ALEA, Lat. Am. J. Probab. Math. Stat. **17**, 675–709 (2020)
DOI: 10.30757/ALEA.v17-26

Poisson and Gaussian fluctuations for the components of the f-vector of high-dimensional random simplicial complexes

Jens Grygierek

Institute of Mathematics
Osnabrück University
Albrechtstraße 28a
D-49076 Osnabrück, Germany.
E-mail address: jgrygierek@uos.de

Abstract. We investigate the high-dimensional asymptotic distributional behavior of the components of the **f**-vector of a random Vietoris-Rips complex that is generated over a Poisson point process in $[-\frac{1}{2}, \frac{1}{2}]^d$ as the space dimension and the intensity tend to infinity while the radius parameter tends to zero simultaneously.

1. Introduction

The field of topological data analysis motivates the study of random simplicial complexes, especially random geometric complexes that are higher-dimensional generalizations of the well known random geometric graph. Naturally, one builds a simplicial complex on data points to study features of the data using combinatorial or topological properties like the **f**-vector, that counts the number of k-dimensional simplices, the Betti-numbers or persistent homology. For a recent introduction into the different opportunities in this research field we refer to the survey article Bobrowski and Kahle (2018).

Let η_d be a Poisson point process on $W := [-\frac{1}{2}, +\frac{1}{2}]^d \subset \mathbb{R}^d$ with dimension dependent intensity $t_d \in (0, \infty)$, i.e. the intensity measure is given by $\mu_d = t_d \Lambda_d$ where Λ_d denotes the d-dimensional Lebesgue measure. We choose a dimension-dependent distance parameter $\delta_d \in (0, \frac{1}{4})$ with $\delta_d \to 0$ for $d \to \infty$.

The points charged by η_d are taken as the vertices of the random Vietoris-Rips complex $\mathrm{VR}^\infty(\eta_d, \delta_d)$, that contains any k-dimensional simplex $\{x_0, \ldots, x_k\} \subseteq \eta_d$, $k \in \mathbb{N}_0$, if and only if the pairwise uniform distances of its vertices are bounded by

---

Received by the editors December 8th, 2019; accepted September 13th, 2020.

2010 *Mathematics Subject Classification.* 60D05, 60F05, 55U10.

Key words and phrases. Poisson limit theorem, central limit theorem, high dimensional random Vietoris-Rips complex, Poisson point process, second-order Poincaré inequality, stochastic geometry, phase transition.

δ_d, i.e.

$$\{x_0, \dots, x_k\} \in \mathrm{VR}^\infty(\eta_d, \delta_d) :\Leftrightarrow \|x_j - x_i\|_\infty \le \delta_d \quad \text{for all} \quad i, j \in \{0, \dots, k\}.$$

The collection of all 1-dimensional simplices coincides with the edges of the well known random geometric graph, where the points of η_d are taken as the vertices and any two vertices are connected by an edge whenever their uniform distance is less than or equal to δ_d, see Penrose (2003, Chapter 3) for more details. We note that the Vietoris-Rips complex $\mathrm{VR}^\infty(\eta_d, \delta_d)$ coincides with the Čech complex whose k-simplices are all subsets $\{x_0, \dots, x_k\}$ admitting a point $y \in \mathbb{R}^d$ with $\|x_i - y\|_\infty \le \frac{\delta_d}{2}$, since we are using the uniform distance. To simplify our notation we will mostly omit the index d in the following. Nevertheless all conditions we impose on the parameter sequences $t := (t_d)_d$ and $\delta := (\delta_d)_d$ in the following have to be treated with respect to $d \to \infty$.

Let $F_k := F_k(\mathrm{VR}^\infty(\eta_d, \delta_d))$, $k \ge 1$, denote the number of k-simplices in the random Vietoris-Rips complex $\mathrm{VR}^\infty(\eta_d, \delta_d)$ that is the U-statistic of order $k+1$ given by

$$F_k(\mathrm{VR}^\infty(\eta_d, \delta_d)) := \frac{1}{(k+1)!} \sum_{(y_0,\dots,y_k)\in\eta_{\neq}^{k+1}} \prod_{i=0}^{k} \prod_{j=i+1}^{k} \mathbb{1}\{\|y_j - y_i\|_\infty \le \delta\},$$

where $\|\cdot\|_\infty$ denotes the uniform norm on $\mathbb{R}^d$ and $\eta_{\neq}^{k+1}$ denotes the set of all $(k+1)$-tuples $(y_0, \dots, y_k) \in \eta^{k+1}$ such that $y_i \ne y_j$ for all $i \ne j$.

Note that F_k is the k-th component of the $\mathbf{f}$-vector of $\mathrm{VR}^\infty(\eta_d, \delta_d)$, i.e.

$$F_k = f_k(\mathrm{VR}^\infty(\eta_d, \delta_d)).$$

Additionally, F_k counts the complete sub-graphs with $k+1$ vertices in the random geometric graph with respect to the uniform distance $\mathrm{d}_\infty(x, y) = \|x - y\|_\infty$. We investigate the asymptotic distributional behavior of F_k as $\delta \to 0$ and the intensity t as well as the space dimension d tend to infinity simultaneously.

Remark 1.1. We use the uniform distance for our model since the Euclidean distance would require the exact calculation of the volume of the intersection of multiple d-dimensional Euclidean balls to achieve sufficiently sharp bounds on the variance. Up to our knowledge, no general formula for this problem is known except for the special case of two d-dimensional Euclidean balls that occurs in the case $k = 1$, see Li (2011). This case was treated in our previous work for a slightly different model, concerning edges that have their midpoints in the d-dimensional Euclidean unit ball, see Grygierek and Thäle (2020) and Grygierek (2019).

1.1. *Main results.* As a preparation for our limit theorems we show asymptotically sharp bounds for the expectation and the variance of our k-simplex counting functional:

Lemma 1.2. *For all $d \ge 1$ and $k \ge 1$ the expected number of k-simplices in the random Vietoris-Rips complex $\mathrm{VR}^\infty(\eta_d, \delta_d)$ is bounded by*

$$(1 - 2\delta)^d \frac{t\big(t\delta^d\big)^k (k+1)^d}{(k+1)!} \le \mathbb{E}[F_k] \le \frac{t\big(t\delta^d\big)^k (k+1)^d}{(k+1)!}.$$

Lemma 1.3. *For all $d \geq 1$ and $k \geq 1$ there exist explicit constants $\mathbf{C}(k,r) \in (0,\infty)$ only depending on k and r such that the variance of the number of k-simplices in the random Vietoris-Rips complex $\mathrm{VR}^{\infty}(\eta_d, \delta_d)$ is bounded by*

$$\mathbb{V}[F_k] \geq \mathbb{E}[F_k] + (1-2\delta)^d t(t\delta^d)^k \sum_{r=1}^{k} \mathbf{C}(k,r)(t\delta^d)^{k+1-r}\Big(\tfrac{2(k+2)(k+1-r)}{r+1} + r\Big)^d,$$

$$\mathbb{V}[F_k] \leq \mathbb{E}[F_k] + t(t\delta^d)^k \sum_{r=1}^{k} \mathbf{C}(k,r)(t\delta^d)^{k+1-r}\Big(\tfrac{2(k+2)(k+1-r)}{r+1} + r\Big)^d.$$

To ensure that the lower and upper bound for the expectation and variance tend to the same limit, we assume that $(\delta_d)_d$ is decreasing sufficiently fast, i.e. we assume

$$\lim_{d\to\infty} d\delta_d = 0,$$

see Remark 4.2 for more details.

The asymptotic behavior of F_k depends on how fast the sequence $(t_d)_d$ increases as $d \to \infty$. This phenomenon is quite common for asymptotic results related to edge counts in fixed dimension and was also shown for edge-counts in high-dimensional random geometric graphs in our previous work Grygierek and Thäle (2020); Grygierek (2019) considering a slightly different model.

In particular, one has to distinguish the following phases that are determined by the limit of the expectation $\mathbb{E}[F_k]$:

$$\lim_{d\to\infty} \frac{1}{(k+1)!} t(t\delta^d)^k (k+1)^d = \infty, \tag{1.1}$$

$$\lim_{d\to\infty} \frac{1}{(k+1)!} t(t\delta^d)^k (k+1)^d = \theta \in (0,\infty), \tag{1.2}$$

$$\lim_{d\to\infty} \frac{1}{(k+1)!} t(t\delta^d)^k (k+1)^d = 0. \tag{1.3}$$

The rate of convergence in the following central limit theorem and Poisson limit theorem will be measured by the Wasserstein distance $\mathrm{d}_W(\cdot,\cdot)$ resp. the total variation distance $\mathrm{d}_{TV}(\cdot,\cdot)$, see Section 2.2 below for a formal definition. We indicate convergence in distribution by writing $\xrightarrow{D}$.

If the expectation tends to infinity (1.1) the k-simplex counting functional satisfies a central limit theorem:

Theorem 1.4 (Gaussian Approximation)**.** *For $k \geq 1$ fixed, we assume $\mathbb{E}[F_k] \to \infty$ for $d \to \infty$. Let $\mathcal{N}(0,1)$ be a standard Gaussian distributed random variable and denote by $\widetilde{F_k} := \frac{F_k - \mathbb{E}[F_k]}{\sqrt{\mathbb{V}[F_k]}}$ the standardized version of F_k.*
If $(t\delta^d) \to 0$ for $d \to \infty$, then

$$\mathrm{d}_W(\widetilde{F_k}, \mathcal{N}(0,1)) = \begin{cases} \mathcal{O}\Big((\mathbb{E}[F_k])^{-\frac{1}{2}}(k+1)^{\frac{3d}{2}} 2^d\Big), & k \leq 3, \\ \mathcal{O}\Big((\mathbb{E}[F_k])^{-\frac{1}{2}}(k+1)^{2d}\Big), & k \geq 3. \end{cases}$$

If $(t\delta^d) \to c \in (0,\infty)$ or $(t\delta^d) \to \infty$ for $d \to \infty$, then

$$\mathrm{d}_W(\widetilde{F_k}, \mathcal{N}(0,1)) = \mathcal{O}\Big(t^{-\frac{1}{2}}\Big(1 + \tfrac{1}{k^2+2k}\Big)^d 2^d\Big).$$

In particular, if $\mathbb{E}[F_k]$ *resp.* t *is increasing sufficiently fast depending on* d *one has that*

$$\widetilde{F_k} \xrightarrow{D} \mathcal{N}(0,1), \quad \textit{as} \quad d \to \infty.$$

If the expectation tends to a finite positive limit (1.2) the k-simplex counting functional satisfies a Poisson limit theorem:

Theorem 1.5 (Poisson Approximation)**.** *For* $k \geq 1$ *fixed, we assume* $\mathbb{E}[F_k] \to \theta \in (0,\infty)$ *for* $d \to \infty$*. Let* $\mathcal{P}(\theta)$ *be a Poisson distributed random variable with expectation and variance* θ*. Then*

$$\begin{aligned} \mathrm{d}_{TV}(F_k, \mathcal{P}(\theta)) &= \mathcal{O}(|\mathbb{E}[F_k] - \theta|) + \mathcal{O}(|\mathbb{V}[F_k] - \theta|) \\ &\quad + \begin{cases} \mathcal{O}\Big(t^{-\frac{1}{2k}}(k+1)^{\frac{d(3k-1)}{2k}} 2^d\Big), & k \leq 3, \\ \mathcal{O}\Big(t^{-\frac{1}{2k}}(k+1)^{\frac{d(4k-1)}{2k}}\Big), & k \geq 3. \end{cases} \end{aligned}$$

In particular, if t *is increasing sufficiently fast depending on* d *one has that*

$$F_k \xrightarrow{D} \mathcal{P}(\theta), \quad \textit{as} \quad d \to \infty.$$

Proposition 1.6. *If the expectation tends to zero* (1.3) *we also have* $\mathbb{V}[F_k] \to 0$*, indicating that the* k*-simplex counting functional vanishes in the limit, since the random Vietoris-Rips complex contains almost surely no* k*-simplices.*

This paper is organized as follows. For the convenience of the reader, we repeat the relevant material on the Malliavin-Stein method for normal approximation and Poisson approximation of Poisson functionals in Section 2.2. Additionally, we give a short introduction to simplicial complexes in Section 2.3. In Section 3 we present a decomposition technique for U-statistics that will be used in the proof of our main results, that are given in Section 4: We start with the expectation and variance bounds, Lemmas 1.2 and 1.3 in Subsection 4.1. In Subsection 4.2 we prepare bounds on the first and second order Malliavin derivatives, that will finally be used in Subsection 4.3 to obtain the central limit theorems, Theorem 1.4, and the Poisson limit theorem, Theorem 1.5.

2. Preliminaries

The d-dimensional Euclidean space is denoted by $\mathbb{R}^d$ and we let $\mathscr{B}^d$ be the Borel σ-field on $\mathbb{R}^d$. The Lebesgue measure on $\mathbb{R}^d$ is indicated by Λ_d. The d-dimensional closed $\mathcal{L}_\infty$-ball with respect to the uniform norm, center in $z \in \mathbb{R}^d$ and radius $r > 0$ is defined by

$$\overline{\mathbb{B}}^d_\infty(z,r) := \big\{x \in \mathbb{R}^d : \|x - z\|_\infty \leq r\big\}.$$

2.1. *Poisson functionals and difference operators.* Let N_σ denote the space of all σ-finite counting measures χ on $\mathbb{R}^d$, i.e. $\chi(B) \in \mathbb{N}_0 \cup \{\infty\}$ for all $B \in \mathscr{B}^d$. We equip the space N_σ with the σ-field $\mathscr{N}_\sigma$ generated by the mappings $\chi \to \chi(B)$, $B \in \mathscr{B}^d$.

Definition 2.1. A Poisson point process η with (non-atomic) intensity measure μ is a random counting measure on $\mathbb{R}^d$, that is a random element in the space N_σ, that satisfies the following properties:

(1) For all $B \in \mathscr{B}^d$ and all $k \in \mathbb{N}_0$ it holds that $\eta(B)$ is a Poisson distributed random variable with expectation $\mu(B)$, i.e.

$$\mathbb{P}(\eta(B) = k) = \frac{\mu(B)^k}{k!} e^{-\mu(B)},$$

where we set $\frac{\infty^k}{k!} e^{-\infty} = 0$ for all k if $\mu(B) = \infty$.

(2) For all $m \in \mathbb{N}_0$ and all pairwise disjoint measurable sets $B_0, \dots, B_m \in \mathscr{B}^d$, the random variables $\eta(B_0), \dots, \eta(B_m)$ are independent.

To simplify our notation we will often handle η as a random set of points using

$$x \in \eta \Leftrightarrow x \in \big\{y \in \mathbb{R}^d : \eta(\{y\}) > 0\big\}.$$

It is well known that such a Poisson point process η satisfies the following multivariate Mecke formula, see Last and Penrose (2018, Theorem 4.4).

Lemma 2.2. *For all $m \in \mathbb{N}^*$ and all non-negative measurable functions $h : (\mathbb{R}^d)^m \times \mathrm{N}_\sigma \to \mathbb{R}$ it holds that*

$$\begin{aligned} &\mathbb{E} \sum_{(y_1,\dots,y_m) \in \eta^m_{\neq}} h(y_1, \dots, y_m; \eta) \\ &\quad = \int\limits_{(\mathbb{R}^d)^m} \mathbb{E}[h(y_1, \dots, y_m; \eta + \delta_{y_1} + \dots + \delta_{y_m})] \mathrm{d}\mu^m(y_1, \dots, y_m), \end{aligned} \tag{2.1}$$

where $\eta^m_{\neq}$ is the collection of m-tuples of pairwise distinct points charged by η.

We call a random variable F a Poisson functional, if there exists a measurable map $f : \mathrm{N}_\sigma \to \mathbb{R}$ such that $F = f(\eta)$ almost surely. The map f is called the representative of F. We define the difference operator or so-called "add-one-cost operator":

Definition 2.3. Let F be a Poisson functional and f its corresponding representative, then the first order difference operator $D_x F : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$D_x F := f(\eta + \delta_x) - f(\eta), \quad x \in \mathbb{R}^d,$$

where δ_x denotes the Dirac measure with mass concentrated in $x \in \mathbb{R}^d$. We say that F belongs to the domain of the difference operator, i.e. $F \in \mathrm{dom}(D)$, if $\mathbb{E}\big[F^2\big] < \infty$ and

$$\int\limits_{\mathbb{R}^d} \mathbb{E}\big[(D_x F)^2\big]\, \mu(\mathrm{d}x) < \infty.$$

The second order difference operator is obtained through iteration:

$$\begin{aligned} D^2_{x_1,x_2} F &:= D_{x_1}(D_{x_2} F) \\ &= f(\eta + \delta_{x_1} + \delta_{x_2}) - f(\eta + \delta_{x_1}) - f(\eta + \delta_{x_2}) + f(\eta), \quad x_1, x_2 \in \mathbb{R}^d. \end{aligned}$$

For a deeper discussion of the underlying theory of Poisson point processes, Malliavin-Calculus, the Wiener-Itô chaos expansion and the Malliavin-Stein method presented below, see Peccati and Reitzner (2016) and Last and Penrose (2018).

2.2. *Malliavin-Stein method.* We will use the Wasserstein-distance for the normal approximation and the total variation distance for the Poisson approximation, see for instance Bourguin and Peccati (2016, Section 2.1).

