\sigma|}} \frac{(\Delta_{\delta_t}[\cdot]^{\alpha_l} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_l} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_m} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_m})_\sigma}{Q_l^2 Q_m^2}\, \mathrm{d}x_0 \ldots \mathrm{d}x_{|\sigma|-1}$$

$$\leq \prod_{i\in\{l,m\}} \left( \int\limits_{W^{|\sigma|}} \frac{\Delta_{\delta_t}[\cdot]^{4\alpha_i}}{Q_i^4} \Delta_{4\delta_t}[\{x_j\}_{j=0}^{|\sigma|-1}]^0 \,\mathrm{d}x_0 \dots \mathrm{d}x_{|\sigma|-1} \right)^{\frac{1}{2}}$$

Now substituting $x_j = \delta_t \tilde{x}_j + x_0$ for $j \geq 1$, changing the order of integration and integrating over $x_0$ gives

$$\int\limits_{W^{|\sigma|}} \frac{(\Delta_{\delta_t}[\cdot]^{\alpha_l} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_l} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_m} \otimes \Delta_{\delta_t}[\cdot]^{\alpha_m})_\sigma}{Q_l^2 Q_m^2} \,\mathrm{d}x_0 \dots \mathrm{d}x_{|\sigma|-1}$$

$$\leq \prod_{i\in\{l,m\}} \delta_t^{2k_i\alpha_i+\frac{1}{2}d(|\sigma|-1)} \left( \int\limits_W \int\limits_{(\delta_t^{-1}(W-x_0)\cap 4B^d)^{|\sigma|-1}} \frac{\Delta_1[0,\{\tilde{x}_j\}_{j=1}^{k_i}]^{4\alpha_i}}{Q_i^4} \,\mathrm{d}\tilde{x}_1 \dots \mathrm{d}\tilde{x}_{|\sigma|-1} \,\mathrm{d}x_0 \right)^{\frac{1}{2}}$$

$$\leq \prod_{i\in\{l,m\}} \delta_t^{2k_i\alpha_i+\frac{1}{2}d(|\sigma|-1)} \left( \int\limits_{(4B^d)^{|\sigma|-1}} \frac{\Delta_1[0,\{\tilde{x}_j\}_{j=1}^{k_i}]^{4\alpha_i}}{Q_i^4} \,\mathrm{d}\tilde{x}_1 \dots \mathrm{d}\tilde{x}_{|\sigma|-1} \right)^{\frac{1}{2}}$$

$$= \frac{\delta_t^{2(k_l\alpha_l+k_m\alpha_m)+d(|\sigma|-1)}}{Q_l^2 Q_m^2} \prod_{i\in\{l,m\}} \left( (4\kappa^d)^{|\sigma|-1-k_i} \mu_{k_i}^{(4\alpha_i)} \right)^{\frac{1}{2}}.$$

By Lemma 4.1, this is finite for $4\alpha_i > -d + k_i - 1$ for $i \in \{l,m\}$. By (5.3) and the definition of $M_{ij}$, this implies

$$\begin{aligned} M_{ij}\left(\frac{\Delta_{\delta_t}[\cdot]^{\alpha_l}}{Q_l}, \frac{\Delta_{\delta_t}[\cdot]^{\alpha_m}}{Q_m}\right) &= \sum_\alpha \frac{t^{|\sigma|}\delta_t^{2(k_l\alpha_l+k_m\alpha_m)+d(|\sigma|-1)}}{(k_l+1)!^2(k_m+1)!^2 Q_l^2 Q_m^2} \prod_{i\in\{l,m\}} \left( (4\kappa^d)^{|\sigma|-1-k_i} \mu_{k_i}^{(4\alpha_i)} \right)^{\frac{1}{2}} \\ &\leq c_2 t \, \frac{\delta_t^{2(k_l\alpha_l+k_m\alpha_m)}}{Q_l^2 Q_m^2} \max\{(t\delta_t^d)^{k_m}, (t\delta_t^d)^{2(k_l+k_m)}\} \end{aligned}$$

with $c_2$ depending on $\boldsymbol{k}, \boldsymbol{\alpha}$. Plugging the definition (4.5) of $Q_i$ into this shows

$$\begin{aligned} M_{ij}\left(\frac{\Delta_{\delta_t}[\cdot]^{\alpha_l}}{Q_l}, \frac{\Delta_{\delta_t}[\cdot]^{\alpha_m}}{Q_m}\right) &\leq c_2 t \, \frac{\delta_t^{2(k_l\alpha_l+k_m\alpha_m)} \max\{(t\delta_t^d)^{k_m}, (t\delta_t^d)^{2(k_l+k_m)}\}}{t\delta_t^{2\alpha_l k_l} \max\{(t\delta_t^d)^{2k_l}, (t\delta_t^d)^{k_l}\} t\delta_t^{2\alpha_m k_m} \max\{(t\delta_t^d)^{2k_m}, (t\delta_t^d)^{k_m}\}} \\ &= c_2 t^{-1} \frac{\max\{(t\delta_t^d)^{k_m}, (t\delta_t^d)^{2(k_l+k_m)}\}}{\max\{(t\delta_t^d)^{k_l+k_m}, (t\delta_t^d)^{2(k_l+k_m)}\}} \\ &= c_2 t^{-1} \max\{1, (t\delta_t^d)^{-k_l}\}. \end{aligned}$$

