TKK Reports in Information and Computer Science
Espoo 2008 TKK-ICS-R9

A BOUNDARY CORRECTED EXPANSION OF THE MOMENTS OF NEAREST NEIGHBOR DISTRIBUTIONS

Elia Liitiäinen, Francesco Corona and Amaury Lendasse

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

TKK Reports in Information and Computer Science
Espoo 2008 TKK-ICS-R9

A BOUNDARY CORRECTED EXPANSION OF THE MOMENTS OF NEAREST NEIGHBOR DISTRIBUTIONS

Elia Liitiäinen, Francesco Corona and Amaury Lendasse

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Information and Computer Science

Teknillinen korkeakoulu
Informaatio- ja luonnontieteiden tiedekunta
Tietojenkäsittelytieteen laitos

Distribution:
Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Information and Computer Science
P.O.Box 5400
FI-02015 TKK
FINLAND
URL: http://ics.tkk.fi
Tel. +358 9 451 1
Fax +358 9 451 3369
E-mail: series@ics.tkk.fi

ISBN 978-951-22-9610-1 (Print)
ISBN 978-951-22-9611-8 (Online)
ISSN 1797-5034 (Print)
ISSN 1797-5042 (Online)
URL: http://lib.tkk.fi/Reports/2008/isbn9789512296118.pdf

TKK ICS
Espoo 2008

Abstract: In this paper, the moments of nearest neighbor distance distributions are examined. While the asymptotic form of such moments is well-known, the boundary effect has this far resisted a rigorous analysis. Our goal is to develop a new technique that allows a closed-form high order expansion, where the boundaries are taken into account up to the first order. The resulting theoretical predictions are tested via simulations and found to be much more accurate than the first order approximation obtained by neglecting the boundaries.

While our results are of theoretical interest, they definitely also have important applications in statistics and physics. As a concrete example, we mention estimating Renyi entropies of probability distributions. Moreover, the algebraic technique developed may turn out to be useful in other, related problems including estimation of the Shannon differential entropy.

Keywords: nearest neighbor, boundary, asymptotics, renyi entropy

CONTENTS

1 INTRODUCTION

In this paper, we examine the moments of nearest neighbor distance distributions. We assume that $(X_i)_{i=1}^M$ is an independent identically distributed (i.i.d.) sample on a probability space $(\Omega, \mathcal{F}, P)$ taking values in an open set $\mathcal{C} \subset \Re^m$ with the common distribution given by a density p. Denoting by d_k the distance between X_i (for some $i > 0$) and its k-th nearest neighbor in the Euclidean metric, we consider $E[d_k^\alpha]$.

The starting point of our work is [2], where it is shown that

$$M^{\alpha/m} E[d_k^\alpha] \to V_m^{-\alpha/m} \frac{\Gamma(k+\alpha/m)}{\Gamma(k)} \int_{\mathcal{C}} p(x)^{1-\alpha/m} dx \qquad (1)$$

under some regularity conditions as $M \to \infty$. The result was obtained by showing that the effect of points close to the boundary of $\mathcal{C}$ can be neglected in the asymptotic limit if a term of order $O(M^{-1/m-\alpha/m+\rho})$ (for an arbitrarily small $\rho > 0$) is accepted. In this paper, our goal is to show that actually the approximation (1) can be improved by taking the boundary effect into account. Our technique leads to predictions that are one order of magnitude more accurate than (1) at the expense of additional regularity assumptions. Moreover, the technique proposed here can probably be applied in other contexts as for example in the analysis of particles in thin plates and films [10].

In addition to [2], there exists a rather large amount of literature on nearest neighbor distances. For example, in [8, 7, 14], asymptotic results were obtained for nearest neighbor graphs as a special case of a more abstract setting. More specific results can be found in [9] and in the aforementioned [2]. The nearest neighbor graph is an important topic of research on computational geometry; often it is analyzed as a special case of a more abstract framework.

In addition to being of mathematical interest, results of the form (1) have some rather concrete applications. One of these follows from the fact that the term

$$\int_{\mathcal{C}} p(x)^{1-\alpha/m} dx$$

is closely related to the Renyi entropies of p. Thus it is not a surprise that $E[d_k^\alpha]$ has been used to estimate such entropies, see for example [4, 13] on methods exploiting the theory of nearest neighbor (and more general) graphs.

Another important functional is the Shannon differential entropy

$$\int_{\mathcal{C}} p(x) \log p(x) dx.$$

The differential entropy is a widely discussed topic in the literature and nearest neighbor estimators have turned out to be useful [5]. We believe that our results and proof techniques will be useful for researchers working on the Shannon entropy as well.

In addition to entropy estimators, nearest neighbors distributions are closely related to other nonparametric estimators like the Gamma test [3] and nonparametric statistics in general. Finally, we mention that nearest neighbor distributions have received some attention in physics as well [12, 11].

The paper is divided into 4 sections. In Section 2 we present our main result and some numerical simulations to test our theoretical predictions. In Sections 3 and 4 we prove some auxiliary results. The theory in this part is of technical nature, but stays at an intuitive and concrete level. Finally, in Section 5 we give proofs to the theorems in Section 2.

2 MAIN RESULTS

2.1 Basic Definitions

To fix some notation, A^T will denote the transpose of a matrix A. The inner product between two vectors u, v is denoted by $\langle u, v\rangle$. We will constantly use the standard $O(\cdot)$ notation typically to denote higher order terms that do not need to be analyzed accurately. By $B(x, r)$ we will denote the open ball with center x, radius r and volume $V_m r^m$. As a general notation, 0 means the zero element of a vector space.

The concept of nearest neighbors is common in the literature on computational geometry, machine learning and statistics. The nearest neighbor of the point X_i is defined simply as the point closest to it with respect to a similarity measure. Using the Euclidean metric, the formal definition is

$$N[i,1] = \text{argmin}_{1\leq j\leq M, j\neq i}\|X_i - X_j\|.$$

The k-th nearest neighbor is defined recursively as

$$N[i,k] = \text{argmin}_{1\leq j\leq M, j\neq i, N[i,1],\ldots,N[i,k-1]}\|X_i - X_j\|,$$

that is, the closest point after removal of the preceeding neighbors. The corresponding distances are defined as

$$d_{i,k} = \|X_i - X_{N[i,k]}\|.$$

Because our sample is i.i.d., we may fix $i = 1$ and use the shorthand d_k for $d_{1,k}$.

Let $\mathcal{H}$ be the half-plane of points with a positive first coordinate, that is,

$$\mathcal{H} = \{(s, s_{m-1}) | s > 0, s_{m-1} \in \Re^{m-1}\}$$

and set (λ denotes the Lebesgue measure)

$$h(r) = \lambda(B((1, 0, \ldots, 0), r) \cap \mathcal{H}).$$

The function h has a rather complicated form due to the cutoff at the boundary. For $r < 1$, $h(r)$ is simply $V_m r^m$; for $r > 1$ this is not true but we still have

$$\frac{1}{2}V_m r^m \leq h(r) \leq V_m r^m. \qquad (2)$$

h is increasing and Lipschitz continuous; thus it is almost everywhere differentiable with the derivative h'.

For any $x \in \Re^m$, the distance between x and and the boundary of $\mathcal{C}$ (denoted by $\partial\mathcal{C}$ in the standard notation), is given by

$$\rho(x, \partial\mathcal{C}) = \inf_{y \in \partial\mathcal{C}} \|x - y\|.$$

For future use, we define $\partial_r\mathcal{C}$ as the set of points for which $\rho(x, \partial\mathcal{C}) \leq r$ and $x \in \mathcal{C}$.

We will need some basic concepts from differential geometry. A nonempty subset $\mathcal{M} \subset \Re^m$ is called a twice differentiable manifold, if for each $x \in \mathcal{M}$ there exists $\delta > 0$ and a twice differentiable homeomorphism

$$\phi : U \to \mathcal{M} \cap B(x, \delta)$$

with U an open subset of $\Re^{m-1}$ such that the Jacobian $J_y\phi$ has linearly independent columns for any $y \in U$. Recall that a homeomorphism is a bijection with both ϕ and ϕ^{-1} continuous.