Definition 2.4. We denote by Lip(1) the class of Lipschitz functions $h : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant less or equal to one, i.e. h is absolutely continuous and almost everywhere differentiable with $\|h'\|_\infty \leq 1$. Given two $\mathbb{R}$-valued random variables X, Y, with $\mathbb{E}|X| < \infty$ and $\mathbb{E}|Y| < \infty$ the Wasserstein distance between the laws of X and Y, written $\mathrm{d}_W(X, Y)$, is defined as

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

Definition 2.5. Given two $\mathbb{N}_0$-valued random variables X, Y, the total variation distance between the laws of X and Y, written $\mathrm{d}_{TV}(X, Y)$, is defined as

$$\mathrm{d}_{TV}(X, Y) := \sup_{A \subseteq \mathbb{N}_0} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

Note that the topologies induced by the metrics d_W and d_{TV} are strictly finer than the one induced by convergence in distribution. Therefore, if a sequence $(X_n)_n$ of random variables satisfies $\lim_{n\to\infty} \mathrm{d}_W(X_n, Y) = 0$ resp. $\lim_{n\to\infty} \mathrm{d}_{TV}(X_n, Y) = 0$ for a random variable Y then it holds that X_n converges to Y in distribution, i.e. $X_n \overset{D}{\to} Y$.

We rephrase a version of the main result from Last et al. (2016), a so-called second order Poincaré inequality for Poisson functionals, see also Last and Penrose (2018, Theorem 2.13), it is the main device in our proof of Theorem 1.4.

Theorem 2.6. *Let $F \in \mathrm{dom}(D)$ be a Poisson functional such that $\mathbb{E}[F] = 0$ and $\mathbb{V}[F] = 1$. Define*

$$\gamma_1(F) := \int_{W^3} \Big(\mathbb{E}\big[(D^2_{x_1,x_3}F)^4\big]\mathbb{E}\big[(D^2_{x_2,x_3}F)^4\big] \times \mathbb{E}\big[(D_{x_1}F)^4\big]\mathbb{E}\big[(D_{x_2}F)^4\big]\Big)^{\frac{1}{4}} \mu^3(\mathrm{d}(x_1, x_2, x_3))$$

$$\gamma_2(F) := \int_{W^3} \Big(\mathbb{E}\big[(D^2_{x_1,x_3}F)^4\big]\mathbb{E}\big[(D^2_{x_2,x_3}F)^4\big]\Big)^{\frac{1}{2}} \mu^3(\mathrm{d}(x_1, x_2, x_3))$$

$$\gamma_{3,N}(F) := \int_W \mathbb{E}|D_x F|^3 \mu(\mathrm{d}x)$$

and let Z be a standard Gaussian random variable, then

$$\mathrm{d}_W(F, Z) \leq 2\sqrt{\gamma_1(F)} + \sqrt{\gamma_2(F)} + \gamma_{3,N}(F), \tag{2.2}$$

where d_W denotes the Wasserstein-distance.

In the proof of Theorem 1.5 we use the analogue of Theorem 2.6 for Poisson approximation from Grygierek (2019):

Theorem 2.7. *Let $F \in \mathrm{dom}(D)$ be an $\mathbb{N}_0$-valued Poisson functional. Define,*

$$\gamma_{3,P}(F) := \int_W \Big(\mathbb{E}|D_x F(D_x F - 1)|^2\Big)^{\frac{1}{2}} \Big(\mathbb{E}|D_x F|^2\Big)^{\frac{1}{2}} \mu(\mathrm{d}x),$$

and let $\mathcal{P}(\theta)$ *be a Poisson distributed random variable with expectation and variance* $\theta > 0$*. Then*

$$\mathrm{d}_{TV}(F, \mathcal{P}(\theta)) \leq \frac{1-e^{-\theta}}{\theta}\Big(2\sqrt{\gamma_1(F)} + \sqrt{\gamma_2(F)} + \frac{\gamma_{3,P}(F)}{\theta} + |\mathbb{E}[F]-\theta| + |\mathbb{V}[F]-\theta|\Big), \tag{2.3}$$

where d_{TV} *denotes the total variation distance.*

2.3. *Simplicial Complexes.* An (abstract) simplicial complex is a collection Δ of subsets of a set V that is closed under taking subsets, i.e.

$$\forall \mathfrak{F} \in \Delta : \forall \mathfrak{L} \subset \mathfrak{F} : \mathfrak{L} \in \Delta.$$

The set V is called the vertex set of Δ, where we assume that $\{v\} \in \Delta$ for all $v \in V$.

The elements of Δ are called faces or simplices. Additionally, every subset $\mathfrak{L} \subseteq \mathfrak{F}$ of a face $\mathfrak{F} \in \Delta$ is called a face of $\mathfrak{F}$, thus the faces of the faces of Δ are faces of Δ themselves.

The dimension of a face $\mathfrak{F} \in \Delta$ is given as the number of vertices of the face minus one, i.e. $\dim(\mathfrak{F}) = |\mathrm{vert}(\mathfrak{F})| - 1$ and the number of i-dimensional faces of Δ will be denoted by $f_i(\Delta)$. We note that the vector $(f_{-1}(\Delta), f_0(\Delta), f_1(\Delta), \ldots)$ is the $\mathbf{f}$-vector of Δ, where $f_{-1}(\Delta)$ is the Euler characteristics of Δ and $f_0(\Delta)$ denotes the number of vertices of Δ, see Ziegler (1995, Definition 8.16, p.245) for more details.

The Vietoris-Rips complex is an example of a simplicial complex that arises naturally from metric spaces.

Definition 2.8 (Vietoris-Rips complex)**.** Let $X = (X, \mathrm{d})$ be a metric space (usually a locally finite subset of $\mathbb{R}^d$) and $\delta \in (0, \infty)$. The Vietoris-Rips complex of (X, d) with respect to δ (and the underlying metric d) is the abstract simplicial complex on vertex set X whose k-simplices are all subsets $\{x_0, \ldots, x_k\} \subseteq X$ with $\mathrm{d}(x_i, x_j) \leq \delta$ for all $i, j \in \{0, \ldots, k\}$.

For more details on simplicial complexes, we refer the reader to the books Stanley (1996); Munkres (1984); Ziegler (1995).

3. Moment-Decomposition for U-statistics

Let η be a simple Poisson point process on $\mathbb{X}$ with intensity measure μ. For $n \geq 1$ we consider the U-statistics F of order n with symmetric and measurable kernel $h : \mathbb{X}^n \to \mathbb{R}$ given by

$$F := \frac{1}{n!} \sum_{(y_0,\ldots,y_{n-1}) \in \eta^n_{\neq}} h(y_0, \ldots, y_{n-1}). \tag{3.1}$$

To prove our main results, we will introduce the following helpful decomposition of the p-th power F^p for $p \in \{2, 3, 4\}$ that allows us to apply Mecke's formula (2.1) to each term in the decomposition and derive the corresponding moments of F.

The main idea is to split the summation over $(\eta^n_{\neq})^p$ into multiple sums over sets that are diagonal free and therefore satisfy the pairwise distinct condition needed for the index set in Mecke's formula (2.1). Identifying variables in the tuples that are assumed to be equal and accounting for all possible combinations and permutations we use the symmetry of h to reduce the different cases to the terms given in the lemmas below.

Note that the constants $\mathbf{C}(n,\cdot) \in (0,\infty)$ are combinatorially constants that do only depend on n and the given indices of the corresponding sum.

Notation 3.1. *To shorten our notation we will use $y_{[m,n]}$ instead of $y_m, \ldots, y_{m+n-1}$ and $y_{[n]}$ instead of $y_0, \ldots, y_{n-1}$. Further $y_{[n,0]}$ resp. $y_{[0]}$ indicates that no y-variables are used.*

Lemma 3.2. *For all $n \geq 1$ there exist explicit constants $\mathbf{C}(n,r) \in (0,\infty)$ only depending on n and r, such that the second moment of F is given by*

$$\mathbb{E}\big[F^2\big] = \sum_{r=0}^{n} \mathbf{C}(n,r) \int\limits_{\mathbb{X}^{2n-r}} h(y_{[n]})h(y_{[r]}, z_{[n-r]})\mathrm{d}\mu^n(y_{[n]})\mathrm{d}\mu^{n-r}(z_{[n-r]}),$$

where r is the number of variables that are shared in both kernel functions in the integral.

Note that $\mathbf{C}(n,0) = \frac{1}{n!^2}$ and that the corresponding integral for $r = 0$ equals $(n!\mathbb{E}[F])^2$, which directly yields the following representation for the variance:

Corollary 3.3. *For all $n \geq 1$ there exist explicit constants $\mathbf{C}(n,r) \in (0,\infty)$ only depending on n and r, such that the variance of F is given by*

$$\mathbb{V}[F] = \sum_{r=1}^{n} \mathbf{C}(n,r) \int\limits_{\mathbb{X}^{2n-r}} h(y_{[n]})h(y_{[r]}, z_{[n-r]})\mathrm{d}\mu^n(y_{[n]})\mu^{n-r}(z_{[n-r]}).$$

For the third and fourth moment we derive similar representations involving the product of three resp. four kernel functions.

Notation 3.4. *To shorten our Notation we denote by $\min\binom{a}{b}$ the minimum of a and b.*

Lemma 3.5. *For all $n \geq 1$ there exist explicit constants $\mathbf{C}(n,\cdot) \in (0,\infty)$ only depending on the given values, such that the third moment of F is given by*

$$\mathbb{E}\big[F^3\big] = \sum_{r=0}^{n} \sum_{s_Y=0}^{n} \sum_{s_Z=0}^{\min\binom{n-r}{n-s_Y}} \mathbf{C}(n,r,s_Y,s_Z)$$
$$\int\limits_{\mathbb{X}^{3n-r-s}} h(y_{[n]})h(y_{[r]}, z_{[n-r]})h(y_{[s_Y]}, z_{[s_Z]}, w_{[n-s]})\mathrm{d}\mu^n(y_{[n]})\mathrm{d}\mu^{n-r}(z_{[n-r]})\mathrm{d}\mu^{n-s}(w_{[n-s]}),$$

where $s := s_Y + s_Z$ and the indices of the sums are denoting the number of variables that are shared in multiple kernel functions in the integral.

Lemma 3.6. *For all $n \geq 1$ there exist explicit constants $\mathbf{C}(n,\cdot) \in (0,\infty)$ only depending on the given values, such that the fourth moment of F is given by*

$$\mathbb{E}\big[F^4\big] = \sum_{r=0}^{n} \sum_{s_Y=0}^{n} \sum_{s_Z=0}^{\min\binom{n-r}{n-s_Y}} \sum_{m_Y=0}^{n} \sum_{m_Z=0}^{\min\binom{n-r}{n-m_Y}} \sum_{m_W=0}^{\min\binom{n-s}{n-m_Y-m_Z}} \mathbf{C}(n,r,s_Y,s_Z,m_Y,m_Z,m_W)$$
$$\int\limits_{\mathbb{X}^{4n-r-s-m}} h(y_{[n]})h(y_{[r]}, z_{[n-r]})h(y_{[s_Y]}, z_{[s_Z]}, w_{[n-s]})h(y_{[m_Y]}, z_{[m_Z]}, w_{[m_W]}, u_{[n-m]})$$
$$\mathrm{d}\mu^n(y_{[n]})\mathrm{d}\mu^{n-r}(z_{[n-r]})\mathrm{d}\mu^{n-s}(w_{[n-s]})\mathrm{d}\mu^{n-m}(u_{[n-m]}),$$

where $s := s_Y + s_Z$, $m := m_Y + m_Z + m_W$ and the indices of the sums are denoting the number of variables that are shared in multiple kernel functions in the integral.

Proof: Moment-Decomposition for U-statistics: We denote the sets of n variables enumerated from 0 to $n-1$ with capital letters

$$Y(n) = \{y_0, \ldots, y_{n-1}\}, \qquad W(n) = \{w_0, \ldots, w_{n-1}\},$$
$$Z(n) = \{z_0, \ldots, z_{n-1}\}, \qquad U(n) = \{u_0, \ldots, u_{n-1}\},$$

and for $X_1, X_2 \subseteq Y(n) \cup Z(n) \cup W(n) \cup U(n)$ we denote by $\Psi(X_1, X_2)$ the set of all injective maps $\psi : X_1 \to X_1 \cup X_2$ such that $\psi(x) = x$ if $x \in X_1$ or $\psi(x) \in X_2$, i.e. it is not allowed to map an element of X_1 onto another element of X_1 but it is allowed to injectively map any element of X_1 to any element of X_2.

Note that we will choose X_1, X_2 such that every map $\psi \in \Psi(X_1, X_2)$ will represent one possible way to choose the variables in the index set of the summation of the corresponding p-th power of F. The case $\psi(x) = x$ represents the case that the variable x is not equal to any other variable in X_2 and therefore gets mapped onto itself and the case $\psi(x) \in X_2$ represents the case where the variables x and $\psi(x)$ are equal, representing a diagonal in the Cartesian product of the index sets of the U-statistic. Further we will use the symmetry of the product and of h to define an equivalence relation on $\Psi(X_1, X_2)$ such that all elements belonging to the same equivalence class yield the same value in the decomposition.

$\mathbf{p = 1}$: Note that the expectation of F can be obtained directly using Mecke's formula (2.1), since the index set already consists of pairwise distinct tuples of points, i.e.

$$\mathbb{E}[F] = \frac{1}{n!}\mathbb{E} \sum_{(y_0,\ldots,y_{n-1}) \in \eta^n_{\neq}} h(y_0, \ldots, y_{n-1})$$
$$= \frac{1}{n!} \int_{\mathbb{X}^n} h(y_0, \ldots, y_{n-1}) \mathrm{d}\mu^n(y_0, \ldots, y_{n-1}).$$

$\mathbf{p = 2}$: For the second moment of F we rewrite the product of the sums as the sums of products renaming the variables in the second factor to obtain:

$$F^2 = \frac{1}{n!^2} \sum_{(y_0,\ldots,y_{n-1}) \in \eta^n_{\neq}} \sum_{(z_0,\ldots,z_{n-1}) \in \eta^n_{\neq}} h(y_0, \ldots, y_{n-1}) h(z_0, \ldots, z_{n-1})$$
$$= \frac{1}{n!^2} \sum_{(y_0,\ldots,y_{n-1},z_0,\ldots,z_{n-1}) \in \eta^n_{\neq} \times \eta^n_{\neq}} h(y_0, \ldots, y_{n-1}) h(z_0, \ldots, z_{n-1}).$$

The index set $\eta^n_{\neq} \times \eta^n_{\neq}$ contains tuples that allow non distinct choices of variables, i.e. $y_0 = z_0$. Therefore we need to decompose the index set into all possible combinations with respect to pairwise distinct choices to apply Mecke's formula (2.1). This yields index sets of the form $\eta^{2n-r}_{\neq}$, where $r = 0, \ldots, n$ denotes the number of variables z that are equal to another variable y. To account for all possible choices in this decomposition we consider the injective maps $\psi \in \Psi(Z(n), Y(n))$. We denote the number of reused variables from the set $Y(n)$ by $r := r_Y(\psi) := |\mathrm{Im}\,\psi \cap Y(n)|$, where $\operatorname{Im}\psi$ denotes the image of ψ. It follows that

$$F^2 = \frac{1}{n!^2} \sum_{\psi\in\Psi} \sum_{(y_0,\dots,y_{n-1},\operatorname{Im}\psi\cap Z(n))\in\eta_{\neq}^{2n-r}} h(y_0,\dots,y_{n-1})h(\psi(z_0),\dots,\psi(z_{n-1})),$$

where the index will use the variables $y_0,\dots,y_{n-1}$ and the variables $\operatorname{Im}\psi\cap Z(n)$ that are not replaced by variables from $Y(n)$. Since h is symmetric, the value of the second sum depends only on the number of reused variables r. Thus we define the equivalence relation $\psi\sim\psi'$ if $r_Y(\psi')=r_Y(\psi)=r$ and obtain $n+1$ equivalence classes $[\psi_r]\in\Psi/\!\sim$ that have to be distinguished. For simplicity we will chose the representative ψ_r such that $\psi(z_i)=y_i$ for all $i=0,\dots,r-1$ indicating that we reuse the first r variables from $Y(n)$ as the first r arguments in the second kernel. We denote the cardinality of the equivalence class $[\psi_r]$ by $|[\psi_r]|$ and observe that $|[\psi_r]|$ only depends on n and r. Thus we define $\mathbf{C}(n,r):=\frac{|[\psi_r]|}{n!^2}$, yielding

$$F^2=\sum_{r=0}^{n}\mathbf{C}(n,r)\sum_{(y_{[n]},z_{[n-r]})\in\eta_{\neq}^{2n-r}} h(y_{[n]})h(y_{[r]},z_{[n-r]}), \tag{3.2}$$

where we renamed the variables from $Z(n)$ that were not replaced to use the shorthand notation $z_{[n-r]}$. Using Mecke's formula (2.1) the claim of Lemma 3.2 is obtained directly.