The sum over all $M_{ij}$ yields

$$\begin{aligned} \sum_{l,m=1}^n \sum_{i=1}^{k_l+1} \sum_{j=1}^{k_m+1} (k_l+1)^{\frac{7}{2}} \sqrt{M_{ij}(f^{(l)}, f^{(m)})} &= \sum_{l,m=1}^n \sum_{i=1}^{k_l+1} \sum_{j=1}^{k_m+1} (k_l+1)^{\frac{7}{2}} \sqrt{c_2}\, t^{-\frac{1}{2}} \max\{1, (t\delta_t^d)^{-\frac{1}{2}k_l}\} \\ &\leq \sum_{l,m=1}^n c_3 t^{-\frac{1}{2}} \max\{1, (t\delta_t^d)^{-\frac{1}{2}k_l}\} \end{aligned}$$

$$\leq \quad c_4 t^{-\frac{1}{2}} \max\{1, (t\delta_t^d)^{-\frac{1}{2}k_n}\}$$

with constants depending on $\boldsymbol{k}$ and $\boldsymbol{\alpha}$. □

# 5.4. The Čech complex

It follows from the definition of the Vietoris-Rips complex and the Čech complex that

$$\mathcal{C}(\eta_t, \delta_t) \subset \mathcal{R}(\eta_t, \delta_t) \subset \mathcal{C}(\eta_t, (\tfrac{2d}{d+1})^{\frac{1}{2}}\ \delta_t).$$

That the optimal factor in the second inclusion is $(\frac{2d}{d+1})^{\frac{1}{2}}$ was shown by de Silva and Ghrist [dSG07, Theorem 2.5]. Hence all bounds obtained for the Vietoris-Rips complex hold true for the Čech complex with constants changed by a factor of $(\frac{2d}{d+1})^{\frac{1}{2}}$. The constants in the expectation and covariance change in the following way. Denote by $\Delta_s^c[x_0, \dots x_k]$ the $k$-dimensional volume of the convex hull of the points $x_0, \dots, x_k$ if the intersection $\bigcap_1^k B^d(x_i, \frac{s}{2})$ is not empty, and set $\Delta_s^c[x_0, \dots, x_k] = 0$ otherwise. In the case $k > d$ we only define $\Delta_s^c[x_0, \dots x_k]^0 = 1$ if the intersection property holds. Thus for all $k \geq 0$,

$$F \in \mathcal{C}_k(\mathcal{R}(\eta_t, \delta_t)) \Leftrightarrow \Delta_{\delta_t}^c(F)^0 = 1.$$

We define

$$\zeta_k^{(\alpha)} = \int\limits_{(B^d)^k} \Delta_1^c[0, \{x_l\}_{l=1}^k]^\alpha \ \mathrm{d}x_1 \cdots \mathrm{d}x_k$$

with $\zeta_0^{(\alpha)} = 1$. In the case $k > d$ the definition only applies to $\alpha = 0$. Again $\alpha > -d+k-1$ ensures that $\zeta_k^{(\alpha)} < \infty$. The volume-power functional of the Čech complex is given by

$$\mathcal{U}_k^{(\alpha)} = \frac{1}{(k+1)!} \sum_{(x_0,\dots,x_k) \in \eta_{t,\neq}^k} \Delta_{\delta_t}^c[\{x_l\}_{l=0}^k]^\alpha.$$

Then the Čech complex version of Theorem 3.6 holds for $\mathcal{U}_k^{(\alpha)}$ with $\mu_k^{(\alpha)}$ replaced by $\zeta_k^{(\alpha)}$. Analogously, define

$$\zeta_{k_1,k_2:m}^{(\alpha_1,\alpha_2)} = \int\limits_{(B^d)^{k_1+k_2+1-m}} \Delta_1^c[0, \{x_l\}_{l=1}^{k_1}]^{\alpha_1} \Delta_1^c[0, \{x_l\}_{l=k_1-m+2}^{k_1+k_2-m+1}]^{\alpha_2}$$
$$\mathrm{d}x_1 \cdots \mathrm{d}x_{k_1+k_2-m+1}.$$

Then Theorem 4.3, Theorem 4.6 and Theorem 4.7 hold for $\mathcal{U}_{k_i}^{(\alpha_i)}$ with a covariance matrix $\Sigma^c$ with $\mu_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}$ replaced by $\zeta_{k_1,k_2:m}^{(\alpha_1,\alpha_2)}$. Finally, the proofs of the central limit theorems only depend on the local behavior of the random simplicial complexes and thus the identical proof holds for the Čech complex. Define the normalized volume-power functionals by

$$\widehat{\mathcal{U}}_{k_i}^{(\alpha_i)} = \frac{1}{Q_i}\mathcal{U}_{k_i}^{(\alpha_i)}$$

with $Q_i$ defined in (4.5), and let $\boldsymbol{U}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})} = (\widehat{\mathcal{U}}_{k_1}^{(\alpha)}, \dots, \widehat{\mathcal{U}}_{k_n}^{(\alpha)})$. Assume that $(k_1, \alpha_1)$, ..., $(k_n, \alpha_n)$ is an admissible sequence with $4\alpha_i > -d + k_i - 1$ for all $i \in \{1, \dots, n\}$.

**Theorem 5.10.** *For $d_\star = d_W$ or $d_\star = d_K$ there is a constant $c_k$ such that*

$$d_\star\left(\frac{\mathcal{U}_k^\alpha - \mathbb{E}\mathcal{U}_k^\alpha}{\sqrt{\mathbb{V}\,\mathcal{U}_k^\alpha}}, N\right) \le c_k t^{-\frac{1}{2}} \max\{(t\delta_t^d)^{-\frac{k}{2}}, 1\}.$$

*And there is a constant $c_{\boldsymbol{k},\boldsymbol{\alpha}}$ such that*

$$d_3\left(\boldsymbol{U}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})} - \mathbb{E}\boldsymbol{U}_{\boldsymbol{k}}^{(\boldsymbol{\alpha})}, \boldsymbol{N}(\Sigma_t^c)\right) \quad \le \quad c_{\boldsymbol{k},\boldsymbol{\alpha}} t^{-\frac{1}{2}} \max\{1, (t\delta_t^d)^{-\frac{1}{2}k_n}\}$$

*where $\Sigma_t^c$ is the covariance matrix $\Sigma_t^c = (\mathbb{C}\mathrm{ov}(\widehat{\mathcal{U}}_{k_l}^{(\alpha_l)}, \widehat{\mathcal{U}}_{k_m}^{(\alpha_m)}))_{lm}$.*

# 6. Poisson Limit Theorem

## 6.1. The $f$-vector

Recall that the $\boldsymbol{f}$-vector of a simplicial complex, $\Delta$, is given by

$$(f_0(\Delta), f_1(\Delta), f_2(\Delta) \dots)$$

where $f_i(\Delta)$ is the number of $i$-dimensional simplices in the simplicial complex.