We will constantly need surface integrals over smooth manifolds. An infinitesimal surface element is denoted by dS, thus the integral over a surface $\mathcal{M}$ looks like

$$\int_{\mathcal{M}} f(x) dS.$$

2.2 Expansions of Nearest Neighbor Distances

Our main result is based on the following assumptions that require regularity both from p and the boundary $\partial\mathcal{C}$. To ensure that there is no boundary points inside $\mathcal{C}$, we state the condition

$$\text{interior}[\bar{\mathcal{C}}] = \mathcal{C}, \tag{3}$$

where $\bar{\mathcal{C}} = \mathcal{C} \cup \partial\mathcal{C}$ is the closure of $\mathcal{C}$.

A1) $\mathcal{C}$ is a bounded open set, Equation (3) holds and the boundary $\partial\mathcal{C}$ is a closed $m-1$ dimensional twice continuously differentiable manifold with $m \geq 2$ (consequently, $\partial\mathcal{C}$ is also a compact set).

A2) There exists a constant $c_p > 0$ such that $c_p^{-1} < p(x) < c_p$ on the closure $\bar{\mathcal{C}}$. Moreover, we assume that for a constant $L > 0$ and some $1 < \xi \leq 2$, the gradient ∇p satisfies

$$\|\nabla p(x) - \nabla p(y)\| \leq L\|x - y\|^{\xi - 1}$$

for all $x, y \in \mathcal{C}$ and

$$|p(x) - p(y)| \leq L\|x - y\|$$

for $x, y \in \bar{\mathcal{C}}$.

Our main result is the following asymptotic expansion of $E[d_k^\alpha]$.

Theorem 1. *Suppose that (A2) holds and alternatively $\mathcal{C}$ is a polytope or (A1) holds. Then for $0 < \gamma < 1$, $M > 2k$ and $\alpha > 0$,*

$$\begin{aligned} E[d_k^\alpha] &= V_m^{-\alpha/m}\frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)}\int_{\mathcal{C}} p(x)^{1-\alpha/m}dx \\ &\quad +(D - V_m^{-\alpha/m-1/m})\frac{\Gamma(k+\alpha/m+1/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m+1/m)}\int_{\partial\mathcal{C}} p(x)^{1-\alpha/m-1/m}dS \\ &\quad +R \end{aligned} \tag{4}$$

with

$$D = \frac{1}{m}\int_0^1 a^{-\alpha-2}h(a^{-1})^{-\alpha/m-1/m-1}h'(a^{-1})da.$$

The remainder term is bounded by ($c, K > 0$ are constants independent of M, γ and k)

$$|R| \le cM^k e^{-K\gamma^m M} + c(\gamma^{2+\alpha} + \gamma^{2+\alpha+m}M + \gamma^{\xi+\alpha} + \gamma^2 k^{\alpha/m}M^{-\alpha/m}).$$

With the choice $\gamma = M^{-1/m}\log M$ we have (for a fixed k)

$$R = O(M^{-\xi/m-\alpha/m}\log^{2+\alpha+m} M).$$

It is worth noticing that for any $\sigma > 0$,

$$\frac{\Gamma(M)}{\Gamma(M+\sigma)} = M^{-\sigma}(1 + O(M^{-1}))$$

as demonstrated in [2]. The expansion of Theorem 1 has a rather interesting form. The boundary correction is contained in the second term at the right side of Equation (4), whereas the first term is the lower order approximation of Theorem 5.4 in [2]. Consequently, the remainder term is one order of magnitude smaller than that in [2]. The correction term is surprisingly simple, as one would have expected highly complicated correction terms due to the nonlinear cutoff at the boundary.

By requiring (A1), we ensure that the boundary $\partial\mathcal{C}$ can be approximated locally by a plane. The proof proceeds by examining planar boundaries and using (A1) to transfer such results to sets with more general boundaries. It seems likely that (A1) and (A2) could be weakened considerably. One drawback of our assumptions is that sets with a non-smooth boundary are excluded. Intuitively it seems that in many cases such singularities do not pose a real problem if they are asymptotically neglible in the limit $M \to \infty$. To demonstrate the application for non-smooth sets, the theorem is stated also for polytopes.

From the practical point of view, the main problem of Theorem 1 is the evaluation of h, h' and the term D. Luckily, numerical evaluation is not too difficult as long as m is not too large.

Example 1. *When $m = 3$, the function h is given by*

$$h(r) = \frac{4}{3}\pi r^3$$

when $r < 1$ *and*

$$h(r) = \frac{2}{3}\pi r^3 - \frac{1}{3}\pi + \pi r^2$$

otherwise. As a concrete example, let us analyze uniformly distributed points in the unit ball. In such a case $p = V_3^{-1}$ *and a numerical evaluation gives*

$$D \approx 0.42.$$

Setting $k = 1$*, Equation (4) takes the form*

$$\begin{aligned} E[d_1^\alpha] \;\approx\; & \Gamma(1+\alpha/3)M^{-\alpha/3} + 3DV_3^{\alpha/3+1/3}\Gamma(4/3+\alpha/3)M^{-\alpha/3-1/3} \\ & -3\Gamma(4/3+\alpha/3)M^{-\alpha/3-1/3}. \end{aligned} \tag{5}$$

Correspondingly, we may calculate the expansion for the unit cube.

Example 2. *If* $\mathcal{C} = (0,1)^3$ *and* $p = 1$*, we may approximate*

$$\begin{aligned} E[d_1^\alpha] \;\approx\; & V^{-\alpha/3}\Gamma(1+\alpha/3)M^{-\alpha/3} + 6D\Gamma(4/3+\alpha/3)M^{-\alpha/3-1/3} \\ & -6V_3^{-\alpha/3-1/3}\Gamma(4/3+\alpha/3)M^{-\alpha/3-1/3}. \end{aligned} \tag{6}$$

2.3 Simulations

We demonstrate the expansions in Examples 1 and 2 via a numerical simulation. The number of samples is increased from 100 to 5000 in steps of 100. For each number of samples, 1000 different configurations are generated and the expected first nearest neighbor distance is estimated as the average of the 1000 different values.

The first experiment involves points uniformly distributed on the unit ball, whereas in the second one, the points are uniform on the unit cube (Examples 1 and 2 respectively). As we want to investigate the effect of the higher order terms, we compare the experimental result also to the approximation

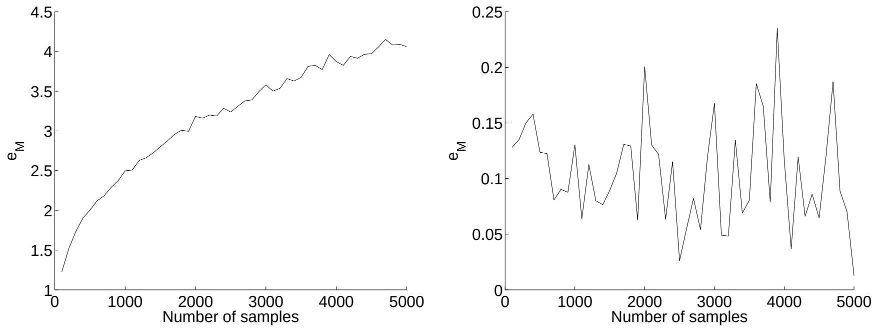
$$E[d_1] \approx V_m^{-1/m}\frac{\Gamma(k+1/m)}{\Gamma(k)}\int_{\mathcal{C}} p(x)^{1-1/m}dx M^{-1/m} \tag{7}$$

in Equation (1).

The results of the unit ball experiment are in Figure 1, whereas those of the second are plotted in Figure 2. As a measure of performance, we use

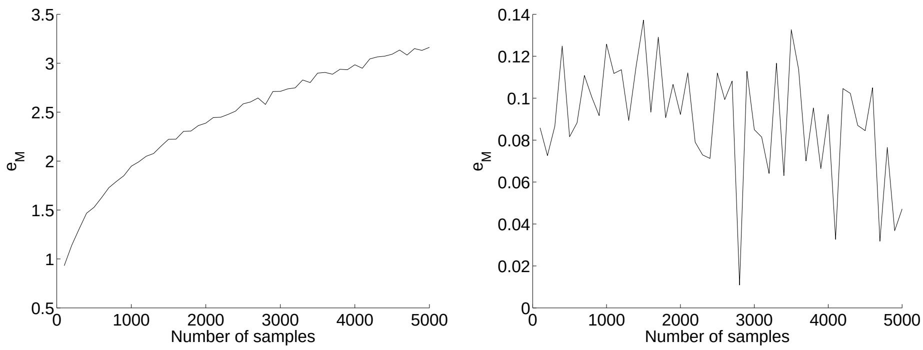
$$e_M = M|E_{experimental}[d_1] - E_{theory}[d_1]|.$$

The results support our theoretical analysis well despite some random fluctuation. The higher order approximation yields estimates that are one order of magnitude more accurate than those predicted by Equation (7), whereas the predictions of Theorem 1 seem to have an error of order $O(M^{-1})$ in this case.