$\mathbf{p=3}$: The proof of Lemma 3.5 is obtained, using the representation of F^2 given by (3.2) and multiplying by F, i.e.

$$\begin{aligned} F^3 &= \sum_{r=0}^{n}\mathbf{C}(n,r)\sum_{(y_{[n]},z_{[n-r]})\in\eta_{\neq}^{2n-r}} h(y_{[n]})h(y_{[r]},z_{[n-r]}) \\ &\qquad\times\frac{1}{n!}\sum_{(w_0,\dots,w_{n-1})\in\eta_{\neq}^{n}} h(w_0,\dots,w_{n-1}) \\ &= \sum_{r=0}^{n}\frac{\mathbf{C}(n,r)}{n!}\sum_{(y_{[n]},z_{[n-r]},w_0,\dots,w_{n-1})\in\eta_{\neq}^{2n-r}\times\eta_{\neq}^{n}} h(y_{[n]})h(y_{[r]},z_{[n-r]})h(w_0,\dots,w_{n-1}). \end{aligned}$$

For every fixed $r=0,\dots,n$ we have to decompose the index set $\eta_{\neq}^{2n-r}\times\eta_{\neq}^{n}$ into index sets of the form $\eta_{\neq}^{3n-r-s}$ where $s=0,\dots,n$ denotes the number of variables w that are equal to another variable y or z. Here, we consider the injective maps $\psi\in\Psi(W(n),Y(n)\cup Z(n-r))$ and denote the number of reused variables from $Y(n)$ resp. $Z(n-r)$ by $s_Y=s_Y(\psi):=|\operatorname{Im}\psi\cap Y(n)|$ resp. $s_Z:=s_Z(\psi):=|\operatorname{Im}\psi\cap Z(n-r)|$. Thus $s=s_Y+s_Z$ and it follows that

$$\begin{aligned} F^3=\sum_{r=0}^{n}\frac{\mathbf{C}(n,r)}{n!}\sum_{\psi\in\Psi}\sum_{(y_{[n]},z_{[n-r]},\operatorname{Im}\psi\cap W(n))\in\eta_{\neq}^{3n-r-s}} & h(y_{[n]})h(z_{[n-r]}) \\ & \times h(\psi(w_0),\dots,\psi(w_{n-1})), \end{aligned}$$

where the index set will use the variables $y_0,\dots,y_{n-1},z_0,\dots,z_{n-r-1}$ and the variables $\operatorname{Im}\psi\cap W(n)$ that are not replaced by variables from $Y(n)$ or $Z(n-r)$. Again, since h is symmetric, the value of the last sum depends only on the numbers of reused variables s_Y and s_Z. We define the equivalence relation $\psi\sim\psi'$ if and only if $s_Y(\psi')=s_Y(\psi)=s_Y$ and $s_Z(\psi')=s_Z(\psi)=s_Z$ and obtain the equivalence classes

$[\psi_{(s_Y,z_Z)}] \in \Psi/\sim$. For simplicity we will chose the representative $\psi_{(s_Y,s_Z)}$ such that $\psi(w_i) = y_i$ for all $i = 0,\dots,s_Y-1$ and $\psi(w_{j+s_Y}) = z_j$ for all $j = 0,\dots,s_Z-1$. Therefore we will reuse the first s_Y variables from $Y(n)$ followed by the first s_Z variables from $Z(n-r)$, before we plug in the leftover variables from $W(n)$ that where not replaced. The cardinality $|[\psi_{(s_Y,s_Z)}]|$ of the equivalence class $[\psi_{(s_Y,s_Z)}]$ depends only on n, r, s_Y and s_Z. Thus, we define $\mathbf{C}(n,r,s_Y,s_Z) := \frac{\mathbf{C}(n,r)}{n!} \times \left|[\psi_{(s_Y,s_Z)}]\right|$ and obtain

$$\begin{aligned} F^3 = &\sum_{r=0}^{n}\sum_{s_Y=0}^{n}\sum_{s_Z=0}^{\min\binom{n-r}{n-s_Y}} \mathbf{C}(n,r,s_Y,s_Z) \\ &\times \sum_{(y_{[n]},z_{[n-r]},w_{[n-s]})\in\eta_{\neq}^{3n-r-s_Y-s_Z}} h(y_{[n]})h(y_{[r]},z_{[n-r]})h(y_{[s_Y]},z_{[s_Z]},w_{[n-s]}), \end{aligned} \tag{3.3}$$

where we renamed the variables from $W(n)$ that were not equal to a variable from $Y(n)$ or $Z(n)$ to use the shorthand notation $w_{[n-s]}$. Note that the number of variables reused from $Z(n-r)$ can not be greater than the number of variables from $W(n)$ left after the replacements with variables from $Y(n)$ are done, thus $0 \le s_Z \le \min(n-r, n-s_Y)$. Using Mecke's formula (2.1) the claim of Lemma 3.5 is obtained directly.

$\mathbf{p = 4}$:

Lemma 3.6 follows by another iteration of this technique, starting with the representation of F^3 given by (3.3) and multiplying by F, i.e.

$$\begin{aligned} F^4 = &\sum_{r=0}^{n}\sum_{s_Y=0}^{n}\sum_{s_Z=0}^{\min\binom{n-r}{n-s_Y}} \tfrac{\mathbf{C}(n,r,s_Y,s_Z)}{n!} \\ &\times \sum_{(y_{[n]},z_{[n-r]},w_{[n-s]})\in\eta_{\neq}^{3n-r-s_Y-s_Z}\times\eta_{\neq}^{n}} h(y_{[n]})h(y_{[r]},z_{[n-r]})h(y_{[s_Y]},z_{[s_Z]},w_{[n-s]})h(u_0,\dots,u_{n-1}). \end{aligned}$$

For every fixed $r, s_Y = 0,\dots,n$ and $s_Z = 0,\dots,\min(n-r,n-s_Y)$ we have to decompose the index set $\eta_{\neq}^{3n-r-s_Y-s_Z} \times \eta_{\neq}^{n}$ into index sets of the form $\eta_{\neq}^{4n-r-s-m}$ where $m = 0,\dots,n$ denotes the number of variables u that are equal to another variables y, z or w. In this case, we consider the injective maps $\psi \in \Psi(U(n), Y(n) \cup Z(n-r) \cup W(n-s))$ and denote the number of reused variables from $Y(n)$, $Z(n-r)$ resp. $W(n-r-s)$ by $m_Y := m_Y(\psi) := |\operatorname{Im}\psi \cap Y(n)|$, $m_Z := m_Z(\psi) := |\operatorname{Im}\psi \cap Z(n-r)|$ resp. $m_W := m_W(\psi) := |\operatorname{Im}\psi \cap W(n-s)|$. Thus $m = m_Y + m_Z + m_W$ and it follows that

$$\begin{aligned} F^4 = &\sum_{r=0}^{n}\sum_{s_Y=0}^{n}\sum_{s_Z=0}^{\min\binom{n-r}{n-s_Y}} \tfrac{\mathbf{C}(n,r,s_Y,s_Z)}{n!} \sum_{\psi\in\Psi} \sum_{(y_{[n]},z_{[n-r]},w_{[n-s]},\operatorname{Im}\cap U(n))\in\eta_{\neq}^{4n-r-s-m}} h(y_{[n]})h(y_{[r]},z_{[n-r]}) \\ &\times h(y_{[s_Y]},z_{[s_Z]},w_{[n-s]})h(\psi(u_0),\dots,\psi(u_{n-1})), \end{aligned}$$

where the index set will use the variables $y_0,\dots,y_{n-1}$, $z_0,\dots,z_{n-r-1}$, $w_0,\dots,w_{n-s-1}$ and the variables $\operatorname{Im}\psi \cap U(n)$ that are not replaced by variables from $Y(n)$, $Z(n-r)$ or $W(n-s)$. Again, since h is symmetric, the value of the last sum depends only on the numbers of reused variables m_Y, m_Z and m_W. We define the equivalence relation $\psi \sim \psi'$ if and only if $m_Y(\psi) = m_Y(\psi')$ and $m_Z(\psi) = m_Z(\psi')$ and $m_W(\psi) =$

$m_W(\psi')$ and obtain the equivalence classes $[\psi_{(m_Y,m_Z,m_W)}] \in \Psi/\sim$. For simplicity we will chose the representative $\psi_{(m_Y,m_Z,m_W)}$ such that $\psi(u_i) = y_i$ for all $i = 0,\dots,m_Y-1$, $\psi(u_{j+m_Y}) = z_j$ for all $j = 0,\dots,s_Z-1$ and $\psi(u_{l+m_Y+m_Z}) = w_l$ for all $l = 0,\dots,m_W$. Therefore we will reuse the first m_Y variables from $Y(n)$ followed by the first m_Z variables from $Z(n-r)$ and the first m_W variables from $W(n-s)$, before we plug in the leftover variables from $U(n)$. The cardinality $|[\psi_{m_Y,m_Z,m_W}]|$ of the equivalence class $[\psi_{m_Y,m_Z,m_W}]$ depends only on n, r, s_Y, s_Z, m_Y, m_Z and m_W. Thus we define $\mathbf{C}(n,r,s_Y,s_Z,m_Y,m_Z,m_W) := \frac{\mathbf{C}(n,r,s_Y,s_Z)}{n!} \times |[\psi_{m_Y,m_Z,m_W}]|$ and obtain

$$F^4 = \sum_{r=0}^{n}\sum_{s_Y=0}^{n}\sum_{s_Z=0}^{\min\binom{n-r}{n-s_Y}}\sum_{m_Y=0}^{n}\sum_{m_Z=0}^{\min\binom{n-r}{n-m_Y}}\sum_{m_W=0}^{\min\binom{n-s}{n-m_Y-m_Z}}\mathbf{C}(n,r,s_Y,s_Z,m_Y,m_Z,m_W)$$
$$\times \sum_{(y_{[n]},z_{[n-r]},w_{[n-s]},u_{[n-m]})\in\eta_{\neq}^{4n-r-s-m}} h(y_{[n]})h(y_{[r]},z_{[n-r]})$$
$$\times h(y_{[s_Y]},z_{[s_Z]},w_{[n-s]})h(y_{[m_Y]},z_{[m_Z]},w_{[m_W]},u_{[n-m]}),$$

where we renamed the variables from $U(n)$ that were not replaced to use the shorthand notation $u_{[n-m]}$. Note that the number of variables reused from $Z(n-r)$ resp. $W(n-s)$ can not be greater than the number of variables from $W(n)$ left after the replacements with variables from $Y(n)$ resp. $Y(n)$ and $Z(n-r)$ are done, thus $0 \le m_Z \le \min(n-r, n-m_Y)$ and $0 \le m_W \le \min(n-s, n-m_Y-m_Z)$. Using Mecke's formula (2.1) the claim of Lemma 3.6 is obtained directly. □

4. Proofs of the main results

For all $n \ge 2$, $d \ge 1$, $\delta > 0$ and all $y_1,\dots,y_n \in \mathbb{R}^d$ we define the indicator

$$\mathbb{1}_{\le\delta}(y_1,\dots,y_n) := \prod_{i=1}^{n}\prod_{j=i+1}^{n}\mathbb{1}\{\|y_j - y_i\|_\infty \le \delta\}.$$

where we set $\mathbb{1}_{\le\delta}(y_1,\dots,y_n) := 1$ for $n \le 1$. Additionally we combine these indicator functions with Notation 3.1 to shorten our notation throughout this section.

Remark 4.1. We note that this indicator function can also be represented using the alternative condition

$$\mathbb{1}_{\le\delta}(y_1,\dots,y_n) := \mathbb{1}\left\{\max_{i,j=1}^{n}\|y_j - y_i\|_\infty \le \delta\right\},$$

and satisfies the factorization inequality

$$\mathbb{1}_{\le\delta}(y_1,\dots,y_r,y_{r+1},\dots y_n) \le \mathbb{1}_{\le\delta}(y_1,\dots,y_r)\mathbb{1}_{\le\delta}(y_1,\dots,y_n) \tag{4.1}$$

and the argument-removal inequality

$$\mathbb{1}_{\le\delta}(y_1,\dots,y_r,y_{r+1},\dots y_n) \le \mathbb{1}_{\le\delta}(y_1,\dots,y_r) \tag{4.2}$$

for all $n \ge 2$ and $r \in \{1,\dots,n\}$.

Having this notations in place, the k-simplex counting functional F_k is a $(k+1)$-order U-statistic with measurable and symmetric kernel $\mathbb{1}_{\le\delta} : W^{k+1} \to \{0,1\}$ given by

$$F_k := \frac{1}{(k+1)!} \sum_{(y_0,\dots,y_k)\in\eta_{\neq}^{k+1}} \mathbb{1}_{\leq\delta}(y_0,\dots,y_k).$$

In the calculation of expectation and variance we will handle boundary effects using the inner parallel set $W_{-\delta}$ of W that is defined by

$$W_{-\delta} := \left\{x \in W : \overline{\mathbb{B}}_\infty^d(x,\delta) \subseteq W\right\} = \left[-\tfrac{1}{2}+\delta, +\tfrac{1}{2}-\delta\right]^d.$$

It is important to notice that

$$\Lambda_d(W_{-\delta}) = (1-2\delta)^d,$$

depends on the dimension d and on δ. Especially, the limit for $d \to \infty$ is determined by the convergence speed of δ and has a major influence on our bounds for the expectation and variance.

Remark 4.2. We will choose the sequence $(\delta_d)_d \in (0,\infty)$ such that $\delta_d \to 0$ and

$$\lim_{d\to\infty} \Lambda_d(W_{-\delta_d}) = \lim_{d\to\infty} (1-2\delta_d)^d = 1 = \Lambda_d(W).$$

Therefore δ_d has to decrease faster than $\frac{1}{d}$, i.e. we require

$$\lim_{d\to\infty} d\delta_d = 0,$$

to ensure that the observation window related factor in the lower and upper bound has the same limit. This condition can be weakened in the Gaussian case to $\lim_{d\to\infty} d\delta_d < \infty$ without changing the convergence rates presented in Section 4.3 below. However, if $\lim_{d\to\infty} d\delta_d = \infty$ rates have to be adjusted for the slower variance bound respecting $(1-2\delta_d)^d \to 0$ and it has to be ensured that $\mathbb{E}[F_k] \to \infty$ and $\widetilde{F_k} \in \mathrm{dom}(D)$ are still satisfied. In the Poisson case, the assumption can be removed completely, as long as the convergence of expectation and variance to the same positive constant is secured otherwise.

We will use $g(d) \ll f(d)$ to indicate that $g(d)$ is of order at most $f(d)$, i.e.

$$\begin{aligned} g(d) \ll f(d) &:\Leftrightarrow g(d) = \mathcal{O}(f(d)) \\ &\Leftrightarrow \exists c > 0, d_0 > 0 : \forall d > d_0 : g(d) \leq cf(d), \end{aligned}$$

where c and d_0 are constants not depending on d.

4.1. *Proof of Lemmas 1.2 and 1.3: expectation and variance.* The proof is divided into three steps, presented here as separate lemmas: First we will use Mecke's formula (2.1) and integral transformations to obtain a bound for the expectation involving an integral that does only depend on k and d. In the second step, we use the same technique combined with the variance decomposition given by Corollary 3.3 to obtain a bound for the variance. Finally we will calculate the exact values of the remaining integrals to complete the proof.

Lemma 4.3. *For $k \geq 1$ the expected number of k-simplices in the random Vietoris-Rips complex* $\mathrm{VR}^\infty(\eta_d, \delta_d)$ *is bounded by*

$$\mathbb{E}[F_k] \geq \frac{\Lambda_d(W_{-\delta})}{(k+1)!} t(t\delta^d)^k \mathcal{I}_{\mathbb{E}}(d,k),$$
$$\mathbb{E}[F_k] \leq \frac{\Lambda_d(W)}{(k+1)!} t(t\delta^d)^k \mathcal{I}_{\mathbb{E}}(d,k),$$

where $\mathcal{I}_{\mathbb{E}}(d,k)$ *denotes the integral*

$$\mathcal{I}_{\mathbb{E}}(d,k) := \int\limits_{\overline{\mathbb{B}}^d_\infty(0,1)^k} \mathbb{1}_{\leq 1}(y_1, \dots, y_k) \mathrm{d}y_1 \cdots \mathrm{d}y_k.$$

Proof: Using Mecke's formula (2.1), $\mu_d = t_d \Lambda_d$ and rewriting the indicator yields

$$\begin{aligned}\mathbb{E}[F_k] &= \tfrac{1}{(k+1)!} \int\limits_{W^{k+1}} \mathbb{1}\left\{\max_{i,j=0}^{k} \|y_j - y_i\|_\infty \leq \delta\right\} \mathrm{d}\mu^{k+1}(y_0, \dots, y_k) \\ &= \tfrac{t^{k+1}}{(k+1)!} \int\limits_W \int\limits_{W^k} \mathbb{1}\left\{\max_{j=1}^{k} \|y_j - y_0\|_\infty \leq \delta\right\} \mathbb{1}\left\{\max_{i,j=1}^{k} \|y_j - y_i\|_\infty \leq \delta\right\} \mathrm{d}y_1 \cdots \mathrm{d}y_k \mathrm{d}y_0.\end{aligned}$$

The linear transformation $y_j = y_j - y_0$, $y_0 \in W$ fixed, for all $j \in \{1, \dots, k\}$ has determinant $\det(J) = 1$, thus

$$\mathbb{E}[F_k] = \tfrac{t^{k+1}}{(k+1)!} \int\limits_W \int\limits_{(W-y_0)^k} \mathbb{1}\left\{\max_{j=1}^{k} \|y_j\|_\infty \leq \delta\right\} \mathbb{1}\left\{\max_{i,j=1}^{k} \|y_j - y_i\|_\infty \leq \delta\right\} \mathrm{d}y_k \cdots \mathrm{d}y_1 \mathrm{d}y_0.$$