We are interested in the Poisson functional which counts the $k$-simplices in the Vietoris-Rips complex, that is, $f_k = f_k(\mathcal{R}(\eta_t, \delta_t))$. So we consider the entries of the $\boldsymbol{f}$-vector $(f_k)_{k\geq 0} = (f_k(\mathcal{R}(\eta_t, \delta_t)))_{k\geq 0}$ of the Vietoris-Rips complex. Once again we consider the Poisson point process, $\eta_t$, with intensity $t > 0$, and intensity measure $\mu$, on a state space, $W$.

We denote by $F_s[x_0, \dots x_k]$ the indicator function asking if, in the convex hull of the points $x_0, \dots, x_k \in W$, all edges have length at most $s$. That is,

$$F_s[x_0, \dots x_k] = \prod_{i,j=0}^{k} \mathbb{1}(\|x_i - x_j\| \leq s).$$

We write $F_s[\{x_l\}_{l=0}^k] = F_s[x_0, \dots x_k]$ for a short notation. We note that $F_s[\{x_l\}_{l=0}^k] = \Delta_s[\{x_l\}_{l=0}^k]^0$ where $\Delta_s[\{x_l\}_{l=0}^k]$ has been defined earlier to be the $k$-dimensional volume of the convex hull of the points $x_0, \dots, x_k$ if all edges have length at most $s$, and 0 otherwise.

Relating these to the Poisson functional $\mathcal{V}_k^{(\alpha)}$ discussed in previous chapters, we observe that

$$f_k = \mathcal{V}_k^{(0)} = \frac{1}{(k+1)!} \sum_{(x_0,\dots,x_k)\in\eta_{t,\neq}^{k+1}} F_{\delta_t}[x_0, \dots, x_k]. \tag{6.1}$$

It is clear that this functional is $\mathbb{N}$-valued.

By Theorem 3.6 and Theorem 4.3, the expectation and variance of $f_k$ is given by

$$\mathbb{E} f_k = \frac{\mu_k}{(k+1)!}\, t(t\delta_t^d)^k (1+O(\delta_t)) \tag{6.2}$$

and

$$\mathbb{V}\, f_k = \sum_{m=1}^{k+1} \frac{\mu_{k,k:m}}{m!((k-m+1)!)^2} t(t\delta_t)^{2k-m+1}(1+O(\delta_t)) \tag{6.3}$$

respectively, where

$$\mu_k = \mu_k^{(0)} = \int\limits_{(B^d)^k} F_1[0,\{x_l\}_{l=1}^k]\, \mathrm{d}x_1 \cdots \mathrm{d}x_k \tag{6.4}$$

and

$$\mu_{k,k:m} = \mu_{k,k:m}^{(0,0)} = \int\limits_{(B^d)^{2k-m+1}} F_1[0,\{x_l\}_{l=1}^k] F_1[0,\{x_l\}_{l=k-m+2}^{2k-m+1}]\, \mathrm{d}x_1 \cdots \mathrm{d}x_{2k-m+1}$$

as defined in (3.3) and (4.1).

We note by (4.3) that for $m = k+1$,

$$\mu_{k,k:k+1} = \mu_k. \tag{6.5}$$

The next Lemma shows the connection between the expectation and variance of $f_k$ in the sparse regime.

**Lemma 6.1.** *In the sparse regime,*

$$\mathbb{V}\, f_k = \mathbb{E} f_k \left[1 + O(t\delta_t^d)\right].$$

*Proof.* By (6.2) and (6.3) together with (6.5), we observe that the summand at $m = k+1$ in $\mathbb{V}\, f_k$ is precisely $\mathbb{E} f_k$, so that we have

$$\begin{aligned}
\mathbb{V}\, f_k &= \mathbb{E} f_k + \left[\sum_{m=1}^{k} \frac{\mu_{k,k:m}}{m!((k-m+1)!)^2} t(t\delta_t^d)^{2k+1-m}\right][1+O(\delta_t)] \\
&= \mathbb{E} f_k + t(t\delta_t^d)^k \left[\sum_{m=1}^{k} \frac{\mu_{k,k:m}}{m!((k-m+1)!)^2} (t\delta_t^d)^{k+1-m}\right][1+O(\delta_t)] \\
&= \mathbb{E} f_k + \frac{(k+1)!}{\mu_k} \mathbb{E} f_k \left[\sum_{m=1}^{k} \frac{\mu_{k,k:m}}{m!((k-m+1)!)^2} (t\delta_t^d)^{k+1-m}\right] \\
&= \mathbb{E} f_k\, [1+O(t\delta_t^d)].
\end{aligned}$$

□

*Remark* 6.2.

(i) Lemma 6.1 implies that if $\mathbb{E} f_k \to c$ in the sparse regime, that is, $\lim_{t\to\infty} t\delta_t^d = 0$, then $\mathbb{V} f_k \to c$ in the limit. In this case, we hope to get a Poisson limit theorem for $f_k$.

(ii) We observe from (6.2) that for $i, j \in \mathbb{N}$, $i \leq k$,

$$\mathbb{E} f_{k-i} = (t\delta_t^d)^{-i}\mathbb{E} f_k \qquad \text{and} \qquad \mathbb{E} f_{k+j} = (t\delta_t^d)^j \mathbb{E} f_k.$$

Thus, in the sparse regime, once $\mathbb{E} f_k$ tends to a constant, $c$, then $\mathbb{E} f_{k-i} \to \infty$ and $\mathbb{E} f_{k+j} \to 0$ immediately. This makes it impossible to get a multivariate Poisson limit theorem for the $\boldsymbol{f}$-vector.