(a) (b)

Figure 1: Results for uniformly distributed points on the unit ball: the prediction error of the approximation (1) in Figure (a) and those of Equation (5) in Figure (b).



(a) (b)

Figure 2: Results for uniformly distributed points on the unit cube: the prediction error of the approximation (1) in Figure (a) and those of Equation (6) in Figure (b).

3 PROPERTIES OF THE SET $\mathcal{C}$

3.1 Linearization of the Boundary

We need some additional notation from differential geometry. For $x \in \partial\mathcal{C}$, we denote by $n(x)$ the normal of the manifold $\partial\mathcal{C}$, orthogonal to the plane spanned by the columns of the Jacobian $J_{\phi^{-1}(x)}\phi$ of some parametrization ϕ. Notice that there are two possible directions for the normal; later in this section we will show that an outer normal is a meaningful concept when (A1) holds solving the orientation problem. Meanwhile, the notation means either of the two possibilities. In any case the normal is continuous in the sense that regardless of the orientation,

$$\min\{\|n(x) - n(x_n)\|, \|n(x) + n(x_n)\|\} \to 0 \tag{8}$$

when $x_n \to x$.

We define the sets

$$T_x = \{x + y : y^T n(x) = 0\}$$

and

$$\mathcal{U}_x = \{x + y : y^T n(x) \leq 0\}$$

corresponding to the tangent plane and the corresponding half-space. Finally, we need the line segments

$$A(x, r) = \{x - sn(x) : \ s \in (0, r]\}, \tag{9}$$

where $x \in \partial\mathcal{C}$ is a point on the boundary and r is a (possibly negative) real number.

Under Assumption (A1), it is intuitively clear that in a small neighborhood of a point on $\partial\mathcal{C}$, $\partial\mathcal{C}$ can be linearized and thus viewed as a plane. Here our goal is to use this idea to show that when $y = x - rn(x)$ for some $x \in \partial\mathcal{C}$ and r small enough, the set $\mathcal{C}$ in the expression

$$B(y, r) \cap \mathcal{C}$$

can be replaced by $\mathcal{U}_x$. Later in Section 4 this observation is necessary to analyze the nearest neighbor distribution close to the boundary as the exact shape of $\partial\mathcal{C}$ is unknown.

Before proceeding to the main result of this section, we need to prove some auxiliary results. The use of them will become clear later; however, the statements are quite intuitive.

Lemma 1. *Suppose that Assumption (A1) holds and let $(x_n)_{n=1}^{\infty} \subset \partial\mathcal{C}$ be a sequence converging to some $x \in \partial\mathcal{C}$. Then for any local parametrization $\phi : U \to B(x, \delta) \cap \partial\mathcal{C}$, there exists a constant $c_x > 0$ and an integer n_0 such that*

$$\|\phi^{-1}(x_n) - \phi^{-1}(x)\| \leq c_x \|x_n - x\|$$

when $n > n_0$.

Proof. Choose a parametrization $\phi : U \to B(x,\delta) \cap \partial\mathcal{C}$ with $\phi(0) = x$. There exists $\epsilon_1 > 0$ such that the closure $\bar{B}(0,\epsilon_1)$ is a subset of U. Notice that

$$(J_y\phi)^T J_y\phi$$

is a continuous matrix valued function with eigenvalues strictly above zero for each fixed $y \in U$ because the columns are linearly independent. This implies that there exists a constant $c > 0$ such that

$$\inf_{y\in\bar{B}(0,\epsilon_1),\|z\|=1} \|(J_y\phi)z\| \geq c. \tag{10}$$

Because ϕ is a homeomorphism, $\phi(B(0,\epsilon_1))$ contains a set of the form $B(x,\epsilon_2) \cap S$ for some $\epsilon_2 > 0$ and consequently $\phi^{-1}(x_n) \in B(0,\epsilon_1)$ for some n_0 and all $n > n_0$. Thus by equation (10) and the mean value theorem,

$$\|x_n - x\| = \|\phi(\phi^{-1}(x_n)) - \phi(\phi^{-1}(x))\| \geq c\|\phi^{-1}(x_n) - \phi^{-1}(x)\|$$

finishing the proof. □

Next we show that for any sequence $(x_n)_{n=1}^\infty$ converging to a point $x \in \partial\mathcal{C}$, the normal $n(x_n)$ is approximately orthogonal to the tangent plane at the point x once x_n is close to x.

Lemma 2. *For any sequence* $(x_n^{(1)}, x_n^{(2)})_{n=1}^\infty \subset \partial\mathcal{C} \times \partial\mathcal{C}$ *with* $\|x_n^{(1)} - x_n^{(2)}\| \to 0$*, Assumption (A1) implies that*

$$\sup_{n>0} \frac{n(x_n^{(1)})^T(x_n^{(1)} - x_n^{(2)})}{\|x_n^{(1)} - x_n^{(2)}\|^2} < \infty. \tag{11}$$

Proof. Let us make the counterassumption, that Equation (11) goes to infinity for the sequence $(x_n^{(1)}, x_n^{(2)})_{n=1}^\infty$. By compactness, we may assume that

$$(x_n^{(1)}, x_n^{(2)}) \to (x,x)$$

for some $x \in \partial\mathcal{C}$. Choose $\phi : U \to \partial\mathcal{C} \cap B(x,\delta)$ as a local parametrization around x and set

$$u_n^{(i)} = \phi^{-1}(x_n^{(i)}).$$

By Lemma 1,

$$u_n^{(1)}, u_n^{(2)} \to \phi^{-1}(x)$$

and using the fact that $(J_{u_n^{(1)}}\phi)^T n(x_n^{(1)}) = 0$, we have

$$\begin{aligned}
\frac{n(x_n^{(1)})^T(x_n^{(1)} - x_n^{(2)})}{\|x_n^{(1)} - x_n^{(2)}\|^2} &= n(x_n^{(1)})^T \frac{J_{u_n^{(1)}}\phi(u_n^{(1)} - u_n^{(2)}) + O(\|u_n^{(1)} - u_n^{(2)}\|^2)}{\|J_{\phi^{-1}(x)}\phi(u_n^{(1)} - u_n^{(2)}) + o(\|u_n^{(1)} - u_n^{(2)}\|)\|^2} \\
&= O(1),
\end{aligned}$$

leading to a contradiction. □

Recall the definition of $A(x, r)$ in Equation (9). We want to show that $A(x, r)$ and $A(y, r)$ are disjoint when $x \neq y$ assuming that r is smaller than some threshold. Intuitively, one would expect this to be true, as the normals $n(x)$ and $n(y)$ become more and more parallel the closer the points x and y are to each other. It turns out that a formal proof is not too difficult as demonstrated by

Lemma 3. *Suppose that (A1) holds. Then there exists a constant $c_1 > 0$ such that*

$$A(x, r) \cap A(y, r) = \emptyset$$

when $y \neq x$, $x, y \in \partial\mathcal{C}$ and $|r| < c_1$. Moreover, when $r < c_1$, $n(x)$ can be chosen in such a way that $A(x, r) \subset \mathcal{C}$ and $A(x, -r) \subset \mathcal{C}^C$.

Proof. Let us make the counterassumption that there exists sequences

$$(x_n, y_n)_{n=1}^{\infty} \subset \partial\mathcal{C} \times \partial\mathcal{C}$$

and $(r_{n,1}, r_{n,2})_{n=1}^{\infty} \to (0, 0)$ such that

$$x_n - y_n = r_{n,1} n(x_n) - r_{n,2} n(y_n)$$

and $x_n \neq y_n$. But then we would have

$$1 = \frac{r_{n,1} n(x_n)^T (x_n - y_n)}{\|x_n - y_n\|^2} - \frac{r_{n,2} n(y_n)^T (x_n - y_n)}{\|x_n - y_n\|^2}$$

leading to a contradiction, because by Lemma 2, the right side should go to zero.