The substitution $\delta y_j = y_j$ for all $j \in \{1, \dots, k\}$ has determinant $\det(J) = \delta^{dk}$, thus

$$\begin{aligned}\mathbb{E}[F_k] = \tfrac{t(t\delta^d)^k}{(k+1)!} \int\limits_W \int\limits_{(\delta^{-1}(W-y_0))^k} &\mathbb{1}\left\{\max_{j=1}^{k} \|y_j\|_\infty \leq 1\right\} \\ &\times \mathbb{1}\left\{\max_{i,j=1}^{k} \|y_j - y_i\|_\infty \leq 1\right\} \mathrm{d}y_k \cdots \mathrm{d}y_1 \mathrm{d}y_0.\end{aligned}$$

Using the inner parallel set to handle the boundary effects arising from y_0 close to ∂W we obtain the lower bound given by

$$\begin{aligned}\mathbb{E}[F_k] &\geq \tfrac{1}{(k+1)!} t(t\delta^d)^k \int\limits_{W_{-\delta}} \int\limits_{(\delta^{-1}(W-y_0)\cap \overline{\mathbb{B}}^d_\infty(0,1))^k} \mathbb{1}\left\{\max_{i,j=1}^{k} \|y_j - y_i\|_\infty \leq 1\right\} \mathrm{d}y_1 \cdots \mathrm{d}y_k \mathrm{d}y_0 \\ &= \tfrac{1}{(k+1)!} t(t\delta^d)^k \int\limits_{W_{-\delta}} \int\limits_{\overline{\mathbb{B}}^d_\infty(0,1)^k} \mathbb{1}\left\{\max_{i,j=1}^{k} \|y_j - y_i\|_\infty \leq 1\right\} \mathrm{d}y_1 \cdots \mathrm{d}y_k \mathrm{d}y_0 \\ &= \tfrac{\Lambda_d(W_{-\delta})}{(k+1)!} t(t\delta^d)^k \int\limits_{\overline{\mathbb{B}}^d_\infty(0,1)^k} \mathbb{1}\left\{\max_{i,j=1}^{k} \|y_j - y_i\|_\infty \leq 1\right\} \mathrm{d}y_1 \cdots \mathrm{d}y_k.\end{aligned}$$

Additionally, using $\delta^{-1}(W - y_0) \cap \overline{\mathbb{B}}^d_\infty(0,1) \subseteq \overline{\mathbb{B}}^d_\infty(0,1)$ we establish the upper bound

$$\begin{aligned}\mathbb{E}[F_k] &\leq \tfrac{1}{(k+1)!} t(t\delta^d)^k \int\limits_W \int\limits_{\overline{\mathbb{B}}^d_\infty(0,1)^k} \mathbb{1}\Big\{\max_{i,j=1}^{k} \|y_j - y_i\|_\infty \leq 1\Big\} \mathrm{d}y_1 \cdots \mathrm{d}y_k \mathrm{d}y_0 \\ &= \tfrac{\Lambda_d(W)}{(k+1)!} t(t\delta^d)^k \int\limits_{\overline{\mathbb{B}}^d_\infty(0,1)^k} \mathbb{1}\Big\{\max_{i,j=1}^{k} \|y_j - y_i\|_\infty \leq 1\Big\} \mathrm{d}y_1 \cdots \mathrm{d}y_k,\end{aligned}$$

which completes the proof. □

Lemma 4.4. *For $k \geq 1$ there exist explicit constants $\mathbf{C}(k,r)$ only depending on k and r such that the variance of the number of k-simplices in the random Vietoris-Rips complex $\mathrm{VR}^\infty(\eta_d, \delta_d)$ is bounded by*

$$\begin{aligned}\mathbb{V}[F_k] &\geq \mathbb{E}[F_k] + \sum_{r=1}^{k} \mathbf{C}(k+1,r)\Lambda_d(W_{-\delta}) t(t\delta^d)^{2k-r+1} \mathcal{I}_{\mathbb{V}}(d,k,r-1), \\ \mathbb{V}[F_k] &\leq \mathbb{E}[F_k] + \sum_{r=1}^{k} \mathbf{C}(k+1,r)\Lambda_d(W) t(t\delta^d)^{2k-r+1} \mathcal{I}_{\mathbb{V}}(d,k,r-1),\end{aligned}$$

where $\mathcal{I}_{\mathbb{V}}(d,k,r)$ denotes the integral

$$\mathcal{I}_{\mathbb{V}}(d,k,r) := \int\limits_{\overline{\mathbb{B}}^d_\infty(0,1)^{2k-r}} \mathbb{1}_{\leq 1}\big(y_{[1,k]}\big) \mathbb{1}_{\leq 1}\big(y_{[1,r]}, z_{[k-r]}\big) \mathrm{d}y_{[1,k]} \mathrm{d}z_{[k-r]}.$$

Proof: We apply Corollary 3.3 to our k-simplices counting statistic F_k to obtain

$$\begin{aligned}\mathbb{V}[F_k] = &\sum_{r=1}^{k+1} \mathbf{C}(k+1,r) \\ &\times \int\limits_{W^{2(k+1)-r}} \mathbb{1}_{\leq\delta}(y_{[k+1]}) \mathbb{1}_{\leq\delta}(y_{[r]}, z_{[k+1-r]}) \mathrm{d}\mu^{k+1}(y_{[k+1]}) \mathrm{d}\mu^{k+1-r}(z_{[k+1-r]}).\end{aligned}$$

For $r = k+1$ the integral is given by

$$\int\limits_{W^{k+1}} \mathbb{1}_{\leq\delta}(y_{[k+1]}) \mathbb{1}_{\leq\delta}(y_{[k+1]}) \mathrm{d}\mu^{k+1}(y_{[k+1]}) = \int\limits_{W^{k+1}} \mathbb{1}_{\leq\delta}(y_{[k+1]}) \mathrm{d}\mu^{k+1}(y_{[k+1]}),$$

which is $(k+1)!\mathbb{E}[F_k]$. Since $\mathbf{C}(k+1,k+1) = \frac{1}{(k+1)!}$ the $r = k+1$ term in the decomposition is equal to $\mathbb{E}[F_k]$. Therefore

$$\begin{aligned}\mathbb{V}[F_k] = \mathbb{E}[F_k] + &\sum_{r=1}^{k} \mathbf{C}(k+1,r) \int\limits_{W^{2(k+1)-r}} \mathbb{1}\Big\{\max_{i,j=0}^{k} \|y_j - y_i\|_\infty \leq \delta\Big\} \mathbb{1}\Big\{\max_{i,j=0}^{k-r} \|z_j - z_i\|_\infty \leq \delta\Big\} \\ &\times \mathbb{1}\Big\{\max_{i=0}^{k-r} \max_{j=0}^{r-1} \|y_j - z_i\|_\infty \leq \delta\Big\} \mathrm{d}\mu^{k+1}(y_{[k]}) \mathrm{d}\mu^{k+1-r}(z_{[k-r]}).\end{aligned}$$

We proceed analogously to the proof of Lemma 4.3 using the linear transformation $y_j = y_j - y_0$, $z_i = z_i - y_0$, $y_0 \in W$ fixed, and the substitution $\delta y_j = y_j$, $\delta z_i = z_i$

for all $j \in \{1,\dots,k\}$ and $i \in \{0,\dots,k-r\}$. Thus the integrals are given for $r \in \{1,\dots,k\}$ by

$$\begin{aligned}
&t(t\delta^d)^{2k-r+1}\int_W \int_{(\delta^{-1}(W-y_0))^{2k-r+1}} \mathbb{1}\Big\{\max_{j=1}^{k}\|y_j\|_\infty \le 1 \wedge \max_{i=0}^{k-r}\|z_i\|_\infty \le 1\Big\}\\
&\qquad\times \mathbb{1}\Big\{\max_{i,j=1}^{k}\|y_j-y_i\|_\infty \le 1 \wedge \max_{i,j=0}^{k-r}\|z_j-z_i\|_\infty \le 1\Big\}\\
&\qquad\times \mathbb{1}\Big\{\max_{i=0}^{k-r}\max_{j=1}^{r-1}\|y_j-z_j\|_\infty \le 1\Big\} \mathrm{d}y_1\cdots \mathrm{d}y_k \mathrm{d}z_0\cdots \mathrm{d}z_{k-r}\mathrm{d}y_0\\
&= t(t\delta^d)^{2k-r+1}\int_W \int_{(\delta^{-1}(W-y_0)\cap \overline{\mathbb{B}}^d_\infty(0,1))^{2k-r+1}} \mathbb{1}\Big\{\max_{i,j=1}^{k}\|y_j-y_i\|_\infty \le 1 \wedge \max_{i,j=0}^{k-r}\|z_j-z_i\|_\infty \le 1\Big\}\\
&\quad\times \mathbb{1}\Big\{\max_{i=0}^{k-r}\max_{j=1}^{r-1}\|y_j-z_j\|_\infty \le 1\Big\} \mathrm{d}y_1\cdots \mathrm{d}y_k \mathrm{d}z_0\cdots \mathrm{d}z_{k-r}\mathrm{d}y_0.
\end{aligned}$$

Using the inner parallel set to handle the boundary effects we obtain the lower bound given by

$$\begin{aligned}
\mathbb{V}[F_k] \ge \mathbb{E}[F_k] &+ \sum_{r=1}^{k} \mathbf{C}(k+1,r)\Lambda_d(W_{-\delta})t(t\delta^d)^{2k-r+1}\\
&\times \int_{\overline{\mathbb{B}}^d_\infty(0,1)^{2k-r+1}} \mathbb{1}\Big\{\max_{i,j=1}^{k}\|y_j-y_i\|_\infty \le 1 \wedge \max_{i,j=0}^{k-r}\|z_j-z_i\|_\infty \le 1\Big\}\\
&\qquad\times \mathbb{1}\Big\{\max_{i=0}^{k-r}\max_{j=1}^{r-1}\|y_j-z_j\|_\infty \le 1\Big\} \mathrm{d}y_1\cdots \mathrm{d}y_k \mathrm{d}z_0\cdots \mathrm{d}z_{k-r}.
\end{aligned}$$

Additionally, we establish the upper bound

$$\begin{aligned}
\mathbb{V}[F_k] \le \mathbb{E}[F_k] &+ \sum_{r=1}^{k} \mathbf{C}(k+1,r)\Lambda_d(W)t(t\delta^d)^{2k-r+1}\\
&\times \int_{\overline{\mathbb{B}}^d_\infty(0,1)^{2k-r+1}} \mathbb{1}\Big\{\max_{i,j=1}^{k}\|y_j-y_i\|_\infty \le 1 \wedge \max_{i,j=0}^{k-r}\|z_j-z_i\|_\infty \le 1\Big\}\\
&\qquad\times \mathbb{1}\Big\{\max_{i=0}^{k-r}\max_{j=1}^{r-1}\|y_j-z_j\|_\infty \le 1\Big\} \mathrm{d}y_1\cdots \mathrm{d}y_k \mathrm{d}z_0\cdots \mathrm{d}z_{k-r},
\end{aligned}$$

which completes the proof. □

We are now left with the task of determining the exact values of the two integrals $\mathcal{I}_{\mathbb{E}}$ in Lemma 4.3 and $\mathcal{I}_{\mathbb{V}}$ in Lemma 4.4:

Lemma 4.5. *For $d \ge 1$ and $k \ge 1$:*

$$\mathcal{I}_{\mathbb{E}}(d,k) = (k+1)^d.$$

Proof: Let us first observe that

$$\mathbb{1}\{\|y_i - y_j\|_\infty \le 1\} = \prod_{n=1}^{d} \mathbb{1}\{|y_{i,n} - y_{j,n}| \le 1\},$$

where $y_{i,n}$ denotes the n-th component of the point $y_i \in \mathbb{R}^d$. Thus

$$\mathcal{I}_{\mathbb{E}}(d,k) = \mathcal{I}_{\mathbb{E}}(1,k)^d = \left(\int\limits_{[-1,1]^k} \mathbb{1}\left\{ \max_{i,j=1}^{k} |y_j - y_i| \le 1 \right\} \mathrm{d}y_1 \cdots \mathrm{d}y_k \right)^d,$$

and we are left to show that $\mathcal{I}_{\mathbb{E}}(1,k) = k+1$. We note that

$$\mathbb{1}\left\{ \max_{i,j=1}^{k} |y_j - y_i| \le 1 \right\} = \mathbb{1}\{\max\{y_1,\dots,y_k\} - \min\{y_1,\dots,y_k\} \le 1\},$$

for all $y_1,\dots,y_k \in [-1,1]$ and thus

$$\mathcal{I}_{\mathbb{E}}(1,k) = \int\limits_{[-1,1]^k} \mathbb{1}\{\max\{y_1,\dots,y_k\} - \min\{y_1,\dots,y_k\} \le 1\} \mathrm{d}y_1 \cdots \mathrm{d}y_k.$$

Since the integrand does only depend on the maximum and the minimum of the variables $y_1,\dots,y_k$ we split $[-1,1]^k$ into $k(k-1)$ different regions that correspond to the different choices of the maximum and the minimum and observe that all regions yield the same contribution to the complete integral. Therefore we can assume without loss of generality that y_1 is the maximum, y_2 is the minimum and $y_3,\dots,y_k \in [y_2,y_1]$. Therefore we calculate the integral

$$\int\limits_{-1}^{1}\int\limits_{-1}^{y_1} \mathbb{1}\{y_1 - y_2 \le 1\}(y_1-y_2)^{k-2} \mathrm{d}y_2 \mathrm{d}y_1 = \int\limits_{-1}^{1} \int\limits_{\max\{-1,y_1-1\}}^{y_1} (y_1-y_2)^{k-2} \mathrm{d}y_2 \mathrm{d}y_1$$
$$= \int\limits_{-1}^{0}\int\limits_{-1}^{y_1} (y_1-y_2)^{k-2} \mathrm{d}y_2 \mathrm{d}y_1 + \int\limits_{0}^{1}\int\limits_{y_1-1}^{y_1} (y_1-y_2)^{k-2} \mathrm{d}y_2 \mathrm{d}y_1$$
$$= \int\limits_{-1}^{0} \frac{(y_1+1)^{k-1}}{k-1} \mathrm{d}y_1 + \int\limits_{0}^{1} \frac{1}{k-1} \mathrm{d}y_1 = \frac{1}{k(k-1)} + \frac{1}{k-1} = \frac{k+1}{k(k-1)}.$$

Multiplying by the number of regions yields $\mathcal{I}_{\mathbb{E}}(1,k) = k+1$ and completes the proof. □

Lemma 4.6. *For $d \ge 1$, $k \ge 1$ and $r \in \{0,\dots,k\}$:*

$$\mathcal{I}_{\mathbb{V}}(d,k,r) = \left(\frac{2(k+2)(k-r)}{r+2} + r + 1 \right)^d.$$

Proof: As in the proof of Lemma 4.5 we reduce the proof to the case $d = 1$ and rewrite the indicators to obtain the integral

$$\mathcal{I}_{\mathbb{V}}(1,k,r) = \int\limits_{[-1,1]^k} \int\limits_{[-1,1]^{k-r}} \mathbb{1}\left\{ \max\left\{ y_{[1,k]} \right\} - \min\left\{ y_{[1,k]} \right\} \le 1 \right\}$$
$$\times \mathbb{1}\left\{ \max\left\{ y_{[1,r]}, z_{[k-r]} \right\} - \min\left\{ y_{[1,r]}, z_{[k-r]} \right\} \le 1 \right\} \mathrm{d}z_{[k-r]} \mathrm{d}y_{[1,k]}.$$

For $r = 0$ the integral factorizes into $\mathcal{I}_{\mathbb{E}}(1,k)^2$ and since $h^2 = h$ the case $r = k$ can be directly reduced to $\mathcal{I}_{\mathbb{E}}(1,k)$ yielding the claim. For $r \in \{1,\dots,k-1\}$ we rearrange the order of integration to obtain

$$\int_{[-1,1]^r} \mathbb{1}\{\max\{y_{[1,r]}\} - \min\{y_{[1,r]}\} \leq 1\}\mathcal{J}(y_1, \dots, y_r)^2 \mathrm{d}y_{[1,r]},$$

where we define

$$\mathcal{J}(y_1, \dots, y_r) := \int_{[-1,1]^{k-r}} \mathbb{1}\{\max\{y_{[1,r]}, z_{[k-r]}\} - \min\{y_{[1,r]}, z_{[k-r]}\} \leq 1\}\mathrm{d}z_{[k-r]}.$$

Let us first examine the condition in the indicator of $\mathcal{J}(y_1, \dots, y_r)$: We note that

$$\max\{y_{[1,r]}, z_{[k-r]}\} - \min\{y_{[1,r]}, z_{[k-r]}\} \leq 1$$

is satisfied if and only if the following three conditions are satisfied at the same time:

$$\begin{aligned} |y_j - y_i| &\leq 1 \quad \forall i, j \in \{1, \dots, r\}, \\ |z_j - z_i| &\leq 1 \quad \forall i, j \in \{0, \dots, k-r-1\}, \\ |y_j - z_i| &\leq 1 \quad \forall i \in \{0, \dots, k-r-1\}, j \in \{1, \dots, r\}. \end{aligned}$$

The first condition is always satisfied since we integrate over $y_1, \dots, y_r$ with respect to the corresponding indicator in the outer integral and the third condition is equivalent to

$$z_i \in [-1 + \max\{0, y_1, \dots, y_r\}, 1 + \min\{0, y_1, \dots, y_r\}], \quad \forall i \in \{0, \dots, k-r-1\}.$$