## 6.2. Moments of First and Second Order Difference Operators

We recall the difference operator, $D_z F$, defined by

$$D_z F = F(\eta + \delta_z) - F(\eta),$$

for a Poisson functional $F$ and $z \in W$. This definition applied to a Poisson U-statistic

$$F = \sum_{(x_1,\dots,x_k)\in\eta^k_{\neq}} f(x_1,\dots,x_k)$$

of order $k$ yields

$$D_z F = k \sum_{(x_1,\dots,x_{k-1})\in\eta^{k-1}_{\neq}} f(z, x_1,\dots,x_{k-1})$$

by the symmetry of $f$, and more generally, for $n \leq k$, we have

$$D_{z_1,\dots,z_n} F = \frac{k!}{(k-n)!} \sum_{(x_1,\dots,x_{k-n})\in\eta^{k-n}_{\neq}} f(z_1,\dots,z_n,x_1,\dots,x_{k-n}).$$

We note by (6.1) that $f_k$ is a Poisson U-statistic of order $k+1$, so that we have by applying the difference operator

$$D_y f_k = \frac{1}{k!} \sum_{(x_0,\dots,x_{k-1})\in\eta^{(k)}_{t,\neq}} F_{\delta_t}[y, \{x_l\}_{l=0}^{k-1}]$$

and

$$D_{y_1,y_2} f_k = \frac{1}{(k-1)!} \sum_{(x_0,\dots,x_{k-2})\in\eta^{(k-1)}_{t,\neq}} F_{\delta_t}[y_1, y_2, \{x_l\}_{l=0}^{k-2}]$$

for $y, y_1, y_2 \in W$.

The following formula was given in [Sch13, Theorem 3.4] for the moments of Poisson U-statistics using the notion of partitions discussed earlier.

**Theorem 6.3.** *Let $F^{(1)}, \dots, F^{(n)}$, $n \geq 2$, be Poisson U-statistics given by*

$$F^{(l)} = \sum_{(x_1,\dots,x_{k_l}) \in \eta_{t,\neq}^{k_l}} f^{(l)}(x_1, \dots, x_{k_l})$$

*with $f^{(l)} \in L_s^1(\mu^{k_l})$, $k_l \in \mathbb{N}$ for $l = 1, \dots, n$ with the assumption that*

$$\int_{W^{|\sigma|}} \left| \left( \otimes_{l=1}^n f^{(l)} \right)_\sigma \right| \, \mathrm{d}\mu^{|\sigma|} < \infty$$

*for all $\sigma \in \Pi(k_1, \dots, k_n)$. Then*

$$\mathbb{E} \prod_{l=1}^n F^{(l)} = \sum_{\alpha \in \Pi(k_1,\dots,k_n)} \int_{W^{|\sigma|}} (\otimes_{l=1}^n f(l))_\alpha \, \mathrm{d}\mu^{|\sigma|}.$$

We seek to evaluate moments of $D_y f_k$ and $D_{y_1,y_2} f_k$ for $y, y_1, y_2 \in W$. We use the following notations for $p \geq 2$.

$$F^{\otimes p} := F \otimes \cdots \otimes F, \ \ p \text{ copies}, \quad \text{and} \quad \{k\}_p := \{k, \dots, k\}, \ \ p \text{ elements}.$$

For $k$ independently and uniformly distributed points in the unit ball, $\{X_l\}_{l=0}^{k-1}$, which form a random $k$-simplex with the origin, we denote by $\mu_{|\sigma|,p}$ its $p$-tensor product partitioned by $\sigma \in \Pi(\{k\}_p)$, if all edges are bounded by one.

$$\mu_{|\sigma|,p} = \int_{(B^d)^{|\sigma|}} \left( F_1[0, \{x_l\}_{l=0}^{k-1}]^{\otimes p} \right)_\sigma \, \mathrm{d}x_1 \dots \mathrm{d}x_{|\sigma|}.$$

Next, we evaluate $\mu_{|\sigma|,p}$ for $|\sigma| = k$ which we would need in the sequel.

For $|\sigma| = k$, each block in $\sigma \in \Pi(\{k\}_p)$ contains $p$ variables, one from each functional. Thus, we have in this case that

$$\left( F_1[0, \{x_l\}_{l=0}^{k-1}]^{\otimes p} \right)_\sigma = \left( F_1[0, \{x_l\}_{l=0}^{k-1}] \right)^p = F_1[0, \{x_l\}_{l=0}^{k-1}]$$

since $F_1[0, \{x_l\}_{l=0}^{k-1}]$ is a product of indicator functions. Thus, $\mu_{k,p} = \mu_k$ for all $p$, where $\mu_k$ is given by (6.4).

**Lemma 6.4.** *Let $y \in W$ and $p \geq 2$, there is a constant $c_{k,p}$ depending only on $k$ and $p$ such that*

$$\mathbb{E}\left[(D_y f_k)^p\right] = (t\delta_t^d)^k \left[ \frac{\mu_k}{k!} + c_{k,p}(t\delta_t^d) + O((t\delta_t^d)^2) \right] (1 + O(\delta_t))$$