By the previous part, we know that $A(x, r)$ must be either a subset of $\mathcal{C}$ or the complement of its closure $\bar{\mathcal{C}}^C = (\mathcal{C} \cup \partial\mathcal{C})^C$ because otherwise it would contain points from $\partial\mathcal{C}$. To see that this would be contradictory, one should observe that the first part of the proof holds for $A(x, r) \cup \{x\}$ as well.

Let us make the counterassumption that

$$A(x, r) \cup A(x, -r) \subset \bar{\mathcal{C}}^C$$

for some $x \in \partial\mathcal{C}$. Choose arbitrarily small $0 < \delta < 1$, define the pair of points $(y^{(1)}, y^{(2)})$ by

$$y^{(i)} = x + \frac{(-1)^i}{4 + |r|^{-1}} \delta n(x)$$

and choose $\epsilon > 0$ in such a way that $B(y^{(i)}, \epsilon) \subset \bar{\mathcal{C}}^C$. Moreover, there exists a sequence $(x_n)_{n=1}^{\infty} \subset \mathcal{C}$ approaching x (when $n \to \infty$) and n_δ such that the set

$$\{x_{n_\delta} - sn(x) : s \in [0, \delta]\} \cup \{x_{n_\delta} - sn(x) : s \in [-\delta, 0]\} \qquad (12)$$

contains two distinct points $(z_\delta^{(1)}, z_\delta^{(2)})$ on $\partial\mathcal{C}$ approaching to (x, x) when $\delta \to 0$. This follows from the fact that for n_δ large enough, the set (12) intersects both $B(y^{(1)}, \epsilon)$ and $B(y^{(2)}, \epsilon)$ with $x_{n_\delta} \in \mathcal{C}$ thus containing points from $\mathcal{C}$ and $\bar{\mathcal{C}}^C$. But this is in contradiction with Lemma 2 because by definition

$$\left|\frac{n(x) \cdot (z_\delta^{(1)} - z_\delta^{(2)})}{\|z_\delta^{(1)} - z_\delta^{(2)}\|^2}\right| = \frac{1}{\|z_\delta^{(1)} - z_\delta^{(2)}\|}$$

and

$$|\frac{(n(x)-n(z_\delta^{(1)}))\cdot(z_\delta^{(1)}-z_\delta^{(2)})}{\|z_\delta^{(1)}-z_\delta^{(2)}\|^2}|\leq\|n(x)-n(z_\delta^{(1)})\|\frac{1}{\|z_\delta^{(1)}-z_\delta^{(2)}\|}$$

the latter being asymptotically neglible because we may choose the normals in such a way that

$$\|n(x)-n(z_\delta^{(1)})\|\to 0$$

as $\delta\to 0$.

To finish, we must examine the opposite case

$$A(x,r)\cup A(x,-r)\subset\mathcal{C}.$$

By Equation (3) we can find $(x_n)_{n=1}^\infty\subset\bar{\mathcal{C}}^C$ approaching x. As in the previous step, it can be seen that again

$$\{x_n-sn(x):s\in[0,\delta]\}\cup\{x_n-sn(x):s\in[-\delta,0]\}$$

contains at least two distinct points from $\partial\mathcal{C}$ for arbitrarily small $\delta>0$ when n is large enough. Analogously to the previous case, this leads to a contradiction. □

From now on, we will always choose $n(x)$ as the outer normal of $\mathcal{C}$, that is, to point outwards from $\mathcal{C}$. Such a function is necessarily continuous and thus measurable by the second part of Lemma 3. Even though the concept of an outer normal is intuitively rather clear, we still needed Lemma 3 to verify the existence of such a normal.

The following lemma is the main result of this section. The idea is simply to linearize the boundary so that locally it can be viewed as a plane by neglecting higher order terms.

Lemma 4. *Choose any $x\in\partial\mathcal{C}$ and define the sets $\Xi_1=B(y,r_2)\cap\mathcal{C}$ and $\Xi_2=B(y,r_2)\cap\mathcal{U}_x$. Then if (A1) holds, there exists constants $c_2,c_3>0$ (depending only on $\mathcal{C}$ and not on x,r_1 and r_2) such that for $0<r_1,r_2<c_2$ and $y=x-r_1n(x)$, we have*

$$\lambda(\Xi_1\setminus\Xi_2)+\lambda(\Xi_2\setminus\Xi_1)\leq c_3(r_1^{m+1}+r_2^{m+1}).\tag{13}$$

Proof. Let us make the counterassumption that there exists a sequence

$$(x_n,y_n,r_{1,n},r_{2,n})_{n=1}^\infty$$

with $r_{1,n},r_{2,n}\to 0$ such that the left side of inequality (13) exceeds $c_3r_{1,n}^{m+1}+c_3r_{2,n}^{m+1}$ for any $c_3>0$ when n is big enough. By compactness we may assume that $x_n\to x$ for some $x\in\partial\mathcal{C}$ and by Lemma 1,

$$\|\phi^{-1}(x_n)-\phi^{-1}(x)\|\leq c_x\|x_n-x\|\tag{14}$$

for some constant $c_x>0$ and a local parametrization $\phi:U\to B(x,\delta)\cap\partial\mathcal{C}$. For each $n>0$, choose an arbitrary point $z_n\in B(x_n,r_{1,n}+r_{2,n})\cap\partial\mathcal{C}$.

When n is large enough, the fact that $B(x_n, r_{1,n} + r_{2,n}) \subset B(x, \delta)$, a Taylor expansion and (14) yield

$$z_n = x_n + J_{\phi^{-1}(x_n)}\phi(\phi^{-1}(z_n) - \phi^{-1}(x_n)) + O(\|z_n - x_n\|^2).$$

The first sum in the right side is a point on the plane T_{x_n}, thus

$$\rho(z_n, T_{x_n}) = O(\|z_n - x_n\|^2). \tag{15}$$

Set $d_n = \sup_{z \in \partial\mathcal{C} \cap B(x_n, r_{1,n}+r_{2,n})} \rho(z, T_{x_n})$ and define the sets (the sum of a vector and a set being defined in the standard way)

$$G_n = \cup_{-d_n \le r \le d_n} (T_{x_n} + rn(x_n)).$$

Then it is clear that

$$\partial\mathcal{C} \cap B(y_n, r_{2,n}) \subset G_n \cap B(x_n, r_{1,n} + r_{2,n}) \tag{16}$$

and by Equation (15)

$$\lambda(G_n \cap B(x_n, r_{1,n} + r_{2,n})) = O(r_{1,n}^{m+1} + r_{2,n}^{m+1}). \tag{17}$$

We may divide $B(y_n, r_{2,n}) \setminus G_n$ into the sets

$$A_1 = (B(y_n, r_{2,n}) \setminus G_n) \cap \mathcal{U}_{x_n}$$

and

$$A_2 = (B(y_n, r_{2,n}) \setminus G_n) \cap \mathcal{U}_{x_n}^C,$$

both of which are open and convex. Now for n big enough,

$$y_n - \frac{1}{2} r_{2,n} n(x_n) \in \mathcal{C} \cap A_1.$$

Thus $A_1 \cap \mathcal{C}$ is non-empty and consequently A_1 must be a subset of $\mathcal{C}$ because it does not contain points from $\partial\mathcal{C}$ (as stated in Equation (16)). On the other hand, by Lemma 3 and the same argument as previously, A_2 is in $\mathcal{C}^C$ when n is big enough and $A_2 \neq \emptyset$. Thus, inevitably

$$A_2 \cap \mathcal{C} = \emptyset.$$

We may conclude that

$$(B(y_n, r_{2,n}) \cap \mathcal{C}) \setminus G_n = (B(y_n, r_{2,n}) \cap \mathcal{U}_{x_n}) \setminus G_n$$

and Equation (17) leads to a contradiction finishing the proof. □

3.2 The Set $\partial_r\mathcal{C}$

$\partial_r\mathcal{C}$ was defined in Section 2.1 to consist of those points in $\mathcal{C}$ for which the distance from the boundary is at most r. Our goal is to reparametrize this set as

$$(x, r) \mapsto x - rn(x), \tag{18}$$

where $x \in \partial\mathcal{C}$. It turns out that such a parametrization is possible as shown by the following lemma.