We define the positive part of the maximum and the negative part of the minimum by

$$\begin{aligned} y_{\max} &:= \max\{0, y_1, \dots, y_r\}, \\ y_{\min} &:= \min\{0, y_1, \dots, y_r\}. \end{aligned}$$

Assuming $y_{\max} - y_{\min} \leq 1$ it follows that

$$\mathcal{J}(y_1, \dots, y_r) = \int_{[-1+y_{\max}, 1+y_{\min}]^{k-r}} \mathbb{1}\{\max\{z_{[k-r]}\} - \min\{z_{[k-r]}\} \leq 1\}\mathrm{d}z_{[k-r]}.$$

Let us first consider the case $r = k-1$ implying $k - r = 1$: It follows immediately that $\mathcal{J}(y_1, \dots, y_r) = 2 + y_{\min} - y_{\max}$. For $r \in \{1, \dots, k-2\}$ we continue similar to the proof of Lemma 4.5: We split the domain of the integral into $(k-r)(k-r-1)$ regions that correspond to the different choices of the maximum and the minimum. Without loss of generality we assume that z_0 is the maximum and z_1 is the minimum, which implies that $z_2, \dots, z_{k-r} \in [z_1, z_0]$. Therefore we obtain

$$\begin{aligned} \mathcal{J}(y_1, \dots, y_r) &= \int_{-1+y_{\max}}^{1+y_{\min}} \int_{-1+y_{\max}}^{z_0} \mathbb{1}\{z_0 - z_1 \leq 1\}(z_0 - z_1)^{k-r-2} \mathrm{d}z_1 \mathrm{d}z_0 \\ &= 1 + (k-r)(1 - (y_{\max} - y_{\min})), \end{aligned}$$

which also represents the equation for $r = k-1$. Our next objective is to evaluate the integral

$$\mathcal{I}_{\mathbb{V}}(1,k,r) = \int\limits_{[-1,1]^r} \mathbb{1}\big\{\max\big\{y_{[1,r]}\big\} - \min\big\{y_{[1,r]}\big\} \le 1\big\} \\ \times \big(1 + (k-r)\big(1 - \big(\max\big\{0, y_{[1,r]}\big\} - \min\big\{0, y_{[1,r]}\big\}\big)\big)\big)^2 \mathrm{d}y_{[1,r]}.$$

For $r = 1$ we simply derive

$$\begin{aligned}\mathcal{I}_{\mathbb{V}}(1,k,1) &= \int\limits_{-1}^{1} (1 + (k-1)(1 - (\max\{0,y\} - \min\{0,y\})))^2 \mathrm{d}y \\ &= \int\limits_{-1}^{0} (1 + (k-1)(1+y))^2 \mathrm{d}y + \int\limits_{0}^{1} (1 + (k-1)(1-y))^2 \mathrm{d}y \\ &= \frac{2}{3}\big(k^2 + k + 1\big),\end{aligned}$$

which is the desired result. For $r \in \{2, \ldots, k-1\}$ we split the domain of the integral into $r(r-1)$ regions. Without loss of generality we assume that y_1 is the maximum and y_2 the minimum, which implies that $y_3, \ldots, y_r \in [y_2, y_1]$. Thus

$$\begin{aligned}&\int\limits_{-1}^{1}\int\limits_{-1}^{y_1} \mathbb{1}\{y_1 - y_2 \le 1\}(y_1 - y_2)^{r-2} J(y_1, \ldots, y_r)^2 \mathrm{d}y_2 \mathrm{d}y_1 \\ &= \int\limits_{-1}^{1} \int\limits_{\max\{-1, y_1 - 1\}}^{y_1} (y_1 - y_2)^{r-2}(1 + (k-r)(1 - (\max\{0, y_1\} - \min\{0, y_2\})))^2 \mathrm{d}y_2 \mathrm{d}y_1 \\ &= \int\limits_{-1}^{0}\int\limits_{-1}^{y_1} (y_1 - y_2)^{r-2}(1 + (k-r)(1 - (0 - y_2)))^2 \mathrm{d}y_2 \mathrm{d}y_1 \\ &\quad + \int\limits_{0}^{1}\int\limits_{y_1 - 1}^{0} (y_1 - y_2)^{r-2}(1 + (k-r)(1 - (y_1 - y_2)))^2 \mathrm{d}y_2 \mathrm{d}y_1 \\ &\quad + \int\limits_{0}^{1}\int\limits_{0}^{y_1} (y_1 - y_2)^{r-2}(1 + (k-r)(1 - (y_1 - 0)))^2 \mathrm{d}y_2 \mathrm{d}y_1 \\ &= \frac{(2(k-r)+3)r + 2(k-r+1)^2 + r^2}{(r-1)r(r+2)}.\end{aligned}$$

Multiplying by the number of regions yields the claim for $\mathcal{I}_{\mathbb{V}}(1,k,r)$, which completes the proof, since $\mathcal{I}_{\mathbb{V}}(d,k,r) = \mathcal{I}_{\mathbb{V}}(1,k,r)^d$. □

Proof of Lemmas 1.2 and 1.3: The bound for the expectation is obtained by combining Lemma 4.3 with Lemma 4.5. Similarly, the combination of Lemma 4.4 with Lemma 4.6 yields the bound for the variance, Lemma 1.3. □

4.2. *Bounds for first and second order Malliavin-Derivatives.* The first order difference operator of our k-simplex counting functional is a U-statistics of order k, given for all $x \in W$ by

$$D_x F_k = \frac{1}{k!} \sum_{(y_0,\dots,y_{k-1}) \in \eta^k_{\neq}} \mathbb{1}_{\leq\delta}(y_0,\dots,y_{k-1},x).$$

The second order difference operator is a U-statistics of order $k-1$, given for all $x_1, x_2 \in W$ by

$$D_{x_1,x_2} F_k = \frac{1}{(k-1)!} \sum_{(y_0,\dots,y_{k-2}) \in \eta^{k-1}_{\neq}} \mathbb{1}_{\leq\delta}(y_0,\dots,y_{k-2},x_1,x_2),$$

if $k \geq 2$ and $D_{x_1,x_2} F_k = \mathbb{1}_{\leq\delta}(x_1,x_2)$ if $k = 1$, see for instance Reitzner and Schulte (2013, Lemma 3.3).

The crucial part in the application of the Malliavin-Stein method, Theorems 2.6 and 2.7, is the control over the moments of the difference operators that are used in γ_1, γ_2 and $\gamma_{3,N}$ resp. $\gamma_{3,P}$. In this section, we will prove the following bounds:

Theorem 4.7. *Let $k \geq 1$ and $d \geq 1$:*

(1) *For all $p \in \{2,3,4\}$ there exist constants $\mathbf{D}_p(k)$ only depending on k and p such that for all $x \in W$ it holds that*

$$\mathbb{E}[(D_x F_k)^p] \leq \mathbf{D}_p(k) \sum_{q=k}^{pk} (t\delta^d)^q \left((k+1) \left(\tfrac{q-k}{p-1} + 1 \right)^{p-1} \right)^d. \tag{4.3}$$

(2) *There exists a constant $\mathbf{D}^*_4(k)$ only depending on k such that for all $x \in W$ it holds that*

$$\mathbb{E}\left[\left((D_x F_k)(D_x F_k - 1) \right)^2 \right] \leq \mathbf{D}^*_4(k) \sum_{q=k+1}^{4k} (t\delta^d)^q \left((k+1) \left(\tfrac{q-k}{3} + 1 \right)^3 \right)^d. \tag{4.4}$$

(3) *There exists a constant $\mathbf{D}'_4(k)$ only depending on k such that for all $x_1, x_2 \in W$ it holds that*

$$\mathbb{E}\left[(D_{x_1,x_2} F_k)^4 \right] \leq \mathbb{1}_{\leq\delta}(x_1,x_2) \mathbf{D}'_4(k) \sum_{q=k-1}^{4(k-1)} (t\delta^d)^q \left(k \left(\tfrac{q-(k-1)}{3} + 1 \right)^3 \right)^d. \tag{4.5}$$

We have divided the proof into a sequence of lemmas. At first we will use the moment decomposition for U-statistics to obtain bounds that only involve integrals that depend on d, k and the indices given by the decomposition, see Section 3.

Therefore, for all $d \geq 1$, $k \geq 1$ we define the integrals

$$\mathcal{I}_2 := \int_{\overline{\mathbb{B}}^d_\infty(0,1)^{2k-r}} \mathbb{1}_{\leq 1}\big(y_{[k]}\big)\mathbb{1}_{\leq 1}\big(y_{[r]}, z_{[k-r]}\big)\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]}, \tag{4.6}$$

$$\mathcal{I}_3 := \int_{\overline{\mathbb{B}}^d_\infty(0,1)^{3k-r-s}} \mathbb{1}_{\leq 1}\big(y_{[k]}\big)\mathbb{1}_{\leq 1}\big(y_{[r]}, z_{[k-r]}\big)\mathbb{1}_{\leq 1}\big(y_{[s_Y]}, z_{[s_Z]}, w_{[k-s]}\big)\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]}\mathrm{d}w_{[k-s]}, \tag{4.7}$$

$$\mathcal{I}_4 := \int_{\overline{\mathbb{B}}^d_\infty(0,1)^{4k-r-s-m}} \mathbb{1}_{\leq 1}\big(y_{[k]}\big)\mathbb{1}_{\leq 1}\big(y_{[r]}, z_{[k-r]}\big)\mathbb{1}_{\leq 1}\big(y_{[s_Y]}, z_{[s_Z]}, w_{[k-s]}\big) \times \mathbb{1}_{\leq 1}\big(y_{[m_Y]}, z_{[m_Z]}, w_{[m_W]}, u_{[k-m]}\big)\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]}\mathrm{d}w_{[k-s]}\mathrm{d}u_{[k-m]}, \tag{4.8}$$

where $s := s_Y + s_Z$, $m := m_Y + m_Z + m_W$ and the indices $r, s_Y, s_Z, m_Y, m_Z, m_W$ are given according to the summations in the corresponding moment-decomposition.

Lemma 4.8. *For all $k \geq 1$, $p \in \{2,3,4\}$ there exist constants $\mathbf{D}_p(k)$ only depending on k and p such that for all $x \in W$ it holds that*

$$\mathbb{E}\big[(D_xF_k)^2\big] \leq \mathbf{D}_2(k)\sum(t\delta^d)^{2k-r}\mathcal{I}_2(d,k,r),$$
$$\mathbb{E}\big[(D_xF_k)^3\big] \leq \mathbf{D}_3(k)\sum\ldots\sum(t\delta^d)^{3k-r-s}\mathcal{I}_3(d,k,r,s_Y,s_Z),$$
$$\mathbb{E}\big[(D_xF_k)^4\big] \leq \mathbf{D}_4(k)\sum\ldots\sum(t\delta^d)^{4k-r-s-m}\mathcal{I}_4(d,k,r,s_Y,s_Z,m_Y,m_Z,m_W),$$

where the summations runs over the indices $r, s_Y, s_Z, m_Y, m_Z, m_W$ given in the corresponding moment-decomposition of the k-order U-statistics D_xF_k.

Proof: Fix $k \geq 1$ and $x \in W$.

p = 2: We use the moment-decomposition for U-statistics, Lemma 3.2, to obtain

$$\mathbb{E}\big[(D_xF_k)^2\big] = \sum_{r=0}^{k}\mathbf{C}(k,r)\int_{W^{2k-r}}\mathbb{1}_{\leq\delta}(y_{[k]},x)\mathbb{1}_{\leq\delta}(y_{[r]},z_{[k-r]},x)\mathrm{d}\mu^k(y_{[k]})\mathrm{d}\mu^{k-r}(z_{[k-r]}).$$

The linear transformation $y_j = y_j - x$, $z_i = z_i - x$ and the substitution $\delta y_j = y_j$, $\delta z_i = z_i$ for all $j \in \{0,\ldots,k-1\}$ and $i \in \{0,\ldots,k-r-1\}$, similar to the proof of Lemma 4.3, yield

$$\mathbb{E}\big[(D_xF_k)^2\big] = \sum_{r=0}^{k}\mathbf{C}(k,r)(t\delta^d)^{2k-r}\int_{(\delta^{-1}(W-x))^{2k-r}}\mathbb{1}\Big\{\max_{j=0}^{k-1}\|y_j\|_\infty \leq 1 \wedge \max_{i=0}^{k-r-1}\|z_i\|_\infty \leq 1\Big\} \times \mathbb{1}_{\leq 1}\big(y_{[k]}\big)\mathbb{1}_{\leq 1}\big(y_{[r]}, z_{[k-r]}\big)\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]}.$$

We set $\mathbf{D}_2(k) := \max_{r=0}^{k}\mathbf{C}(k,r)$ and use the first indicator to bound the domain of integration to obtain

$$\mathbb{E}\big[(D_xF_k)^2\big] \leq \mathbf{D}_2(k)\sum_{r=0}^{k}(t\delta^d)^{2k-r}\int_{\overline{\mathbb{B}}^d_\infty(0,1)^{2k-r}}\mathbb{1}_{\leq 1}\big(y_{[k]}\big)\cdot\mathbb{1}_{\leq 1}\big(y_{[r]}, z_{[k-r]}\big)\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]},$$

which is our claim.

p = 3: Using Lemma 3.5 it follows that

$$\mathbb{E}\big[(D_x F_k)^3\big] = \sum_{r=0}^{k} \sum_{s_Y=0}^{k} \sum_{s_Z=0}^{\min\binom{k-r}{k-s_Y}} \mathbf{C}(k, r, s_Y, s_Z) \int_{W^{3k-r-s}} \mathbb{1}_{\leq\delta}(y_{[k]}, x) \mathbb{1}_{\leq\delta}(y_{[r]}, z_{[k-r]}, x)$$
$$\times \mathbb{1}_{\leq\delta}(y_{[s_Y]}, z_{[s_Z]}, w_{[k-s]}, x) \mathrm{d}\mu^n(y_{[k]}) \mathrm{d}\mu^{k-r}(z_{[k-r]}) \mathrm{d}\mu^{k-s}(w_{[k-s]}).$$

The linear transformation $y_j = y_j - x$, $z_i = z_i - x$, $w_l = w_l - x$ and the substitution $\delta y_j = y_j$, $\delta z_i = z_i$, $\delta w_l = w_l$ for all $j \in \{0, \ldots, k-1\}$, $i \in \{0, k-r-1\}$ and $l \in \{0, \ldots, k-s-1\}$ yield

$$\mathbb{E}\big[(D_x F_k)^3\big] = \sum_{r=0}^{k} \sum_{s_Y=0}^{k} \sum_{s_Z=0}^{\min\binom{k-r}{k-s_Y}} \mathbf{C}(k, r, s_Y, s_Z)(t\delta^d)^{3k-r-s}$$
$$\times \int_{(\delta^{-1}(W-x))^{3k-r-s}} \mathbb{1}\left\{ \max_{j=0}^{k-1} \|y_j\|_\infty \leq 1 \wedge \max_{i=0}^{k-r-1} \|z_i\|_\infty \leq 1 \wedge \max_{l=0}^{k-s-1} \|w_l\|_\infty \leq 1 \right\}$$
$$\times \mathbb{1}_{\leq 1}(y_{[k]}) \mathbb{1}_{\leq 1}(y_{[r]}, z_{[k-r]}) \mathbb{1}_{\leq 1}(y_{[s_Y]}, z_{[s_Z]}, w_{[k-s]}) \mathrm{d}y_{[k]} \mathrm{d}z_{[k-r]} \mathrm{d}w_{[k-s]}.$$

We set $\mathbf{D}_3(k)$ to be the maximal constant $\mathbf{C}(k, r, s_Y, s_Z)$ occurring in the sum and use the first indicator to bound the domain of integration to obtain

$$\mathbb{E}\big[(D_x F_k)^3\big] \leq \mathbf{D}_3(k) \sum_{r=0}^{k} \sum_{s_Y=0}^{k} \sum_{s_Z=0}^{\min\binom{k-r}{k-s_Y}} (t\delta^d)^{3k-r-s}$$
$$\times \int_{\overline{\mathbb{B}}^d_\infty(0,1)^{3k-r-s}} \mathbb{1}_{\leq 1}(y_{[k]}) \mathbb{1}_{\leq 1}(y_{[r]}, z_{[k-r]}) \mathbb{1}_{\leq 1}(y_{[s_Y]}, z_{[s_Z]}, w_{[k-s]}) \mathrm{d}y_{[k]} \mathrm{d}z_{[k-r]} \mathrm{d}w_{[k-s]},$$

which establishes the formula.

p = 4: It follows from Lemma 3.6 that

$$\mathbb{E}\big[(D_x F_k)^4\big]$$
$$= \sum_{r=0}^{k} \sum_{s_Y=0}^{k} \sum_{s_Z=0}^{\min\binom{k-r}{k-s_Y}} \sum_{m_Y=0}^{k} \sum_{m_Z=0}^{\min\binom{k-r}{k-m_Y}} \sum_{m_W=0}^{\min\binom{k-s}{k-m_Y-m_Z}} \mathbf{C}(k, r, s_Y, s_Z, m_Y, m_Z, m_W)$$
$$\times \int_{W^{4k-r-s-m}} \mathbb{1}_{\leq\delta}(y_{[k]}, x) \mathbb{1}_{\leq\delta}(y_{[r]}, z_{[k-r]}, x)$$
$$\times \mathbb{1}_{\leq\delta}(y_{[s_Y]}, z_{[s_Z]}, w_{[k-s]}, x) \mathbb{1}_{\leq\delta}(y_{[m_Y]}, z_{[m_Z]}, w_{[m_W]}, u_{[k-m]}, x)$$
$$\mathrm{d}\mu^k(y_{[k]}) \mathrm{d}\mu^{k-r}(z_{[k-r]}) \mathrm{d}\mu^{k-s}(w_{[k-s]}) \mathrm{d}\mu^{k-m}(u_{[k-m]}).$$