*Proof.* The functionals $F_{\delta_t}[y,\cdot]$ are indeed a product of indicator functions and depend on $k$-simplices, where the vertex $y$ connects all vertices of a $(k-1)$-simplex. We note that $\left(F_{\delta_t}[y,\{x_l\}_{l=0}^{k-1}]^{\otimes p}\right)_\sigma$ is a $p$-product of these functionals with variables determined by the partition $\sigma$. Also, since $\sigma \in \Pi(\{k\}_p)$, these $p$ functionals are not necessarily connected in these variables, they only need be connected in $y$. Below, we will substitute $x_i = \delta_t \tilde{x}_i + y$ in the integral and recall that $\delta_t^{-1}(W-y) \cap B^d = B^d$ for $y \in W_{-\delta_t}$. Thus we have

$$\begin{aligned}
\mathbb{E}\left[(D_y f_k)^p\right] &= \frac{1}{k!^p} \sum_{\sigma \in \Pi(\{k\}_p)} \int_{W^{|\sigma|}} \left(F_{\delta_t}[y,\{x_l\}_{l=0}^{k-1}]^{\otimes p}\right)_\sigma \, \mathrm{d}\mu^{|\sigma|} \\
&= \frac{1}{k!^p} \sum_{\sigma \in \Pi(\{k\}_p)} t^{|\sigma|} \delta_t^{d|\sigma|} \int_{(\delta_t^{-1}(W-y)\cap B^d)^{|\sigma|}} \left(F_1[0,\{\tilde{x}_l\}_{l=0}^{k-1}]^{\otimes p}\right)_\sigma \, \mathrm{d}\tilde{x}_1 \dots \mathrm{d}\tilde{x}_{|\sigma|} \\
&= \frac{1}{k!^p} \sum_{\sigma \in \Pi(\{k\}_p)} \mu_{|\sigma|,p} (t\delta_t^d)^{|\sigma|} (1 + O(\delta_t))
\end{aligned}$$

Now, by (2.8), we have $k \le |\sigma| \le pk$. Since the sparse regime is under consideration, the maximum summand is the one for which $|\sigma|$ is least, that is $|\sigma| = k$. By Theorem 2.9 and Remark 2.10 (*iii*), the total number of such partitions $\sigma \in \Pi(\{k\}_p)$ such that $|\sigma| = k$, is $k!^{(p-1)}$.

We now move a step further. For $\sigma \in \Pi(\{k\}_p)$ such that $|\sigma| = k+1$, again by Remark 2.10 (*iv*), the total number of such partitions is given by

$$kk!^{p-1} \sum_{i=0}^{p-2} (k+1)^i.$$

We note that for $p = 2$, this number is $kk!$.

We thus have

$$\begin{aligned}
\mathbb{E}\left[(D_y f_k)^p\right] &= \frac{1}{k!^p} \sum_{\sigma \in \Pi(\{k\}_p)} \mu_{|\sigma|,p} (t\delta_t^d)^{|\sigma|} (1 + O(\delta_t)) \\
&= k!^{(p-1)} \frac{\mu_{k,p}(t\delta_t^d)^k}{k!^p} + kk!^{(p-1)} \sum_{i=0}^{p-2} (k+1)^i \frac{\mu_{k+1,p}(t\delta_t^d)^{k+1}}{k!^p} + \cdots (1 + O(\delta_t)) \\
&= (t\delta_t^d)^k \left[\frac{\mu_k}{k!} + c_{k,p}(t\delta_t^d) + O((t\delta_t^d)^2)\right] (1 + O(\delta_t))
\end{aligned}$$

where $c_{k,p} = \dfrac{\mu_{k+1,p} \sum_{i=0}^{p-2} (k+1)^i}{(k-1)!}$.

$\square$

*Remark* 6.5.

(i) We were particular about getting an equality in the above lemma and not an upper bound. This will be useful in the Poisson approximation which we will encounter later.

(ii) For any value of $p$ we have the following upper bound since $t\delta_t^d \to 0$.

$$\mathbb{E}\big[(D_y f_k)^p\big] \le c'_{k,p} \frac{\mu_k (t\delta_t^d)^k}{k!}.$$

**Lemma 6.6.** *For $y, y_1, y_2 \in W$,*

$$\mathbb{E}\left[(D^2_{y_1,y_2} f_k)^4\right] \le c_k (t\delta_t^d)^{k-1}$$

*Proof.* For $\left(F_{\delta_t}[y_1, y_2, \{x_l\}_{l=0}^{k-2}]^{\otimes 4}\right)_\sigma$, $\sigma \in \Pi(\{k-1\}_4)$, we observe that all functionals are linked by $y_1$ and $y_2$ but not necessarily by other variables, so that $k-1 \le |\sigma| \le 4(k-1)$. Also max $\|x_i - y_1\| \le 4\delta_t$ for all the variables, and although this holds for $y_2$, it is enough since $y_1$ and $y_2$ are connected. We evaluate as follows.

$$\begin{aligned}
\mathbb{E}\big[(D^2_{y_1,y_2} f_k)^4\big] &= \frac{1}{(k-1)!^4} \sum_{\sigma\in\Pi(\{k-1\}_4)} \int_{W^{|\sigma|}} \left(F_{\delta_t}[y_1, y_2, \{x_l\}_{l=0}^{k-2}]^{\otimes 4}\right)_\sigma \, \mathrm{d}\mu^{|\sigma|} \\
&\le \frac{1}{(k-1)!^4} \sum_{\sigma\in\Pi(\{k-1\}_4)} t^{|\sigma|} \int_{W^{|\sigma|}} \mathbb{1}(\forall i\colon \|x_i - y_1\| \le 4\delta_t, \|y_1 - y_2\| \le \delta_t) \\
&\qquad\qquad \mathrm{d}x_0 \dots \mathrm{d}x_{|\sigma|-1} \\
&\le \frac{1}{(k-1)!^4} \sum_{\sigma\in\Pi(\{k-1\}_4)} t^{|\sigma|} (4\delta_t)^{d|\sigma|} \; \mathbb{1}(y_1 \in B(y_2, \delta_t)) \\
&\le c_k \max\{(t\delta_t^d)^{k-1}, (t\delta_t^d)^{4(k-1)}\} \; \mathbb{1}(y_1 \in B(y_2, \delta_t)) \\
&= c_k (t\delta_t^d)^{k-1} \; \mathbb{1}(y_1 \in B(y_2, \delta_t)),
\end{aligned}$$

being in the sparse regime, where $c_k$ is a constant depending only on $k$ and $d$. □

## 6.3. Poisson limit theorem for Poisson Functionals

The Stein's method together with Malliavin calculus have been effective in get a normal approximation for a Poisson functional. Its analogue known as the Chen-Stein method is used in Poisson approximations, that is, how close a random variable is to a Poisson distribution. The total variation distance has been effective for these kind of approximations. It is essential that the random variables considered in this case are $\mathbb{Z}_+$-valued with positive expectations.