Lemma 5. *There exists a constant* $c_4 > 0$ *depending only on* $\mathcal{C}$ *such that for* $0 < r < c_4$,

$$\partial_r \mathcal{C} = \cup_{x \in \partial \mathcal{C}} A(x, r).$$

Proof. Suppose that $(x_i)_{i=1}^{\infty}$ is a sequence with $d_i = \rho(x_i, \partial \mathcal{C}) \to 0$ and $x_i \in \mathcal{C}$. By compactness we may assume that $x_i \to x \in \partial \mathcal{C}$.

Choose a local parametrization $\phi : U \to \partial \mathcal{C} \cap B(x, \epsilon)$ ($\phi(0) = x$) at the point x and define the injective mapping $g : U \times [-\epsilon, \epsilon] \to \Re^m$ (similar considerations as in Lemma 3 can be applied here for small $\epsilon > 0$) by

$$g(y, r) = \phi(y) - rn(\phi(y)).$$

To see that g is continuously differentiable at the origin, let $(v_i(\phi(y)))_{i=1}^{m-1}$ be an orthonormal basis for the tangent space at the point $\phi(y)$ obtained by Gram-Schmidt orthonormalization of the columns of $J_y \phi$. Then each $v_i(\phi(y))$ is a continuously differentiable function on U. For y close to 0, we obtain

$$n(\phi(y)) = \frac{n(x) - \sum_{i=1}^{m-1} \langle n(x), v_i(\phi(y)) \rangle \, v_i(\phi(y))}{\| n(x) - \sum_{i=1}^{m-1} \langle n(x), v_i(\phi(y)) \rangle \, v_i(\phi(y)) \|}.$$

Clearly $n(\phi(y))$ is continuously differentiable with respect to y, because the denominator is bounded away from zero when y is close enough to 0. Moreover, the Jacobian of g at $(0, 0)$ is

$$J_{(0,0)} g = [J_0 \phi, -n(0)],$$

which is non-singular. Thus by the implicit function theorem, the set

$$g(B((0, 0), \delta))$$

is open for any small $\delta > 0$ and it contains x. Consequently there exists an integer i such that x_i belongs to the range of g.

Now assume that (z_i, t_i) is the pair with $g(z_i, t_i) = x_i$. For any $\epsilon > 0$, we may choose a point $y_{i,\epsilon} \in \partial \mathcal{C}$ such that $\|y_{i,\epsilon} - x_i\| \leq d_i + \epsilon$. Then by Lemma 2,

$$\begin{aligned} (d_i + \epsilon)^2 \geq \|x_i - y_{i,\epsilon}\|^2 &= \|\phi(z_i) - t_i n(\phi(z_i)) - y_{i,\epsilon}\|^2 \\ &= t_i^2 + \|\phi(z_i) - y_{i,\epsilon}\|^2 + O(t_i \|\phi(z_i) - y_{i,\epsilon}\|^2), \end{aligned}$$

because $\phi(z_i) - y_{i,\epsilon}$ is approximately parallel to the tangent plane at $\phi(z_i)$. The remainder term $O(t_i \|\phi(z_i) - y_{i,\epsilon}\|^2)$ is actually due to second order effects; again we applied a linearization argument. Now, because δ can be chosen as arbitrarily small (and thus t_i as well), the error term can be essentially neglected (we assumed it is at most half of $\|\phi(z_i) - y_{i,\epsilon}\|^2$ in absolute value). But such an argument implies that $t_i \leq d_i$.

To summarize, for any choice $(x_i)_{i=1}^{\infty}$, x_i belongs to the set $\cup_{x \in \partial \mathcal{C}} A(x, d_i)$ for large i. In other words, for small r

$$\partial_r \mathcal{C} \subset \cup_{x \in \partial \mathcal{C}} A(x, r).$$

The other direction is easier rather trivially:

$$\cup_{x \in \partial \mathcal{C}} A(x, r) \subset \partial_r \mathcal{C}.$$

□

The mapping of Equation (18) has a rather complex Jacobian due to the nonlinearity of $n(x)$. This poses some problems when doing a change of variable in integrals over $\partial_r \mathcal{C}$. However, it turns out that such difficulties only affect higher order terms and can be neglected in our nearest neighbor analysis.

Lemma 6. *Set* $\Xi = \cup_{x \in \partial\mathcal{C}} A(x, r)$. *Then for any function* $f : \mathcal{C} \to \Re$ *with* $|f| \leq 1$, *we have for* $r > 0$

$$\int_{\Xi} f(x)dx = \int_{\partial\mathcal{C}} \int_0^r f(x - \tilde{r}n(x))d\tilde{r}dS + O(r^2),$$

where the outer integral is the surface integral over $\partial\mathcal{C}$. *The remainder term* $O(r^2)$ *can be bounded by* $c_5 r^2$ *with the constant* c_5 *depending only on* $\mathcal{C}$.

Proof. Choose $x_0 \in \partial\mathcal{C}$ and a local parametrization $\phi : U \to B(x_0, \delta) \cap \partial\mathcal{C}$. Then $g(y, r) = \phi(y) - rn(\phi(y))$ is an injection. It has the Jacobian

$$J_{(y,r)}g = \left[J_y\phi - rJ_y n(\phi(y)), -n(\phi(y))\right].$$

Because all submatrices in the expression are bounded (ϕ can be chosen in such a way by restricting it into a subset of U if necessary, see the proof of Lemma 1), we may use (for the L^2-matrix norm)

$$\sup_{\|D\|,\|E\| \leq 1} \det(D + \epsilon E) - \det(D) = O(\epsilon),$$

when ϵ approaches zero to conclude that the determinant of $J_{(y,r)}g$ is

$$|J_{(y,r)}g| = |\left[J_y\phi, n(\phi(y))\right]| + O(r) = \sqrt{|(J_y\phi)^T J_y\phi|} + O(r).$$

Thus by a change of variables we obtain

$$\begin{aligned} & \int_{\cup_{x \in B(x_0,\delta)\cap\partial\mathcal{C}} A(x,r)} f(x)dx \\ = & \int_U \int_0^r f(\phi(y) - \tilde{r}n(y))\sqrt{|(J_y\phi)^T J_y\phi|}d\tilde{r}dy + O(r^2) \\ = & \int_{\partial\mathcal{C}\cap B(x_0,\delta)} \int_0^r f(x - \tilde{r}n(y))d\tilde{r}dS + O(r^2). \end{aligned}$$

To finish the proof, notice that by compactness, $\partial\mathcal{C}$ can be divided into a finite number of sets $\partial\mathcal{C} \cap B(x_i, \delta_i)$ with corresponding local parametrizations ϕ_i. One can examine each ball separately by replacing f by

$$f_i(x) = I(x \notin \cup_{j=1}^{i-1} \cup_{z \in B(x_j,\delta_j)\cap\partial\mathcal{C}} A(z, r))f(x)$$

to take into account the overlap between the sets. □

4 AUXILIARY RESULTS FOR NEAREST NEIGHBORS

4.1 Nearest Neighbor Distributions

Let us define the probability mass function

$$\omega_x(r) = P(X_1 \in B(x,r))$$

corresponding to the probability that a point belongs to the open ball $B(x,r)$. For any bounded function f, it is well-known [2] that the distribution of the k-nearest neighbor distance conditional on X_1 is given by

$$E[f(d_k)|X_1 = x] = k\binom{M-1}{k}\int_0^\infty f(r)\omega_x(r)^{k-1}(1-\omega_x(r))^{M-k-1}d\omega_x(r); \tag{19}$$

here $d\omega_x(r)$ corresponds to the Lebesgue-Stieltjes measure. The derivation of this relation proceeds (informally) by observing that

$$P(d_k \in [r, r+dr]|X_1 = x)$$

is equal to the probability that one vector belongs to $B(x, r+dr) \setminus B(x,r)$, $k-1$ vectors to $B(x,r)$ and $M-k-1$ to the complement $B(x, r+dr)^C$. Then (19) follows by combinatorics.

4.2 Uniformly Distributed Points in the Unit Cube

Here we will analyze points uniformly distributed in the unit cube $[0,1] \times [-1/2, 1/2]^{m-1}$. It turns out that the more general case is not much different to the one analyzed here. As a first step, we examine the situation, where the boundary effect can be neglected.