Using the linear transformation $y_j = y_j - x$, $z_i = z_i - x$, $w_l = w_l - x$, $u_n = u_n - x$ and the substitution $\delta y_j = y_j$, $\delta z_i = z_i$, $\delta w_l = w_l$, $\delta u_n = u_n$ for all $j \in \{0, \ldots, k-1\}$,

$i \in \{0,\dots,k-r-1\}$, $l \in \{0,\dots,k-s-1\}$ and $n \in \{0,\dots,k-m-1\}$ we obtain

$$\mathbb{E}\big[(D_xF_k)^4\big] = \sum_{r=0}^{k}\sum_{s_Y=0}^{k}\sum_{s_Z=0}^{\min\binom{k-r}{k-s_Y}}\sum_{m_Y=0}^{k}\sum_{m_Z=0}^{\min\binom{k-r}{k-m_Y}}\sum_{m_W=0}^{\min\binom{k-s}{k-m_Y-m_Z}}\mathbf{C}(k,\dots)(t\delta^d)^{4k-r-s-m}$$
$$\times \int\limits_{(\delta^{-1}(W-x))^{4k-r-s-m}} \mathbb{1}\left\{\max_{j=0}^{k-1}\|y_j\|_\infty \le 1 \wedge \max_{i=0}^{k-r-1}\|z_i\|_\infty \le 1\right\}$$
$$\times \mathbb{1}\left\{\max_{l=0}^{k-s-1}\|w_l\|_\infty \le 1 \wedge \max_{n=0}^{k-m-1}\|u_n\|_\infty \le 1\right\}$$
$$\times \mathbb{1}_{\le 1}(y_{[k]})\mathbb{1}_{\le 1}(y_{[r]}, z_{[k-r]})\mathbb{1}_{\le 1}(y_{[s_Y]}, z_{[s_Z]}, w_{[k-s]})$$
$$\times \mathbb{1}_{\le 1}(y_{[m_Y]}, z_{[m_Z]}, w_{[m_W]}, u_{[k-m]})\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]}\mathrm{d}w_{[k-s]}\mathrm{d}u_{[k-m]}.$$

Setting $\mathbf{D}_4(k)$ to be the maximal constant $\mathbf{C}(k, r, s_Y, s_Z, m_Y, m_Z, m_W)$ occurring in the sum and bounding the domain of integration with the first indicator yields

$$\mathbb{E}\big[(D_xF_k)^4\big] \le \mathbf{D}_4(k)\sum_{r=0}^{k}\sum_{s_Y=0}^{k}\sum_{s_Z=0}^{\min\binom{k-r}{k-s_Y}}\sum_{m_Y=0}^{k}\sum_{m_Z=0}^{\min\binom{k-r}{k-m_Y}}\sum_{m_W=0}^{\min\binom{k-s}{k-m_Y-m_Z}}(t\delta^d)^{4k-r-s-m}$$
$$\times \int\limits_{\overline{\mathbb{B}}^d_\infty(0,1)^{4k-r-s-m}} \mathbb{1}_{\le\delta}(y_{[k]})\mathbb{1}_{\le\delta}(y_{[r]}, z_{[k-r]})\mathbb{1}_{\le\delta}(y_{[s_Y]}, z_{[s_Z]}, w_{[k-s]})$$
$$\times \mathbb{1}_{\le\delta}(y_{[m_Y]}, z_{[m_Z]}, w_{[m_W]}, u_{[k-m]})\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]}\mathrm{d}w_{[k-s]}\mathrm{d}u_{[k-m]}.$$

which completes the proof. □

Lemma 4.9. *For all $k \ge 2$ there exists a constant $\mathbf{D}'_4(k)$ only depending on k such that for all $x_1, x_2 \in W$ it holds that*

$$\mathbb{E}\big[(D_{x_1,x_2}F_k)^4\big] \le \mathbb{1}\{\|x_1 - x_2\|_\infty \le \delta\}\mathbf{D}'_4(k)$$
$$\times \sum\dots\sum(t\delta^d)^{4(k-1)-r-s-m}\mathcal{I}_4(d, k-1, r, s_Y, s_Z, m_Y, m_z, m_W),$$

where the summation runs over the indices $r, s_Y, s_Z, m_Y, m_Z, m_W$ given in the corresponding moment-decomposition of the $(k-1)$-order U-statistics $D_{x_1,x_2}F_k$. For $k = 1$ it holds that $\mathbb{E}\big[(D_{x_1,x_2}F_1)^4\big] = \mathbb{1}\{\|x_1 - x_2\|_\infty \le \delta\}$.

Proof: Since $D_{x_1,x_2}F_1 = \mathbb{1}\{\|x_1 - x_2\|_\infty \le \delta\}$, the claim for $k = 1$ follows immediately. Fix $k \ge 2$ and $x_1, x_2 \in W$. We use the moment-decomposition for U-statistics, Lemma 3.6, on $D_{x_1,x_2}F_k$ to obtain

$$\mathbb{E}\big[(D_{x_1,x_2}F_k)^4\big] = \sum\dots\sum\mathbf{C}(k-1, r, s_Y, s_Z, m_Y, m_Z, m_W)$$
$$\times \int\limits_{W^{4(k-1)-r-s-m}} \mathbb{1}_{\le\delta}\big(y_{[k-1]}, x_1, x_2\big)\mathbb{1}_{\le\delta}\big(y_{[r]}, z_{[k-1-r]}, x_1, x_2\big)$$
$$\times \mathbb{1}_{\le\delta}\big(y_{[s_Y]}, z_{[s_Z]}, w_{[k-1-s]}, x_1, x_2\big)$$
$$\times \mathbb{1}_{\le\delta}\big(y_{[m_Y]}, z_{[m_Z]}, w_{[m_W]}, u_{[k-1-m]}, x_1, x_2\big)\mathrm{d}\mu^{4(k-1)-r-s-m}(y, z, w, u).$$

Using the factorization inequality (4.1) and the argument-removal inequality (4.2) for $\mathbb{1}_{\leq\delta}(\cdot,x_1,x_2)$ it follows that $\mathbb{1}_{\leq\delta}(\cdot,x_1,x_2) \leq \mathbb{1}_{\leq\delta}(x_1,x_2)\mathbb{1}_{\leq\delta}(\cdot,x_1)$ and thus

$$\mathbb{E}\big[(D_{x_1,x_2}F_k)^4\big] \leq \sum \cdots \sum \mathbf{C}(k-1,r,s_Y,s_Z,m_Y,m_Z,m_W)$$
$$\times\, \mathbb{1}_{\leq\delta}(x_1,x_2)\int_{W^{4(k-1)-r-s-m}} \mathbb{1}_{\leq\delta}\big(y_{[k-1]},x_1\big)\mathbb{1}_{\leq\delta}\big(y_{[r]},z_{[k-1-r]},x_1\big)\mathbb{1}_{\leq\delta}\big(y_{[s_Y]},z_{[s_Z]},w_{[k-1-s]},x_1\big)$$
$$\times\, \mathbb{1}_{\leq\delta}\big(y_{[m_Y]},z_{[m_Z]},w_{[m_W]},u_{[k-1-m]},x_1\big)\mathrm{d}\mu^{4(k-1)-r-s-m}(y,z,w,u).$$

Using the linear transformation $y_j = y_j - x_1$, $z_i = z_i - x_1$, $u_l = u_l - x_1$ and $w_n = w_n - x_1$ and the substitution $\delta y_j = y_j$, $\delta z_i = z_i$, $\delta u_l = u_l$ and $\delta w_n = w_n$ for all $j \in \{0,\ldots,k-2\}$, $i \in \{0,\ldots,k-r-2\}$, $l \in \{0,\ldots,k-s-2\}$ and $n \in \{0,\ldots,k-m-2\}$ we obtain

$$\mathbb{E}\big[(D_{x_1,x_2}F_k)^4\big] \leq \sum \cdots \sum \mathbf{C}(k-1,r,s_Y,s_Z,m_Y,m_Z,m_W)(t\delta^d)^{4(k-1)-r-s-m}$$
$$\times\, \mathbb{1}_{\leq\delta}(x_1,x_2)\int_{(\delta^{-1}(W-x))^{4(k-1)-r-s-m}} \mathbb{1}\Big\{\max_{j=0}^{k-2}\|y_j\|_\infty \leq 1 \wedge \max_{i=0}^{k-r-2}\|z_i\|_\infty \leq 1\Big\}$$
$$\times\, \mathbb{1}\Big\{\max_{l=0}^{k-s-2}\|w_l\|_\infty \leq 1 \wedge \max_{n=0}^{k-m-2}\|u_n\|_\infty \leq 1\Big\}$$
$$\times\, \mathbb{1}_{\leq\delta}(y_{[k-1]})\mathbb{1}_{\leq\delta}(y_{[r]},z_{[k-1-r]})\mathbb{1}_{\leq\delta}(y_{[s_Y]},z_{[s_Z]},w_{[k-1-s]})$$
$$\times\, \mathbb{1}_{\leq\delta}(y_{[m_Y]},z_{[m_Z]},w_{[m_W]},u_{[k-m-1]})\mathrm{d}y_{[k-1]}\mathrm{d}z_{[k-1-r]}\mathrm{d}w_{[k-1-s]}\mathrm{d}u_{[k-1-m]}.$$

Setting $\mathbf{D}_4(k)$ to be the maximal constant $\mathbf{C}(k-1,r,s_Y,s_Z,m_Y,m_Z,m_W)$ occurring in the sum and bounding the domain of integration using the first indicator yields

$$\mathbb{E}\big[(D_{x_1,x_2}F_k)^4\big] \leq \sum \cdots \sum \mathbf{C}(k-1,r,s_Y,s_Z,m_Y,m_Z,m_W)(t\delta^d)^{4(k-1)-r-s-m}$$
$$\times\, \mathbb{1}_{\leq\delta}(x_1,x_2)\int_{\overline{\mathbb{B}}^d_\infty(0,1)^{4(k-1)-r-s-m}} \mathbb{1}_{\leq\delta}(y_{[k-1]})\mathbb{1}_{\leq\delta}(y_{[r]},z_{[k-1-r]})\mathbb{1}_{\leq\delta}(y_{[s_Y]},z_{[s_Z]},w_{[k-1-s]})$$
$$\times\, \mathbb{1}_{\leq\delta}(y_{[m_Y]},z_{[m_Z]},w_{[m_W]},u_{[k-m-1]})\mathrm{d}y_{[k-1]}\mathrm{d}z_{[k-1-r]}\mathrm{d}w_{[k-1-s]}\mathrm{d}u_{[k-1-m]},$$

which completes the proof. □

In the next step we derive a bound for the integrals depending on the indices in the moment-decomposition:

Lemma 4.10. *For all $d \geq 1$, $k \geq 1$ and all $r, s_Y, s_Z, m_Y, m_Z, m_W \in \{0,\ldots,k\}$ we have*

$$\mathcal{I}_2(d,k,r) \leq ((k+1)(k-r+1))^d$$
$$\mathcal{I}_3(d,k,r,s_Y,s_Z) \leq ((k+1)(k-r+1)(k-s+1))^d$$
$$\mathcal{I}_4(d,k,r,s_Y,s_Z,m_Y,m_Z,m_W) \leq ((k+1)(k-r+1)(k-s+1)(k-m+1))^d,$$

where $s := s_Y + s_Z$ and $m := m_Y + m_Z + m_W$.

Proof: Similarly to the proof of Lemma 4.5, we factorize the integral to obtain

$$\mathcal{I}_2(d,k,r) = \mathcal{I}_2(1,k,r)^d.$$

Therefore we only have to consider the case $d = 1$:

$$\mathcal{I}_2(1,k,r) = \int\limits_{[-1,1]^{2k-r}} \mathbb{1}_{\le 1}\big(y_{[k]}\big)\mathbb{1}_{\le 1}\big(y_{[r]}, z_{[k-r]}\big)\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]}.$$

Using (4.1) and (4.2) we obtain

$$\mathbb{1}_{\le 1}\big(y_{[r]}, z_{[k-r]}\big) \le \mathbb{1}_{\le 1}\big(y_{[r]}\big)\cdot \mathbb{1}_{\le 1}\big(z_{[k-r]}\big),$$

for all $r \in \{0,\dots,k\}$. Hence

$$\mathcal{I}_2(1,k,r) \le \int\limits_{[-1,1]^{2k-r}} \mathbb{1}_{\le 1}\big(y_{[k]}\big)\mathbb{1}_{\le 1}\big(y_{[r]}\big)\mathbb{1}_{\le 1}\big(z_{[k-r]}\big)\mathrm{d}y_{[k]}\mathrm{d}z_{[k-r]}.$$

Since $\mathbb{1}_{\le 1}\big(y_{[k]}\big)\mathbb{1}_{\le 1}\big(y_{[r]}\big) = \mathbb{1}_{\le 1}\big(y_{[k]}\big)$ this integral factorizes into two integrals separating the y and z variables. Thus

$$\mathcal{I}_2(1,k,r) \le \int\limits_{[-1,1]^{k}} \mathbb{1}_{\le 1}\big(y_{[k]}\big)\mathrm{d}y_{[k]} \times \int\limits_{[-1,1]^{k-r}} \mathbb{1}_{\le 1}\big(z_{[k-r]}\big)\mathrm{d}z_{[k-r]}.$$

The claim follows directly from Lemma 4.5, since the two factors are given by $\mathcal{I}_{\mathbb{E}}(1,k)$ and $\mathcal{I}_{\mathbb{E}}(1,k-r)$. In the same manner we bound the integrals $\mathcal{I}_3$ and $\mathcal{I}_4$ using

$$\begin{aligned}\mathcal{I}_3(1,k,r,s_Y,s_Z) \le &\int\limits_{[-1,1]^{k}} \mathbb{1}_{\le 1}\big(y_{[k]}\big)\mathrm{d}y_{[k]} \times \int\limits_{[-1,1]^{k-r}} \mathbb{1}_{\le 1}\big(z_{[k-r]}\big)\mathrm{d}z_{[k-r]}\\ &\times \int\limits_{[-1,1]^{k-s}} \mathbb{1}_{\le 1}\big(w_{[k-s]}\big)\mathrm{d}w_{[k-s]},\end{aligned}$$

and

$$\begin{aligned}\mathcal{I}_4(1,k,r,s_Y,s_Z) \le &\int\limits_{[-1,1]^{k}} \mathbb{1}_{\le 1}\big(y_{[k]}\big)\mathrm{d}y_{[k]} \times \int\limits_{[-1,1]^{k-r}} \mathbb{1}_{\le 1}\big(z_{[k-r]}\big)\mathrm{d}z_{[k-r]}\\ &\times \int\limits_{[-1,1]^{k-s}} \mathbb{1}_{\le 1}\big(w_{[k-s]}\big)\mathrm{d}w_{[k-s]} \times \int\limits_{[-1,1]^{k-m}} \mathbb{1}_{\le 1}\big(u_{[k-m]}\big)\mathrm{d}u_{[k-m]}.\end{aligned}$$

□

Finally, we simplify the bounds given by the previous lemma using only the number of variables in the integral, removing the dependencies on the specific choice of indices:

Lemma 4.11. *For all $d \ge 1$, $k \ge 1$ and $p \in \{2,3,4\}$ we denote by $q_p \in \{k,\dots,pk\}$ the number of variables in the integral $\mathcal{I}_p$.*

(1) *For all indices r such that $2k - r = q_2$ it holds that*

$$\mathcal{I}_2(k,r) \le ((k+1)(q_2 - k + 1))^d. \tag{4.9}$$

(2) *For all indices r, s_Y, s_Z such that $3k - r - s = q_3$ it holds that*

$$\mathcal{I}_3(k,r,s_Y,s_Z) \le \left((k+1)\left(\frac{q_3 - k}{2} + 1\right)^2\right)^d. \tag{4.10}$$

(3) *For all indices* $r, s_Y, s_Z, m_Y, m_Z, m_W$ *such that* $4k-r-s-m = q_4$ *it holds that*

$$\mathcal{I}_4(k, r, s_Y, s_Z, m_Y, m_Z, m_W) \le \left((k+1)\left(\frac{q_4-k}{3}+1\right)^3\right)^d. \tag{4.11}$$

Proof: We give the proof only for the case $p = 4$; the other cases are similar. We note that it is sufficient to show the claim for $d = 1$: For fixed $k \ge 1$ and $q_4 \in \{k, \dots, 4k\}$ we define $g_3 : [0,k]^3 \to \mathbb{R}$ by

$$g_3(r, s, m) := (k+1)(k-r+1)(k-s+1)(k-m+1).$$

Maximizing over $[0,k]^3$ with respect to the condition $F(r,s,m) := 4k-r-s-m-q_4 = 0$ yields the maximal value for

$$r = s = m = \frac{4k-q_4}{3}.$$

Thus, using Lemma 4.10, we have for all indices $r, s_Y, s_Z, m_Y, m_Z, m_W \in \{0, \dots, k\}$, with $4k-r-s-m = q_4$ the bound

$$\mathcal{I}_4(k, r, s_Y, s_Z, m_Y, m_Z, m_W) \le g\left(\frac{4k-q_4}{3}, \frac{4k-q_4}{3}, \frac{4k-q_4}{3}\right),$$

which is our claim. □

We are now in a position, to show the main result of this section, Theorem 4.7:

Proof of Theorem 4.7 a) and c): We give the proof only for the case $p = 4$; the other cases are similar. We reorganize the summation in the bound given by Lemma 4.8 according to the number q of variables in the integral $\mathcal{I}_4$, i.e.

$$\begin{aligned}
&\mathbb{E}\big[(D_xF_k)^4\big] \\
&\le \mathbf{D}_4(k)\sum\dots\sum(t\delta_d)^{4k-r-s-m}\mathcal{I}_4(d,k,r,s_Y,s_Z,m_Y,m_Z,m_W) \\
&= \mathbf{D}_4(k)\sum_{q=k}^{4k}\sum\dots\sum \mathbb{1}\{4k-r-s-m=q\}(t\delta^d)^q\mathcal{I}_4(d,k,r,s_Y,s_Z,m_Y,m_Z,m_W).
\end{aligned}$$

Applying (4.11), yields

$$\begin{aligned}
&\mathbb{E}\big[(D_xF_k)^4\big] \\
&\le \mathbf{D}_4(k)\sum_{q=k}^{4k}(t\delta^d)^q\Big((k+1)(\tfrac{q-k}{3}+1)^3\Big)^d\sum\dots\sum\mathbb{1}\{4k-r-s-m=q\}.
\end{aligned}$$

Finally, the last summations over the indicator yields a constant that does only depend on k and q. Thus we redefine the constant $\mathbf{D}_4(k)$ and the claim a) follows. The proof of claim c) is similar. □

Proof of Theorem 4.7 b): Since

$$\mathbb{E}\Big[\big((D_xF)(D_xF-1)\big)^2\Big] = \mathbb{E}\big[(D_xF)^4\big] - 2\mathbb{E}\big[(D_xF)^3\big] + \mathbb{E}\big[(D_xF)^2\big] \tag{4.12}$$

we use the moment decomposition for U-statistics, Lemmas 3.2, 3.5 and 3.6 on D_xF to obtain the decomposition for $\mathbb{E}[((D_xF)(D_xF-1))^2]$. Recall that the decomposition of the second moment is given by

$$\mathbb{E}\big[(D_xF)^2\big] = \sum_{r=0}^{k} \mathbf{C}(k,r) \int\limits_{W^{2k-r}} \mathbb{1}_{\leq\delta}(y_{[k]},x)\mathbb{1}_{\leq\delta}(y_{[r]},z_{[k-r]},x)\mathrm{d}\mu^k(y_{[k]})\mathrm{d}\mu^{k-r}(z_{[k-r]}).$$

For $r=k$ the integral is equal to the integral in $\mathbb{E}[D_xF]$ an further $\mathbf{C}(k,k)=\frac{1}{k!}$, thus we rewrite this decomposition into

$$\begin{aligned}\mathbb{E}\big[(D_xF)^2\big] &= \mathbb{E}[(D_xF)] \\ &\quad + \sum_{r=0}^{k-1} \mathbf{C}(k,r) \int\limits_{W^{2k-r}} \mathbb{1}_{\leq\delta}(y_{[k]},x))\mathbb{1}_{\leq\delta}(y_{[r]},z_{[k-r]},x)\mathrm{d}\mu^k(y_{[k]})\mathrm{d}\mu^{k-r}(z_{[k-r]}).\end{aligned}$$

Similarly, for $r=k$, $s_Y=k$, which forces us to set $s_Z=0$ by definition, the term in the decomposition of the third moment is equal to $\mathbb{E}[D_xF]$. For $r=k$, $s_Y=k$, $m_Y=k$, which forces us to set $s_Z=0$, $m_Z=0$ and $m_W=0$, the term in the decomposition of the fourth moment is also equal to $\mathbb{E}[D_xF]$. Therefore, these terms cancel out in the decomposition of (4.12). We further note, that these combinations are the only choices of indices in the decomposition yielding integrals that do only involve k distinct variables. It follows immediately that all other integrals appearing in the decomposition have at least $k+1$ distinct variables. Denoting the integrals in Lemmas 3.2, 3.5 and 3.6 by $\mathcal{J}_2$, $\mathcal{J}_3$ resp. $\mathcal{J}_4$, we use the decomposition to rewrite the mixed moments. Since $D_xF_k \geq 0$ for all $x\in\mathbb{R}^d$ we obtain the bound

$$\begin{aligned}\mathbb{E}\Big[\big((D_xF)(D_xF-1)\big)^2\Big] &= \sum_{q=k+1}^{2k} \sum \mathbb{1}\{2k-r=q\}\mathbf{C}(k,r)\mathcal{J}_2(d,k,r) \\ &- \sum_{q=k+1}^{3k} \sum \dots \sum \mathbb{1}\{3k-r-s=q\}\mathbf{C}(k,r,s_Y,s_Z)\mathcal{J}_3(k,r,s_Y,s_Z) \\ &+ \sum_{q=k+1}^{4k} \sum \dots \sum \mathbb{1}\{4k-r-s-m=q\}\mathbf{C}(k,r,s_Y,s_Z,m_Y,m_Z,m_W) \\ &\qquad \times \mathcal{J}_4(d,k,r,s_Y,s_Z,m_Y,m_Z,m_W) \\ &\leq \sum_{q=k+1}^{2k} \sum \mathbb{1}\{2k-r=q\}\mathbf{C}(k,r)\mathcal{J}_2(d,k,r) \\ &+ \sum_{q=k+1}^{4k} \sum \dots \sum \mathbb{1}\{4k-r-s-m=q\}\mathbf{C}(k,r,s_Y,s_Z,m_Y,m_Z,m_W) \\ &\qquad \times \mathcal{J}_4(d,k,r,s_Y,s_Z,m_Y,m_Z,m_w)\end{aligned}$$

where we just removed the negative term to simplify the upper bound. Analysis similar to that in the proof of Theorem 4.7 a) before shows

$$\begin{aligned}
&\mathbb{E}\Big[\big((D_xF)(D_xF-1)\big)^2\Big]\\
&\le \mathbf{D}_2(k)\sum_{q=k+1}^{2k}(t\delta^d)^q((k+1)(q-k+1))^d+\mathbf{D}_4(k)\sum_{q=k+1}^{4k}(t\delta^d)^q\Big((k+1)\Big(\tfrac{q-k}{3}+1\Big)^3\Big)^d\\
&\le (\mathbf{D}_2(k)+\mathbf{D}_4(k))\sum_{q=k+1}^{4k}(t\delta^d)^q\Big((k+1)\Big(\tfrac{q-k}{3}+1\Big)^3\Big)^d,
\end{aligned}$$

since $(q-k+1)\le(\frac{q-k}{3}+1)^3$. Defining $\mathbf{D}_4^*(k):=\mathbf{D}_2(k)+\mathbf{D}_4(k)$ completes the proof. □

4.3. *Proof of Theorems 1.4 and 1.5: Gaussian and Poisson limit.* Let us first investigate the asymptotic behavior of the variance $\mathbb{V}[F_k]$ in the three different phases determined by the limit of the expectation $\mathbb{E}[F_k]$. Using the bound given by Lemma 1.3 we obtain

$$\mathbb{E}[F_k]+(1-2\delta)^dR(d,k,t,\delta)\le\mathbb{V}[F_k]\le\mathbb{E}[F_k]+R(d,k,t,\delta),$$

where we defined

$$R(d,k,t,\delta):=t(t\delta^d)^k\sum_{r=1}^{k}\mathbf{C}(k,r)(t\delta^d)^{k+1-r}\Big(\tfrac{2(k+2)(k+1-r)}{r+1}+r\Big)^d,$$

and note that $R(d,k,t,\delta)\ge0$ holds for $d\ge1$, $k\ge1$, $t>0$ and $\delta>0$.

Lemma 4.12. *For all $k\ge1$ the variance of the k-simplex counting functional satisfies*

$$\lim_{d\to\infty}\mathbb{V}[F_k]=\lim_{d\to\infty}\mathbb{E}[F_k],$$

if the limit of the expectation is infinite (1.1), a positive constant (1.2) or zero (1.3).

Proof: Let us first assume that the limit of the expectation is either a positive constant (1.2) or zero (1.3). Thus we have

$$\lim_{d\to\infty}\tfrac{1}{(k+1)!}t(t\delta^d)^k(k+1)^d=\theta\in[0,\infty)$$

and it follows that

$$\begin{aligned}
&R(d,k,t,\delta)\\
&=t(t\delta^d)^k\tfrac{(k+1)^d}{(k+1)^d}\sum_{r=1}^{k}\mathbf{C}(k,r)\Big((t\delta^d)^k\tfrac{(k+1)^d}{(k+1)^d}\Big)^{\frac{k+1-r}{k}}\left(\frac{2(k+2)(k+1-r)}{r+1}+r\right)^d\\
&=t(t\delta^d)^k(k+1)^d\sum_{r=1}^{k}\mathbf{C}(k,r)\big((t\delta^d)^k(k+1)^d\big)^{\frac{k+1-r}{k}}\left(\tfrac{2(k+2)(k+1-r)+r(r+1)}{(r+1)(k+1)(k+1)^{\frac{k+1-r}{k}}}\right)^d.
\end{aligned}$$

We define the function $g_k:[0,\infty)\to[0,\infty)$ by

$$g_k(r):=\tfrac{2(k+2)(k+1-r)+r(r+1)}{(r+1)(k+1)(k+1)^{\frac{k+1-r}{k}}}$$

and a straightforward calculation shows that $g_k''(r) \geq 0$ for $r \in [1, k+1]$ implying convexity of g_k on $[1, k+1]$. Using $g_k(1) = 1$ and $g_k(k+1) = 1$ we obtain

$$g_k(\alpha + (1-\alpha)(k+1)) \leq \alpha g_k(1) + (1-\alpha) g_k(k+1) \leq 1,$$

for all $\alpha \in [0,1]$ which implies $g_k(r) \in [0,1]$ for all $r \in \{1, \dots, k+1\}$. Therefore we have

$$R(d,k,t,\delta) \leq t(t\delta^d)^k (k+1)^d \sum_{r=1}^{k} \mathbf{C}(k,r) \big((t\delta^d)^k (k+1)^d\big)^{\frac{k+1-r}{k}}.$$

Since $t \to \infty$ and we assume (1.2) or (1.3) we have $(t\delta^d)^k (k+1)^d \to 0$, thus

$$R(d,k,t,\delta) \leq \underbrace{t(t\delta^d)^k (k+1)^d}_{\to (k+1)!\theta} \sum_{r=1}^{k} \mathbf{C}(k,r) \Bigg(\underbrace{(t\delta^d)^k (k+1)^d}_{\to 0} \Bigg)^{\frac{k+1-r}{k}} \to 0,$$

which yields our claim.

If the expectation tends to infinity (1.1) we directly obtain the claim from the lower variance bound. □

In the next step we check that our k-simplex counting functional F_k or its standardization $\widetilde{F_k}$ satisfy the condition $F \in \mathrm{dom}(D)$ resp. $\widetilde{F_k} \in \mathrm{dom}(D)$ imposed through the Malliavin-Stein method:

Lemma 4.13. *For all $d \geq 1$ and all $k \geq 1$ there exists a constant $\mathbf{D}^{\mathrm{dom}}(k) \in (0, \infty)$ only depending on k such that*

$$\int_W \mathbb{E}\big[(D_x F_k)^2\big] \leq \mathbf{D}^{\mathrm{dom}}(k) \mathbb{V}[F_k].$$

Furthermore, if the expectation tends to infinity (1.1), then

$$\widetilde{F_k} \in \mathrm{dom}(D).$$

If the expectation converges to a positive constant (1.2), then

$$F_k \in \mathrm{dom}(D).$$

Proof: We use the moment decomposition Lemma 3.2 on $D_x F$ to obtain

$$\begin{aligned}
&\int_W \mathbb{E}\big[(D_x F_k)^2\big] \mathrm{d}\mu(x) \\
&= \int_W \sum_{r=0}^{k} \mathbf{C}(k,r) \int_{W^{2k-r}} \mathbb{1}_{\leq \delta}\big(y_{[k]}, x\big) \mathbb{1}_{\leq \delta}\big(y_{[r]}, z_{[k-r]}, x\big) \mathrm{d}\mu^k(y_{[k]}) \mathrm{d}\mu^{k-r}(z_{[k-r]}) \mathrm{d}\mu(x) \\
&= \sum_{r=0}^{k} \mathbf{C}(k,r) \int_{W^{2k-r+1}} \mathbb{1}_{\leq \delta}\big(y_{[k]}, x\big) \mathbb{1}_{\leq \delta}\big(y_{[r]}, z_{[k-r]}, x\big) \mathrm{d}\mu^k(y_{[k]}) \mathrm{d}\mu^{k-r}(z_{[k-r]}) \mathrm{d}\mu(x)
\end{aligned}$$

Since x is used in both kernels we rename the variables $y_i \to y_{i+1}$ for all $i \in \{0, \dots, k\}$ and $x \to y_0$ and shift the index r of the sum to obtain

$$\begin{aligned}
&\int\limits_W \mathbb{E}\big[(D_x F_k)^2\big] \mathrm{d}\mu(x) \\
&= \sum_{r=0}^{k} \mathbf{C}(k,r) \int\limits_{W^{2k-r+1}} \mathbb{1}_{\leq\delta}\big(y_{[k+1]}\big) \mathbb{1}_{\leq\delta}\big(y_{[r+1]}, z_{[k-r]}\big) \mathrm{d}\mu^{k+1}(y_{[k+1]}) \mathrm{d}\mu^{k-r}(z_{[k-r]}) \\
&= \sum_{r=1}^{k+1} \mathbf{C}(k,r-1) \int\limits_{W^{2(k+1)-r}} \mathbb{1}_{\leq\delta}\big(y_{[k+1]}\big) \mathbb{1}_{\leq\delta}\big(y_{[r]}, z_{[k+1-r]}\big) \mathrm{d}\mu^{k+1}(y_{[k+1]}) \mathrm{d}\mu^{k+1-r}(z_{[k+1-r]}) \\
&= \sum_{r=1}^{k+1} \tfrac{\mathbf{C}(k,r-1)\mathbf{C}(k+1,r)}{\mathbf{C}(k+1,r)} \int\limits_{W^{2(k+1)-r}} \mathbb{1}_{\leq\delta}\big(y_{[k+1]}\big) \\
&\qquad\qquad \times \mathbb{1}_{\leq\delta}\big(y_{[r]}, z_{[k+1-r]}\big) \mathrm{d}\mu^{k+1}(y_{[k+1]}) \mathrm{d}\mu^{k+1-r}(z_{[k+1-r]}).
\end{aligned}$$

Defining

$$\mathbf{D}^{\mathrm{dom}}(k) = \max_{r=1}^{k+1} \frac{\mathbf{C}(k,r-1)}{\mathbf{C}(k+1,r)}$$

the first claim follows immediately. Using Lemma 4.12 completes the proof. □

We proceed to derive the upper bounds for γ_1, γ_2 and $\gamma_{3,N}$ resp. $\gamma_{3,P}$.