The Chen-Stein method was used in [BP16] to get a Poisson approximation for Poisson functionals as given below.

**Theorem 6.7.** *Let $Z \sim \mathrm{Po}(c)$, $c > 0$ and assume that $F \in L^2(\mathbb{P})$ is an element of dom $D$ such that $\mathbb{E}F = c$ and $F$ takes values in $\mathbb{Z}_+$. Then*

$$\begin{aligned} d_{TV}(F, Z) &\leq \frac{1-e^{-c}}{c}\mathbb{E}|c - \langle DF, -DL^{-1}F\rangle_{L^2(\mu)}| \\ &\quad + \frac{1-e^{-c}}{c^2}\mathbb{E}\Big[\int_W |D_zF(D_zF-1)D_zL^{-1}F| \,\mathrm{d}\mu(z)\Big] \\ &\leq \frac{1-e^{-c}}{c}\sqrt{\Big[\mathbb{E}\big(c - \langle DF, -DL^{-1}F\rangle_{L^2(\mu)}\big)^2\Big]} \\ &\quad + \frac{1-e^{-c}}{c^2}\mathbb{E}\Big[\int_W |D_zF(D_zF-1)D_zL^{-1}F| \,\mathrm{d}\mu(z)\Big]. \end{aligned}$$

We compare the above theorem with the normal approximation given in Theorem 5.2. The expectations in the first term have similar evaluations since the value of the variance is 1 in the normal approximation, being a normalised random variable and the variance in the Poisson approximation is $c$. This informs the similarity in the first two summands of [LPS16, Theorem 1.1] and [Gry19, Theorem 1.2]. The later is given below.

First, we define

$$\begin{aligned} \gamma_1(F) &:= \int_{W^3} \Big(\mathbb{E}\big[(D^2_{x_1,x_3}F)^4\big]\mathbb{E}\big[(D^2_{x_2,x_3}F)^4\big]\mathbb{E}\big[(D_{x_1}F)^4\big]\mathbb{E}\big[(D_{x_2}F)^4\big]\Big)^{\frac{1}{4}} \,\mathrm{d}\mu^3(x_1,x_2,x_3) \\ \gamma_2(F) &:= \int_{W^3} \Big(\mathbb{E}\big[(D^2_{x_1,x_3}F)^4\big]\mathbb{E}\big[(D^2_{x_2,x_3}F)^4\big]\Big)^{\frac{1}{2}} \,\mathrm{d}\mu^3(x_1,x_2,x_3) \\ \gamma_3(F) &:= \int_W \Big(\mathbb{E}|D_xF(D_xF-1)|^2\Big)^{\frac{1}{2}}\Big(\mathbb{E}|D_xF|^2\Big)^{\frac{1}{2}} \,\mathrm{d}\mu(x) \end{aligned} \tag{6.6}$$

**Theorem 6.8.** *Let $F$ be a $\mathbb{N}$-valued Poisson functional satisfing $F \in \mathrm{dom}D$, and let $Z \sim \mathcal{P}(c)$, that is, a Poisson distributed random variable with parameter $c > 0$. Then*

$$d_{TV}(F, Z) \leq \frac{1-e^{-c}}{c}\big(2\sqrt{\gamma_1(F)} + \sqrt{\gamma_2(F)} + \frac{\gamma_3(F)}{c} + |\mathbb{E}F - c| + |\mathbb{V}\,F - c|\big).$$

We now present the main theorem in this section, that is the Poisson approximation for $f_k$, the $k$-th component of the $\boldsymbol{f}$-vector. First, we have the following.

**Lemma 6.9.** *There are contants $\boldsymbol{C}_1$, $\boldsymbol{C}_2$ and $\boldsymbol{C}_3$ depending on $k$ such that*

*(i)* $\gamma_1(f_k) \leq \boldsymbol{C}_1 t(t\delta_t^d)^{k+\frac{3}{2}}$,

*(ii)* $\gamma_2(f_k) \leq \boldsymbol{C}_2 t(t\delta_t^d)^{k+1}$,

*(iii)* $\gamma_3(f_k) \leq \boldsymbol{C}_3 t(t\delta_t^d)^{k+\frac{1}{2}}$,

*Proof.* We apply 6.6, Remark 6.5 and Lemmas 6.4 and 6.6 to get the following.

(i)

$$\begin{aligned}\gamma_1(f_k) &\le t^3 \int_{W^3} \Big( c_k (t\delta_t^d)^{k-1}\ \mathbb{1}(x_1 \in B(x_3,\delta_t)) c_k (t\delta_t^d)^{k-1}\ \mathbb{1}(x_2 \in B(x_3,\delta_t)) \\ &\qquad\qquad c'_{k,4} \frac{\mu_k (t\delta_t^d)^k}{k!}\ \ c'_{k,4}\frac{\mu_k (t\delta_t^d)^k}{k!}\Big)^{\frac{1}{4}}\ \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \\ &\le t^3 \int_{W^3} c'_k \Big( (t\delta_t^d)^{2(k-1)} (t\delta_t^d)^{2k}\Big)^{\frac{1}{4}} \mathbb{1}(x_1, x_2 \in B(x_3,\delta_t))\ \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \\ &\le \boldsymbol{C}_1 t (t\delta_t^d)^2 (t\delta_t^d)^{k-\frac{1}{2}} \\ &= \boldsymbol{C}_1 t (t\delta_t^d)^{k+\frac{3}{2}}.\end{aligned}$$