Consider a point x far from the boundaries of the cube. Then, because $p = 1$, $\omega_x(r)$ takes the simple form

$$\omega_x(r) = V_m r^m, \tag{20}$$

or equivalently

$$r = V_m^{-1/m}\omega_x(r)^{1/m}.$$

Using the relation between $\omega_x(r)$ and r, a good approximation would be (see Equation 19)

$$\begin{aligned} &E[d_k^\alpha|X_1 = x] \\ &\quad\approx V_m^{-\alpha/m}k\binom{M-1}{k}\int_0^\infty \omega_x(r)^{\alpha/m+k-1}(1-\omega_x(r))^{M-k-1}d\omega_x(r) \\ &\quad= V_m^{-\alpha/m}k\binom{M-1}{k}\int_0^1 z^{\alpha/m+k-1}(1-z)^{M-k-1}dz. \end{aligned} \tag{21}$$

Of course, this approximation is never exact as Equation (20) does not hold for large values of r due to the boundary effect. Even though Equation (21) looks complex, it can actually be written in terms of Gamma functions. The interesting part in such a representation is the approximation

$$\frac{\Gamma(M)}{\Gamma(M+\delta)} = M^{-\delta} + O(M^{-1-\delta}),$$

which becomes very accurate even for relatively small values of M and thus one can write the right side of Equation (21) in the following intuitive form.

Lemma 7. *For any $\alpha > 0$, the integral in Equation (21) can be represented using the Gamma function as*

$$k\binom{M-1}{k}\int_0^1 z^{\alpha/m+k-1}(1-z)^{M-k-1}dz = \frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)}$$
$$= \frac{\Gamma(k+\alpha/m)}{\Gamma(k)}M^{-\alpha/m} + O(k^{\alpha/m}M^{-\alpha/m-1}).$$

Proof. Denoting by β the Beta function,

$$\int_0^1 z^{\alpha/m+k-1}(1-z)^{M-k-1}dz = \beta(\alpha/m+k, M-k); \tag{22}$$

on the other hand,

$$\beta(\alpha/m+k, M-k) = \frac{\Gamma(k+\alpha/m)\Gamma(M-k)}{\Gamma(M+\alpha/m)} \tag{23}$$

and

$$k\binom{M-1}{k} = \frac{\Gamma(M)}{\Gamma(k)\Gamma(M-k)}. \tag{24}$$

The proof is finished by combining Equations (22), (23) and (24). □

Problems arise once x is close to the boundaries, because the approximation in (21) is not valid as the relation between $\omega_x(r)$ and r becomes much more complicated than (20). One idea would be to assume that the number of such points in a sample is asymptotically neglible and thus neglect the boundary effect in the analysis, see [2]. Here we show the rather surprising result that actually the effect of the boundaries can be estimated in an analytic way.

When points close to the corners of the cube are neglected, we can restrict ourselves to vectors of the form

$$x = (s, 0, \ldots, 0). \tag{25}$$

Recalling the function h defined in Section 2.1, for $r < 1/2$ we have (the notation $(s, 0)$ is a shorthand for (25))

$$\omega_{(s,0)}(r) = s^m h(\frac{r}{s}) \tag{26}$$

and

$$d\omega_{(s,0)}(r) = s^{m-1}h'(\frac{r}{s})dr. \tag{27}$$

Given x close to the boundary, $\omega_x(r)$ has a rather complicated form and at first sight it seems that evaluating $E[d_k^\alpha|X_1 = x]$ is rather difficult. This is indeed the case if X_1 is held fixed. Here we propose another approach: we let the first coordinate $x^{(1)}$ vary and thus instead of a single integral, we obtain a double integral, which can be represented in a simple form.

Before stating the main result of this section, we define the function g by

$$g_\gamma(z) = 1 \quad \text{if } z < \gamma$$

and 0 otherwise. We will replace d_k by $d_k g_\gamma(d_k)$; in the next section we will analyze the additional error caused by adding g_γ.

Theorem 2. *Let us define the constant*

$$D = \frac{1}{m}\int_0^1 a^{-\alpha-2} h(a^{-1})^{-\alpha/m-1/m-1} h'(a^{-1}) da.$$

If the variables $(X_i)_{i=1}^M$ are uniform on the cube $[0,1]\times[-1/2,1/2]^{m-1}$, we have for any $\alpha > 0$ and $\gamma < (2+V_m^{1/m})^{-1}$,

$$\int_0^\gamma E[g_\gamma(d_k)d_k^\alpha | X_1 = (s,0,\dots,0)]ds = V_m^{-\alpha/m}\frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)}\gamma$$
$$+ (D - V_m^{-\alpha/m-1/m})\frac{\Gamma(k+\alpha/m+1/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m+1/m)} + R_1.$$

The remainder term is bounded by

$$|R_1| \le M^k e^{-V_m\gamma^m(M-k-1)}(D + V_m^{-\alpha/m} + V_m^{-\alpha/m-1/m}).$$

Proof. Recalling the nearest neighbor distribution in Equation (19) and applying Equation (27) we obtain

$$k\binom{M-1}{k}\int_0^\gamma\int_0^\gamma r^\alpha \omega_{(s,0)}(r)^{k-1}(1-\omega_{(s,0)}(r))^{M-k-1} d\omega_{(s,0)}(r)ds$$
$$= k\binom{M-1}{k}\int_0^\gamma\int_0^\gamma t(s,r)drds \tag{28}$$

with the definition

$$t(s,r) = r^\alpha s^{m-1}\omega_{(s,0)}(r)^{k-1}(1-\omega_{(s,0)}(r))^{M-k-1}h'(\frac{r}{s}).$$

The integral (28) can be divided into two parts by considering sets with $s > r$ and $s < r$ separately. We use Equation (20) and the identity (valid for $s > r$)

$$h'(\frac{r}{s}) = \frac{mV_m r^{m-1}}{s^{m-1}}$$

to write the contribution from the first set as

$$\begin{aligned} I_1 &= k\binom{M-1}{k}\int_0^\gamma\int_r^\gamma t(s,r)dsdr \\ &= mkV_m^k\binom{M-1}{k}\int_0^\gamma\int_r^\gamma r^{\alpha+km-1}(1-V_m r^m)^{M-k-1}dsdr \\ &= mkV_m^k\binom{M-1}{k}\int_0^\gamma r^{\alpha+km-1}(\gamma-r)(1-V_m r^m)^{M-k-1}dr. \end{aligned}$$

By making the change of variable $y = V_m r^m$, I_1 can be written as

$$I_1 = V_m^{-\alpha/m}\gamma k\binom{M-1}{k}\int_0^{V_m\gamma^m} y^{\alpha/m+k-1}(1-y)^{M-k-1}dy$$
$$- V_m^{-\alpha/m-1/m}k\binom{M-1}{k}\int_0^{V_m\gamma^m} y^{\alpha/m+1/m+k-1}(1-y)^{M-k-1}dy.$$

The integrals from 0 to $V_m\gamma^m$ can be extended to integrals from 0 to 1 at the expense of an error term roughly bounded by

$$|R_1^{(a)}| \leq M^k(V_m^{-\alpha/m} + V_m^{-\alpha/m-1/m})\int_{V_m\gamma^m}^1 y^{\alpha/m+k-1}(1-y)^{M-k-1}dy$$
$$\leq M^k(V_m^{-\alpha/m} + V_m^{-\alpha/m-1/m})e^{-V_m\gamma^m(M-k-1)}.$$

Applying Lemma 7, we obtain the final form of I_1:

$$I_1 = V_m^{-\alpha/m}\gamma\frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)} - V_m^{-\alpha/m-1/m}\frac{\Gamma(k+\alpha/m+1/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m+1/m)} + R_1^{(a)}.$$

Next we proceed to the slightly more difficult case $r > s$. One possible approach is to make the change of variable $(s,r) = (ar,r)$ to obtain

$$I_2 = k\binom{M-1}{k}\int_0^\gamma\int_0^r t(s,r)dsdr = k\binom{M-1}{k}\int_0^1 r\int_0^\gamma t(ar,r)drda$$
$$= k\binom{M-1}{k}\int_0^1 a^{km-1}h(a^{-1})^{k-1}h'(a^{-1})$$
$$\times\int_0^\gamma r^{\alpha+mk}(1-a^m h(a^{-1})r^m)^{M-k-1}drda$$
$$= k\binom{M-1}{k}\int_0^1 a^{-\alpha-2}h(a^{-1})^{-\alpha/m-1-1/m}h'(a^{-1})$$
$$\times\int_0^{a^m h(a^{-1})\gamma^m} y^{\alpha/m+k+1/m-1}(1-y)^{M-k-1}dyda.$$