Lemma 4.14. *For all $d \geq 1$ and all $k \geq 1$ the error terms for the k-simplex counting functional in the Malliavin-Stein limit theorem are bounded by*

$$\gamma_1(F_k) \ll 4^d t (t\delta^d)^{k+\frac{3}{2}} (k+1)^{4d} \sum_{q=0}^{12k-6} (t\delta^d)^{\frac{q}{4}},$$

$$\gamma_2(F_k) \ll 4^d t (t\delta^d)^{k+1} k^{4d} \sum_{q=0}^{6k-6} (t\delta^d)^{\frac{q}{2}},$$

$$\gamma_{3,P}(F_k) \ll t (t\delta^d)^{k+\frac{1}{2}} (k+1)^{3d} \sum_{q=0}^{4k-1} (t\delta^d)^{\frac{q}{2}}.$$

The error terms for the standardized k-simplex counting functional in the Malliavin-Stein limit theorem are bounded by

$$\gamma_1(\widetilde{F_k}) \ll (\mathbb{V}[F_k])^{-2} 4^d t (t\delta^d)^{k+\frac{3}{2}} (k+1)^{4d} \sum_{q=0}^{12k-6} (t\delta^d)^{\frac{q}{4}},$$

$$\gamma_2(\widetilde{F_k}) \ll (\mathbb{V}[F_k])^{-2} 4^d t (t\delta^d)^{k+1} k^{4d} \sum_{q=0}^{6k-6} (t\delta^d)^{\frac{q}{2}},$$

$$\gamma_{3,N}(\widetilde{F_k}) \ll (\mathbb{V}[F_k])^{-\frac{3}{2}} \mathbf{C} t (t\delta^d)^{k} (k+1)^{3d} \sum_{q=0}^{2k} (t\delta^d)^{q}.$$

Proof: Since the Malliavin difference operator is linear and invariant under addition of a constant we have $D_x \widetilde{F_k} = D_x F_k / \sqrt{\mathbb{V}[F_k]}$. Therefore we can calculate the estimates for F_k first and re-scale them later to obtain the results for $\widetilde{F_k}$. Fix $x_1, x_2, x_3 \in W$. Using (4.3) and (4.5) we bound the integrand of γ_1 by

$$\begin{aligned}
&\mathbb{E}\big[(D^2_{x_1,x_3}F)^4\big]\mathbb{E}\big[(D^2_{x_2,x_3}F)^4\big]\mathbb{E}\big[(D_{x_1}F)^4\big]\mathbb{E}\big[(D_{x_2}F)^4\big] \\
&\ll \mathbf{D}_4(k)^2\mathbf{D}_4'(k)^2 \mathbb{1}_{\le\delta}(x_1,x_3)\mathbb{1}_{\le\delta}(x_2,x_3) \sum_{q_1,q_2=k-1}^{4(k-1)} \sum_{q_3,q_4=k}^{4k} (t\delta^d)^{q_1+q_2+q_3+q_4} \\
&\quad \times ((k+1)k)^{2d}\Big(\Big(\tfrac{q_1-(k-1)}{3}+1\Big)\Big(\tfrac{q_2-(k-1)}{3}+1\Big)\Big(\tfrac{q_3-k}{3}+1\Big)\Big(\tfrac{q_4-k}{3}+1\Big)\Big)^{3d} \\
&\ll \mathbb{1}_{\le\delta}(x_1,x_3)\mathbb{1}_{\le\delta}(x_2,x_3) \sum_{q_1,q_2=k-1}^{4(k-1)} \sum_{q_3,q_4=k}^{4k} (t\delta^d)^{q_1+q_2+q_3+q_4}(k+1)^{16d},
\end{aligned}$$

and the integrand of γ_2 by

$$\begin{aligned}
&\mathbb{E}\big[(D^2_{x_1,x_3}F)^4\big]\mathbb{E}\big[(D^2_{x_2,x_3}F)^4\big] \\
&\ll \mathbf{D}_4'(k)^2 \mathbb{1}_{\le\delta}(x_1,x_3)\mathbb{1}_{\le\delta}(x_2,x_3) \\
&\qquad \times \sum_{q_1,q_2=k-1}^{4(k-1)} (t\delta^d)^{q_1+q_2} k^{2d}\Big(\Big(\tfrac{q_1-(k-1)}{3}+1\Big)\Big(\tfrac{q_2-(k-1)}{3}+1\Big)\Big)^{3d} \\
&\ll \mathbb{1}_{\le\delta}(x_1,x_3)\mathbb{1}_{\le\delta}(x_2,x_3) \sum_{q_1,q_2=k-1}^{4(k-1)} (t\delta^d)^{q_1+q_2} k^{8d}.
\end{aligned}$$

Note that

$$\int_{W^3} \mathbb{1}_{\le\delta}(x_1,x_3)\mathbb{1}_{\le\delta}(x_2x_3)\mathrm{d}\mu^3(x_1,x_2,x_3) \le \Lambda_d(W)t^3\delta^{2d}4^d = t^3\delta^{2d}4^d.$$

Since $\sqrt[n]{x}$ is subadditive for all $x > 0$ and $n > 1$, it follows that

$$\begin{aligned}
\gamma_1(F_k) &\ll \int_{W^3} \mathbb{1}_{\le\delta}(x_1,x_3)\mathbb{1}_{\le\delta}(x_2,x_3)\mathrm{d}\mu^3(x_1,x_2,x_3) \\
&\qquad \times \sum_{q_1,q_2=k-1}^{4(k-1)} \sum_{q_3,q_4=k}^{4k} (t\delta^d)^{\frac{q_1+q_2+q_3+q_4}{4}}(k+1)^{4d} \\
&\ll 4^d t^3\delta^{2d}(k+1)^{4d} \sum_{q=4k-2}^{16k-8} (t\delta^d)^{\frac{q}{4}},
\end{aligned}$$

and

$$\begin{aligned}
\gamma_2(F_k) &\ll \int_{W^3} \mathbb{1}_{\le\delta}(x_1,x_3)\mathbb{1}_{\le\delta}(x_2,x_3)\mathrm{d}\mu^3(x_1,x_2,x_3) \sum_{q_1,q_2=k-1}^{4(k-1)} (t\delta^d)^{\frac{q_1+q_2}{2}} k^{4d} \\
&\ll 4^d t^3\delta^{2d}k^{4d} \sum_{q=2k-2}^{8k-8} (t\delta^d)^{\frac{q}{2}},
\end{aligned}$$

where we have simplified the sums using only the possible exponents of $(t\delta^d)$, since the number of terms with the corresponding exponent in the double sum does only depend on k. Shifting the indices of the sums such that the summations start at

$q = 0$ establishes the γ_1 and γ_2 bounds for F_k and re-scaling yields the corresponding bounds for $\widetilde{F_k}$.

Fix $x \in W$. Using (4.4) and (4.3) we bound the integrand of $\gamma_{3,P}$ by

$$\begin{aligned}
&\mathbb{E}\Big[((D_xF_k)(D_xF_k-1))^2\Big]\mathbb{E}\big[(D_xF_k)^2\big] \\
&\ll \mathbf{D}_4^*(k) \sum_{q_1=k+1}^{4k} \sum_{q_2=k}^{2k} (t\delta^d)^{q_1+q_2}(k+1)^{2d}\Big(\tfrac{q_1-k}{3}+1\Big)^{3d}(q_2-k+1)^d \\
&\ll \sum_{q_1=k+1}^{4k} \sum_{q_2=k}^{2k} (t\delta^d)^{q_1+q_2}(k+1)^{6d}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\gamma_{3,P}(F_k) &\ll \int\limits_W 1\mathrm{d}\mu(x) \sum_{q_1=k+1}^{4k} \sum_{q_2=k}^{2k} (t\delta^d)^{\frac{q_1+q_2}{2}}(k+1)^{3d} \\
&\ll t(k+1)^{3d} \sum_{q=2k+1}^{6k} (t\delta^d)^{\frac{q}{2}}.
\end{aligned}$$

Using (4.3) together with $D_xF_k \geq 0$, we bound the integrand of $\gamma_{3,N}$ by

$$\begin{aligned}
\mathbb{E}\big|(D_xF)^3\big| &\ll \mathbf{D}_3(k) \sum_{q=k}^{3k} (t\delta^d)^q(k+1)^d\Big(\tfrac{q-k}{2}+1\Big)^{2d} \\
&\ll \sum_{q=k}^{3k} (t\delta^d)^q(k+1)^{3d}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\gamma_{3,N}(F_k) &\ll \int\limits_W 1\mathrm{d}\mu(x) \sum_{q=k}^{3k} (t\delta^d)^q(k+1)^{3d} \\
&\ll t(k+1)^{3d} \sum_{q=k}^{3k} (t\delta^d)^q.
\end{aligned}$$

Re-scaling this bound for $\widetilde{F_k}$ completes the proof. □

We are now in a position to use the Malliavin-Stein method, in particular Theorem 2.6 and Theorem 2.7, to prove our main results, the central limit theorem and the Poisson limit theorem for the k-simplex counting functional.

Proof of Theorem 1.4: Assume that the expectation of F_k tends to infinity (1.1), i.e.

$$\lim_{d\to\infty} \frac{1}{(k+1)!} t(t\delta^d)^k(k+1)^d = \infty.$$

Lemma 4.12 yields $\mathbb{V}[F_k] \to \infty$ and Lemma 4.13 gives $\widetilde{F_k} \in \mathrm{dom}(D)$. This enables us, to use the Malliavin-Stein method to derive a bound on the Wasserstein-distance between $\widetilde{F_k}$ and a standard Gaussian distributed random variable $\mathcal{N}(0,1)$ from Theorem 2.6. We have to distinguish the following three cases based on the limit of $(t\delta^d)$:

Case 1: If $(t\delta^d) \to 0$, the variance is bounded from below by

$$t(t\delta^d)^k(k+1)^d \ll \mathbb{V}[F_k].$$

It follows that

$$\begin{aligned}\gamma_1(\widetilde{F_k}) &\ll 4^d(k+1)^{2d}t^{-1}(t\delta^d)^{\frac{3}{2}-k},\\ \gamma_2(\widetilde{F_k}) &\ll 4^d k^{2d}\left(\tfrac{k}{k+1}\right)^{2d}t^{-1}(t\delta^d)^{1-k},\\ \gamma_{3,N}(\widetilde{F_k}) &\ll (k+1)^{\frac{3d}{2}}t^{-\frac{1}{2}}(t\delta^d)^{-\frac{k}{2}},\end{aligned}$$

and further

$$\sqrt{\gamma_1}+\sqrt{\gamma_2}+\gamma_{3,N} \ll \begin{cases} t^{-\frac{1}{2}}(t\delta^d)^{-\frac{k}{2}}2^d(k+1)^d, & k\le 3,\\ t^{-\frac{1}{2}}(t\delta^d)^{-\frac{k}{2}}(k+1)^{\frac{3d}{2}}, & \ge 3.\end{cases}$$

Since Lemma 1.2 yields $\mathbb{E}[F_k] \ll t(t\delta^d)^k(k+1)^d$ it follows that

$$t^{-\frac{1}{2}}(t\delta^d)^{-\frac{k}{2}}(k+1)^{-\frac{d}{2}} \ll (\mathbb{E}[F_k])^{-\frac{1}{2}},$$

and thus for $k \le 3$ we have

$$\begin{aligned}\sqrt{\gamma_1}+\sqrt{\gamma_2}+\gamma_{3,N} &\ll t^{-\frac{1}{2}}(t\delta^d)^{-\frac{k}{2}}(k+1)^{-\frac{d}{2}} \times (k+1)^{\frac{d}{2}}2^d(k+1)^d\\ &\ll (\mathbb{E}[F_k])^{-\frac{1}{2}} \times 2^d(k+1)^{\frac{3d}{2}},\end{aligned}$$

and for $k \ge 3$ we have

$$\begin{aligned}\sqrt{\gamma_1}+\sqrt{\gamma_2}+\gamma_{3,N} &\ll t^{-\frac{1}{2}}(t\delta^d)^{-\frac{k}{2}}(k+1)^{-\frac{d}{2}} \times (k+1)^{\frac{d}{2}}(k+1)^{\frac{3d}{2}}\\ &\ll (\mathbb{E}[F_k])^{-\frac{1}{2}} \times (k+1)^{2d}.\end{aligned}$$

which yields our claim using the Malliavin-Stein bound (2.2).

Case 2: If $(t\delta^d) \to c \in (0,\infty)$, the variance is bounded from below by

$$t(k^2+2k)^d \ll \mathbb{V}[F_k]$$

It follows that

$$\begin{aligned}\gamma_1(\widetilde{F_k}) &\ll 4^d t^{-1}\left(1+\tfrac{1}{k^2+2k}\right)^{2d},\\ \gamma_2(\widetilde{F_k}) &\ll 4^d t^{-1}\left(\tfrac{k^2}{k^2+2k}\right)^{2d},\\ \gamma_{3,N}(\widetilde{F_k}) &\ll t^{-\frac{1}{2}}\left(1+\tfrac{1}{k^2+2k}\right)^{\frac{3d}{2}},\end{aligned}$$

and further

$$\sqrt{\gamma_1}+\sqrt{\gamma_2}+\gamma_{3,N} \ll 2^d t^{-\frac{1}{2}}\left(1+\tfrac{1}{k^2+2k}\right)^{d},$$

which yields our claim using the Malliavin-Stein bound (2.2).

Case 3: If $(t\delta^d) \to \infty$, the variance is bounded from below by

$$t(t\delta^d)^{2k}(k^2+2k)^d \ll \mathbb{V}[F_k].$$

It follows that

$$\gamma_1(\widetilde{F_k}) \ll 4^d t^{-1}\Big(1+\tfrac{1}{k^2+2k}\Big)^{2d},$$
$$\gamma_2(\widetilde{F_k}) \ll 4^d t^{-1}(t\delta^d)^{-2}\Big(\tfrac{k^2}{k^2+2k}\Big)^{2d},$$
$$\gamma_{3,N}(\widetilde{F_k}) \ll t^{-\frac{1}{2}}\Big(1+\tfrac{1}{k^2+2k}\Big)^{\frac{3d}{2}},$$

and further

$$\begin{aligned}\sqrt{\gamma_1}+\sqrt{\gamma_2}+\gamma_{3,N} &\ll 2^d t^{-\frac{1}{2}}\Big(1+\tfrac{1}{k^2+2k}\Big)^{d} + t^{-\frac{1}{2}}\Big(1+\tfrac{1}{k^2+2k}\Big)^{\frac{3d}{2}}\\ &\ll 2^d t^{-\frac{1}{2}}\Big(1+\tfrac{1}{k^2+2k}\Big)^{d}.\end{aligned}$$

Using the Malliavin-Stein bound (2.2) completes the proof. □

Proof of Theorem 1.5: Assume that the expectation of F_k converges to a positive constant (1.2), i.e.

$$\lim_{d\to\infty}\frac{1}{(k+1)!}t(t\delta^d)^k(k+1)^d = \theta \in (0,\infty).$$

Lemma 4.12 yields $\mathbb{V}[F_k] \to \theta$ and Lemma 4.13 gives $F_k \in \mathrm{dom}(D)$. This enables us, to use the Malliavin-Stein method to derive a bound on the total variation distance between F_k and a Poisson-distributed random variable $\mathcal{P}(\theta)$ from Theorem 2.7. We note that our assumption implies $(t\delta^d) \to 0$ and $t(t\delta^d)^k(k+1)^d \to (k+1)!\theta$.

Therefore

$$\gamma_1(F_k) \ll 4^d (t\delta^d)^{\frac{3}{2}}(k+1)^{3d},$$
$$\gamma_2(F_k) \ll 4^d (t\delta^d)k^{3d}\Big(\tfrac{k}{k+1}\Big)^{d},$$
$$\gamma_{3,P}(F_k) \ll (t\delta^d)^{\frac{1}{2}}(k+1)^{2d},$$

and further

$$\sqrt{\gamma_1}+\sqrt{\gamma_2}+\gamma_{3,P} \ll \begin{cases}(t\delta^d)^{\frac{1}{2}}2^d(k+1)^{\frac{3d}{2}}, & k\le 3,\\ (t\delta^d)^{\frac{1}{2}}(k+1)^{2d}, & k\ge 3.\end{cases}$$

It follows from the Malliavin-stein bound (2.3) that

$$\begin{aligned}\mathrm{d}_{TV}(F_k,\mathcal{P}(\theta)) &\ll |\mathbb{E}[F_k]-\theta| + |\mathbb{V}[F_k]-\theta| + \begin{cases}(t\delta^d)^{\frac{1}{2}}2^d(k+1)^{\frac{3d}{2}}, & k\le 3,\\ (t\delta^d)^{\frac{1}{2}}(k+1)^{2d}, & k\ge 3.\end{cases}\\ &\ll |\mathbb{E}[F_k]-\theta| + |\mathbb{V}[F_k]-\theta| + \begin{cases}t^{-\frac{1}{2k}}(k+1)^{\frac{d(3k-1)}{2k}}2^d, & k\le 3,\\ t^{-\frac{1}{2k}}(k+1)^{\frac{d(4k-1)}{2k}}, & k\ge 3.\end{cases}\end{aligned}$$

which yields our claim. □

Acknowledgements

The author would like to thank the anonymous referees for many useful comments and corrections. Further the author wishes to express his thanks to Alexander Nover for helpful comments concerning simplicial complexes.

References

Bobrowski, O. and Kahle, M. Topology of random geometric complexes: a survey. *J. Appl. Comput. Topol.*, **1** (3-4), 331–364 (2018). MR3975557.

Bourguin, S. and Peccati, G. The Malliavin-Stein method on the Poisson space. In *Stochastic analysis for Poisson point processes*, volume 7 of *Bocconi Springer Ser.*, pp. 185–228. Bocconi Univ. Press (2016). MR3585401.

Grygierek, J. Poisson fluctuations for edge counts in high-dimensional random geometric graphs. *ArXiv Mathematics e-prints* (2019). arXiv: 1905.11221.

Grygierek, J. and Thäle, C. Gaussian fluctuations for edge counts in high-dimensional random geometric graphs. *Statist. Probab. Lett.*, **158**, 108674, 10 (2020). MR4038665.

Last, G., Peccati, G., and Schulte, M. Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Probab. Theory Related Fields*, **165** (3-4), 667–723 (2016). MR3520016.

Last, G. and Penrose, M. *Lectures on the Poisson process*, volume 7 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, Cambridge (2018). ISBN 978-1-107-45843-7; 978-1-107-08801-6. MR3791470.

Li, S. Concise formulas for the area and volume of a hyperspherical cap. *Asian J. Math. Stat.*, **4** (1), 66–70 (2011). MR2813331.

Munkres, J. R. *Elements of algebraic topology*. Addison-Wesley Publishing Company, Menlo Park, CA (1984). ISBN 0-201-04586-9. MR755006.

Peccati, G. and Reitzner, M., editors. *Stochastic analysis for Poisson point processes*, volume 7 of *Bocconi & Springer Series*. Bocconi University Press (2016). MR3444831.

Penrose, M. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford (2003). ISBN 0-19-850626-0. MR1986198.

Reitzner, M. and Schulte, M. Central limit theorems for U-statistics of Poisson point processes. *Ann. Probab.*, **41** (6), 3879–3909 (2013). MR3161465.

Stanley, R. P. *Combinatorics and commutative algebra*, volume 41 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, second edition (1996). ISBN 0-8176-3836-9. MR1453579.

Ziegler, G. M. *Lectures on polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, New York (1995). ISBN 0-387-94365-X. MR1311028.