(ii)

$$\begin{aligned}\gamma_2(f_k) &\le t^3 \int_{W^3} \Big( c_k (t\delta_t^d)^{k-1}\ \mathbb{1}(x_1 \in B(yx_3,\delta_t)) c_k (t\delta_t^d)^{k-1}\ \mathbb{1}(x_2 \in B(x_3,\delta_t))\Big)^{\frac{1}{2}} \\ &\qquad\qquad \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \\ &\le t^3 \int_{W^3} c_k \Big( (t\delta_t^d)^{2(k-1)}\Big)^{\frac{1}{2}}\ \mathbb{1}(x_1, x_2 \in B(x_3,\delta_t))\ \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \\ &\le \boldsymbol{C}_1 t (t\delta_t^d)^2 (t\delta_t^d)^{k-1} \\ &= \boldsymbol{C}_1 t (t\delta_t^d)^{k+1}.\end{aligned}$$

(iii) Here the first terms in the expectations cancel out which is the reason why we needed an equality in Lemma 6.4.

$$\begin{aligned}\gamma_3(f_k) &= t\int_W \Big(\mathbb{E}(D_x f_k)^4 - 2\mathbb{E}(D_x f_k)^3 + \mathbb{E}(D_x f_k)^2\Big)^{\frac{1}{2}} \Big(\mathbb{E}(D_x f_k)^2\Big)^{\frac{1}{2}}\ \mathrm{d}x \\ &= t\int_W \Big( (t\delta_t^d)^k \Big[\frac{\mu_k}{k!} + c_{k,4}(t\delta_t^d) - 2\Big(\frac{\mu_k}{k!} + c_{k,3}(t\delta_t^d)\Big) + \frac{\mu_k}{k!} + c_{k,2}(t\delta_t^d) + O((t\delta_t^d)^2)\Big]\Big)^{\frac{1}{2}} \\ &\qquad\qquad \Big( (t\delta_t^d)^k \Big[\frac{\mu_k}{k!} + c_{k,2}(t\delta_t^d) + O((t\delta_t^d)^2)\Big]\Big)^{\frac{1}{2}} (1 + O(\delta_t))\ \ \mathrm{d}x \\ &\le t(t\delta_t^d)^k \int_W c_k (t\delta_t^d)^{\frac{1}{2}} (1 + O(\delta_t) + O(t\delta_t^d))\ \mathrm{d}x \\ &\le \boldsymbol{C}_3 t (t\delta_t^d)^{k+\frac{1}{2}}.\end{aligned}$$

□

In the sparse regime, we consider the case where $\mathbb{E}f_k$ tends to a constant in $(0,\infty)$. This happens when $t(t\delta_t^d)^k$ tends to a constant in the limit.

We have the following error term for the Poisson approximation.

**Theorem 6.10.**

(i) *Let* $\mathbb{E} f_k = c_t$, $c_t > 0$ *and* $Z_t \sim \mathcal{P}(c_t)$. *Then in the sparse regime, for some constant* $\boldsymbol{c_{k,t}}$ *depeding only on* $k$ *and* $t$, *we have*

$$d_{TV}(f_k, Z_t) \leq \boldsymbol{c_{k,t}} t^{-\frac{1}{2k}}$$

*in the limit.*

(ii) *From* $(i)$, *let* $c_t \xrightarrow{t\to\infty} c$ *in the sparse regime, and* $Z \sim \mathcal{P}(c)$, $c > 0$. *Then for some constant* $\boldsymbol{c_k}$ *depeding only on* $k$, *we have*

$$d_{TV}(f_k, Z) \leq \boldsymbol{c_k} t^{-\frac{1}{2k}} + |\mathbb{E} f_k - c|$$

*in the limit.*

*Proof.*

(i) We note by the hypothesis of the theorem that $t(t\delta_t^d)^k$ is a constant, which implies that $t\delta_t^d = t^{-\frac{1}{k}}$ up to a constant. By Theorem 6.8 and Lemma 6.1, we have,

$$\begin{aligned}
d_{TV}(f_k, Z_t) &\leq \frac{1-e^{-c_t}}{c_t}\left(2(\boldsymbol{C}_1 t(t\delta_t^d)^{k+\frac{3}{2}})^{\frac{1}{2}} + (\boldsymbol{C}_2 t(t\delta_t^d)^{k+1})^{\frac{1}{2}} + \frac{\boldsymbol{C}_3}{c_t} t(t\delta_t^d)^{k+\frac{1}{2}} + |\mathbb{V} f_k - c_t|\right) \\
&\leq \frac{1-e^{-c_t}}{c_t} c_{k,t}\left(t^{-\frac{3}{4k}} + t^{-\frac{1}{2k}} + t^{-\frac{1}{2k}} + O(t^{-\frac{1}{k}})\right) \\
&\leq \frac{1-e^{-c_t}}{c_t} c_{k,t} t^{-\frac{1}{2k}}\left(t^{-\frac{1}{4k}} + 2 + O(t^{-\frac{1}{2k}})\right) \\
&\leq \boldsymbol{c_{k,t}} t^{-\frac{1}{2k}}\ O(t^{-\frac{1}{4k}}).
\end{aligned}$$

(ii) We apply the triangle inequality and evaluate as before to get

$$\begin{aligned}
d_{TV}(f_k, Z) &\leq \frac{1-e^{-c}}{c}\Big(2(\boldsymbol{C}_1 t(t\delta_t^d)^{k+\frac{3}{2}})^{\frac{1}{2}} + (\boldsymbol{C}_2 t(t\delta_t^d)^{k+1})^{\frac{1}{2}} + \frac{\boldsymbol{C}_3}{c} t(t\delta_t^d)^{k+\frac{1}{2}} \\
&\qquad + |\mathbb{E} f_k - c| + |\mathbb{V} f_k - \mathbb{E} f_k| + |\mathbb{E} f_k - c|\Big) \\
&\leq \frac{1-e^{-c}}{c} c_k t^{-\frac{1}{2k}}\left(t^{-\frac{1}{4k}} + 2 + O(t^{-\frac{1}{2k}})\right) + 2|\mathbb{E} f_k - c| \\
&\leq \boldsymbol{c_k} t^{-\frac{1}{2k}}\ O(t^{-\frac{1}{4k}}) + 2|\mathbb{E} f_k - c|.
\end{aligned}$$

□

*Remark:* This shows that the Poisson limit theorem holds for $f_k$ in the sparse regime, since $t \to \infty$.