By Equation (4),

$$a^m h(a^{-1})\gamma^m > \frac{1}{2}V_m\gamma^m$$

and consequently

$$I_2 = Dk\binom{M-1}{k}\int_0^1 y^{\alpha/m+k+1/m-1}(1-y)^{M-k-1}dy + R_1^{(b)}$$
$$= D\frac{\Gamma(k+\alpha/m+1/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m+1/m)} + R_1^{(b)}$$

the error term being bounded by

$$|R_1^{(b)}| \leq DM^k\int_{\frac{1}{2}V_m\gamma^m}^1 y^{\alpha/m+k+1/m-1}(1-y)^{M-k-1}dy$$
$$\leq DM^k e^{-\frac{1}{2}V_m\gamma^m(M-k-1)}.$$

□

4.3 General Distributions

Analyzing points distributed according to more general measures than the uniform distribution on a cube is rather challenging. However, when the boundary effect can be neglected, the problem simplifies considerably as in such a situation the techniques for uniform random variables can be applied in a straightforward manner.

Lemma 8. *Assume that (A2) holds. Then there exists a constant c_6 independent of M, k, x and γ, such that for any $\gamma > 0$ and $x \in \mathcal{C} \setminus \partial_\gamma \mathcal{C}$,*

$$E[d_k^\alpha g_\gamma(d_k)|X_1 = x] = V_m^{-\alpha/m} p(x)^{-\alpha/m} \frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)} + R_2,$$

with

$$|R_2| \leq c_6 \gamma^{\xi+\alpha} + M^k e^{-c_p^{-1} V_m \gamma^m (M-k-1)}.$$

Proof. Let $\nabla_x p$ be the gradient of p which exists by (A1). Then by symmetry we know that

$$\int_{B(x,r)} \langle y - x, \nabla_x p \rangle \, dy = 0.$$

Thus we may find a constant $c > 1$ such that

$$|\omega_x(r) - V_m r^m p(x)| \leq c r^{m+\xi}.$$

Because p is bounded from below and the previous inequality holds for any choice $x \in \mathcal{C} \setminus \partial_\gamma \mathcal{C}$, we have on the event $X_1 \in \mathcal{C} \setminus \partial_\gamma \mathcal{C}$ (I refers to the indicator function)

$$g_\gamma(d_k) I(X_1 \in \mathcal{C} \setminus \partial_\gamma \mathcal{C}) |\frac{V_m^{\alpha/m} p(X_1)^{\alpha/m} d_k^\alpha}{\omega_{X_1}(d_k)^{\alpha/m}} - 1| \leq c\gamma^\xi.$$

Consequently, we only need to examine $W = E[\omega_{X_1}(d_k)^{\alpha/m} g_\gamma(d_k)|X_1 = x]$ which can be evaluated using Equation (19):

$$\begin{aligned} W &= k\binom{M-1}{k} \int_0^\gamma \omega_x(r)^{\alpha/m+k-1}(1-\omega_x(r))^{M-k-1} d\omega_x(r) \\ &= k\binom{M-1}{k} \int_0^{\omega_x(\gamma)} z^{\alpha/m+k-1}(1-z)^{M-k-1} dz. \end{aligned}$$

Because $\omega_x(\gamma) \geq c_p^{-1} V_m \gamma^m$,

$$k\binom{M-1}{k} \int_{\omega_x(\gamma)}^1 z^{\alpha/m+k-1}(1-z)^{M-k-1} dz \leq M^k e^{-c_p^{-1} V_m \gamma^m (M-k-1)}$$

and we may finish the proof by the approximation

$$W \approx k\binom{M-1}{k} \int_0^1 z^{\alpha/m+k-1}(1-z)^{M-k-1} dz$$

and Lemma 7. □

For points close to $\partial\mathcal{C}$ the shape of the boundary plays an important role in a higher order expansion. It turns out that under sufficient regularity, a local linearization can be used for a reduction to a simpler case with the expense of some higher order terms. In contrary to Lemma 8, both $\partial\mathcal{C}$ and p are linearized.

Let $(X_{i,unif})_{i=1}^M$ be a set of uniformly distributed random variables as in Section 4.2. The corresponding nearest neighbor distances are denoted by $d_{k,unif}$. The proof of the following theorem is based on a coupling argument for the total variation distance [1, 6], which has turned out to be rather powerful.

Lemma 9. *Assume that (A1)-(A2) hold. Then there exists a constant $c_7 > 0$ depending only on $\mathcal{C}$, α, and p but not on M, k, y or γ, such that for all $x \in \partial\mathcal{C}$, $0 < r < \gamma < 1$ and $y = x - rn(x)$,*

$$|p(x)^{1-\alpha/m} E[d_{k,unif}^\alpha g_\gamma(d_{k,unif}) | X_1^{unif} = (p(x)^{1/m} r, 0, \ldots, 0)] \\ - E[d_k^\alpha g_\gamma(d_k) | X_1 = y] p(y)| \leq c_7(\gamma^{\alpha+1} + M\gamma^{m+\alpha+1}).$$

Proof. By rotation and translation, we may assume without losing generality that

$$x = (0, 0, \ldots, 0)$$

and similarly $n(x) = (-1, 0, \ldots, 0)$. We will use the notation Ξ_1 for the set $B(y,\gamma) \cap \mathcal{U}_x$ and Ξ_2 for $B(y,\gamma) \cap \mathcal{C}$. Define a new density $\tilde{p}$ by setting

$$\tilde{p}(z) = p(x) I(z \in \Xi_1)$$

for $z \in B(y,\gamma)$ and

$$\tilde{p}(z) = \frac{(1 - p(x)\lambda(\Xi_1))p(z)}{1 - P(X_1 \in \Xi_1)}$$

otherwise. Lemma 4 and Assumption (A2) ensure the existence of a constant $\tilde{c} > 0$ such that

$$\begin{aligned} \int_{B(y,\gamma)} |\tilde{p}(z) - p(z)| dz &\leq L\gamma^{m+1} + p(x)\lambda(\Xi_1 \setminus \Xi_2) + p(x)\lambda(\Xi_2 \setminus \Xi_1) \\ &\leq \tilde{c}\gamma^{m+1}. \end{aligned}$$

Moreover, for γ small enough to ensure $P(X_1 \in \Xi_1) \leq 1/2$, we have

$$|1 - \frac{1 - p(x)\lambda(\Xi_1)}{1 - P(X_1 \in \Xi_1)}| \leq 2\tilde{c}\gamma^{m+1}.$$

This gives us

$$\int_{\Re^m} |\tilde{p}(z) - p(z)| dz \leq c\gamma^{m+1},$$

which is a bound on the total variation distance between the two probability measures. By a classical coupling argument (see [1] or [6]), there exists an i.i.d. sample $(\tilde{X}_i^{(1)})_{i=1}^M$ distributed according to the density $\tilde{p}$ (we may of course extend the probability space $(\Omega, \mathcal{F}, P)$ in an appropriate way) such that for each $i > 1$,

$$P(X_i \neq \tilde{X}_i^{(1)}) \leq 2c\gamma^{m+1}$$

and consequently

$$P((X_i)_{i=2}^M \neq (\tilde{X}_i^{(1)})_{i=2}^M) \leq \sum_{i=2}^M P(X_i \neq \tilde{X}_i^{(1)}) \leq 2cM\gamma^{m+1}. \qquad (29)$$

The new sample has a convenient uniformity property in the neighborhood of y. Taking $X_1 = \tilde{X}_1^{(1)}$ independent of $(X_i, \tilde{X}_i^{(1)})_{i=2}^M$, we obtain by Equation (29)

$$|E[d_k^\alpha g_\gamma(d_k)|X_1 = y] - E[\tilde{d}_{k,1}^\alpha g_\gamma(\tilde{d}_{k,1})|\tilde{X}_1^{(1)} = y]| \leq 2cM\gamma^{m+\alpha+1}, \qquad (30)$$

because on the event $(X_i)_{i=2}^M = (\tilde{X}_i)_{i=2}^M$ the nearest neighbor distances are the same for both samples.