# Bibliography


[AR20] G. Akinwande and M. Reitzner. Multivariate central limit theorems for random simplicial complexes. *Advances in Applied Mathematics*, 121(102076), 2020.

[BP14] S. Bourguin and G. Peccati. Portmanteau inequalities on the Poisson space: mixed limits and multidimensional clustering. *Electron. J. Probab.*, 19(66):1–42, 2014.

[BP16] S. Bourguin and G. Peccati. The Malliavin-Stein method on the Poisson space. *Springer*, 2016.

[Car09] G. Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46:255–308, 2009.

[DFRV14] L. Decreusefond, E. Ferraz, H. Randriam, and A. Vergne. Simplicial homology of random configurations. *Adv. in Appl. Probab.*, 46:1–20, 2014.

[dSG07] V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.*, 7:339–358, 2007.

[ER59] P. Erdös and A. Rényi. On random graphs. I. *Publicationes Mathematicae*, 6:290–297, 1959.

[Ghr08] R. Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)*, 45:61–75, 2008.

[Gil61] E. N. Gilbert. Random plane networks. *J. Soc. Indust. Appl. Math.*, 9, 1961.

[Gry19] Jens Grygierek. Poisson fluctuations for edge counts in high-dimensional random geometric graphs. *arXiv 1905.11221 math.PR*, 2019.

[Las16] G. Last. Stochastic analysis for Poisson processes. *Bocconi University Press*, 2016.

[Lau09] M. Laurent. Sums of squares, moment matrices and optimization over polynomials in: M. Putinar, S. Sullivant (eds.). Emerging applications of algebraic geometry. *IMA Vol. Math. Appl.*, 149:157–270, 2009.

[LNS16] G. Last, F. Nestmann, and M. Schulte. The random connection model and functions of edge-marked Poisson processes: second order properties and normal approximation. *arXiv:1808.01203*, 2016.

[LP11] G. Last and M. D. Penrose. Poisson process Fock space representation, chaos expansion and covariance inequalities. *Probab. Theory Related Fields*, 150:663–690, 2011.

[LP18] G. Last and M. D. Penrose. Lectures on the Poisson process. In *Stochastic Geometry*, Lectures Notes in Mathematics. Institute of Mathematical Statistics Textbooks, Cambridge University Press, Cambridge, 2018.

[LPS16] G. Last, G. Peccati, and M. Schulte. Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Probab. Theory Related Fields*, 165(3-4):667–723, 2016.

[LRP13a] R. Lachiéze-Rey and G. Peccati. Fine Gaussian fluctuations on the Poisson space, i: contractions, cumulants and geometric random graphs. *Electron. J. Probab.*, 18(32), 2013.

[LRP13b] R. Lachiéze-Rey and G. Peccati. Fine Gaussian fluctuations on the Poisson space, ii: rescaled kernels, marked processes and geometric U-statistics. *Stochastic Process. Appl.*, 123:4186–4218, 2013.

[LRSY19] R. Lachiéze-Rey, M. Schulte, and J. E. Yukich5. Normal approximation for stabilizing functionals. *Ann. Appl. Probab.*, 29:931–993, 2019.

[Pe16] G. Peccati and M. Reitzner (eds.). Stochastic analysis for Poisson point processes. *Bocconi University Press*, 2016.

[Pen03] M.D. Penrose. Random geometric graphs. *Oxford University Press, Oxford*, 2003.

[PN12] G. Peccati and I. Nourdin. Normal approximations with Malliavin calculus. from Stein's method to universality. *Cambridge University Press, Cambridge*, 2012.

[PSTU10] G. Peccati, J. L. Solé, M. S. Taqqu, and F. Utzet. Stein's method and normal approximation of Poisson functionals. *Ann. Probab.*, 38:443–478, 2010.

[PT11] G. Peccati and M. S. Taqqu. Wiener chaos: moments, cumulants and diagrams: a survey with computer implementation. *Bocconi University Press*, 2011.

[PZ10] G. Peccati and C. Zheng. Multi-dimensional Gaussian fluctuations on the Poisson space. *Electron. J. Probab.*, 15:1487–1527, 2010.

[RS13] M. Reitzner and M. Schulte. Central limit theorems for U-statistic of Poisson point processes. *Ann. Probab.*, 41:3879–3909, 2013.

[RST17] M. Reitzner, M. Schulte, and C. Thäle. Limit theory for the Gilbert graph. *Adv. in Appl. Math.*, 88:26–61, 2017.

[Sch13] M. Schulte. Malliavin-Stein method in stochastic geometry. *PhD thesis, Universität Osnabrück*, 2013.

[Sch16] M. Schulte. Normal approximation of Poisson functionals in Kolmogorov distance. *J. Theor. Probab.*, 29:96–117, 2016.

[Ste72] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, II: Probability theory:583–602, 1972.

[Ste86] C. Stein. Approximate computation of expectations. *Institute of Mathematical Statistics, Hayward*, 1986.

[Sur84] S. Surgailis. On multiple Poisson stochastic integrals and associated Markov semigroups. *Probab. Math. Statist.*, 3:217–239, 1984.

[SW08] R. Schneider and Weil W. Stochastic and integral geometry. *Springer, Berlin*, 2008.

[SY19] M. Schulte and J. E. Yukich. Multivariate second order Poincaré inequalities for Poisson functionals. *Electron. J. Probab.*, 24(130), 2019.