Fixing the notation $\Delta = [0, p(x)^{-1/m}] \times [-p(x)^{-1/m}/2, p(x)^{-1/m}/2]^{m-1}$, we introduce a third sample, $(\tilde{X}_i^{(2)})_{i=1}^M$:

When $\tilde{X}_i^{(1)} \in B(y,\gamma)$, set $\tilde{X}_i^{(1)} = \tilde{X}_i^{(2)}$.
Otherwise choose $\tilde{X}_i^{(2)}$ uniformly from the set $\Delta \setminus B(y,\gamma)$.

Conditional on $\tilde{X}_1^{(1)} = y$, the variable $\tilde{d}_{k,1}^\alpha g_\gamma(\tilde{d}_{k,1})$ depends only on points in the ball $B(y,\gamma)$. Thus the nearest neighbors in the second and third samples are the same:

$$E[\tilde{d}_{k,1}^\alpha g_\gamma(\tilde{d}_{k,1})|\tilde{X}_1^{(1)} = y] = E[\tilde{d}_{k,2}^\alpha g_\gamma(\tilde{d}_{k,2})|\tilde{X}_1^{(2)} = (r,0)]. \qquad (31)$$

The proof is finished by combining Equations (30) and (31) because $(\tilde{X}_i^{(2)})_{i=1}^M$ is uniformly distributed and (A2) implies the bound

$$E[d_k^\alpha g_\gamma(d_k)|X_1 = y]|p(y) - p(x)| \leq L\gamma^{1+\alpha}.$$

□

This far we have used the threshold function g_γ to bound the largest nearest neighbor value. The probability of the event $g_\gamma(d_k) = 1$ can be bounded in a rather straightforward way.

Lemma 10. *Suppose that (A1) and (A2) hold. Then there exists a constant $c_8 > 0$ independent of M and k such that*

$$P(d_k > \gamma|X_1) \leq M^k e^{-c_8\gamma^m(M-k-1)}.$$

Proof. As a consequence of Lemma 4 we obtain

$$\lambda(B(x,r) \cap \mathcal{C}) \geq cr^m$$

for some $c > 0$ and all $x \in \mathcal{C}$. Using the nearest neighbor distribution (21) we obtain the result

$$\begin{aligned} P(d_{1,k} > \gamma|X_1 = x) &= \binom{M-1}{k} \int_\gamma^\infty \omega_x(r)^{k-1}(1-\omega_x(r))^{M-k-1} d\omega_x(r) \\ &\leq M^k e^{-(M-k-1)\omega_x(\gamma)} \leq M^k e^{-c\gamma^m(M-k-1)}, \end{aligned}$$

which finishes the proof. □

5 PROOF OF THEOREM 1

5.1 Smooth Sets

(Proof of Theorem 1 for sets satisfying (A1) and (A2)). By Lemma 10, there exists a constants $c > 0$ such that

$$|R^{(a)}| = |E[d_k^\alpha] - E[d_k^\alpha g_\gamma(d_k)]| \leq M^k e^{-c\gamma^m M}.$$

The expectation of $d_k^\alpha g_\gamma(d_k)$ can be divided into an integral over the set of points close to the boundary and its complement:

$$\begin{aligned} E[d_k^\alpha g_\gamma(d_k)] &= \int_{\partial_\gamma \mathcal{C}} E[d_k^\alpha g_\gamma(d_k)|X_1 = x]p(x)dx \\ &\quad + \int_{\mathcal{C}\setminus\partial_\gamma \mathcal{C}} E[d_k^\alpha g_\gamma(d_k)|X_1 = x]p(x)dx \\ &= I_1 + I_2. \end{aligned}$$

As a first step to solve I_1, we reparametrize it using Lemma 6 to obtain

$$I_1 = \int_{\partial \mathcal{C}} \int_0^\gamma E[d_k^\alpha g_\gamma(d_k)|X_1 = x - rn(x)]p(x - rn(x))drdS + R^{(b)}$$

with $R^{(b)} \leq c_5 \gamma^{2+\alpha}$. An application of Lemma 9 and a change of variables leads to

$$\begin{aligned} I_1 &= \int_{\partial \mathcal{C}} p(x)^{1-\alpha/m-1/m} \int_0^{p(x)^{1/m}\gamma} E[d_{k,unif}^\alpha g_\gamma(d_{k,unif})|X_1^{unif} = (r,0)]drdS \\ &\quad +R^{(c)} \end{aligned}$$

with

$$|R^{(c)}| \leq c_5 \gamma^{2+\alpha} + c_7(\gamma^{\alpha+2} + M\gamma^{m+\alpha+2}).$$

To apply Theorem 2 for solving the inner integral, one more observation is required:

$$\begin{aligned} &|E[d_{k,unif}^\alpha g_\gamma(d_{k,unif})|X_1^{unif} = (r,0)] - E[d_{k,unif}^\alpha g_{p(x)^{1/m}\gamma}(d_{k,unif})|X_1^{unif} = (r,0)]| \\ &\leq (c_p^{\alpha/m} + 1)\gamma^\alpha P(d_{k,unif} > (1 + c_p^{1/m})^{-1}\gamma|X_1^{unif} = (r,0)) \\ &\leq M^k (c_p^{\alpha/m} + 1)\gamma^\alpha e^{-\frac{1}{2}V_m(1+c_p^{1/m})^{-m}\gamma^m(M-k-1)}. \end{aligned}$$

Now Theorem 2 can be applied to obtain

$$\begin{aligned} I_1 &= V_m^{-\alpha/m} \frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)} \gamma \int_{\partial \mathcal{C}} p(x)^{1-\alpha/m} dS \\ &\quad +(D - V_m^{-\alpha/m-1/m}) \frac{\Gamma(k+\alpha/m+1/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m+1/m)} \int_{\partial \mathcal{C}} p(x)^{1-\alpha/m-1/m} dS \\ &\quad +R^{(d)} \end{aligned}$$

with

$$\begin{aligned} |R^{(d)}| &\leq |R_3^{(c)}| + M^k e^{-V_m c_p^{-1} \gamma^m (M-k-1)} (D + V_m^{-\alpha/m} + V_m^{-\alpha/m-1/m}) \\ &\quad + M^k (c_p^{\alpha/m} + 1)\gamma^\alpha e^{-\frac{1}{2}V_m(1+c_p^{1/m})^{-m}\gamma^m(M-k-1)}. \end{aligned}$$

By assumption (A2) and Lemma 6, we find a constant $c > 0$ such that

$$
\begin{aligned}
&|\gamma \int_{\partial\mathcal{C}} p(x)^{1-\alpha/m} dS - \int_{\partial_\gamma \mathcal{C}} p(x)^{1-\alpha/m} dx| \\
&\leq |\gamma \int_{\partial\mathcal{C}} p(x)^{1-\alpha/m} dS - \int_{\partial\mathcal{C}} \int_0^\gamma p(x - rn(x))^{1-\alpha/m} dr dS| \\
&+ |\int_{\partial\mathcal{C}} \int_0^\gamma p(x - rn(x))^{1-\alpha/m} dr dS - \int_{\partial_\gamma \mathcal{C}} p(x)^{1-\alpha/m} dx| \leq c\gamma^2 \qquad (32)
\end{aligned}
$$

and thus

$$
\begin{aligned}
I_1 &= V_m^{-\alpha/m} \frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)} \int_{\partial_\gamma \mathcal{C}} p(x)^{1-\alpha/m} dx \\
& +(D - V_m^{-\alpha/m-1/m}) \frac{\Gamma(k+\alpha/m+1/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m+1/m)} \int_{\partial\mathcal{C}} p(x)^{1-\alpha/m-1/m} dS \\
& +R_3^{(e)}.
\end{aligned}
$$

Lemma 8 allows a straightforward estimation of I_2 by

$$
I_2 = V_m^{-\alpha/m} \frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)} \int_{\mathcal{C}\backslash\partial_\gamma \mathcal{C}} p(x)^{1-\alpha/m} dx + R^{(f)}
$$

with

$$
|R^{(f)}| \leq c_6 \gamma^{\xi+\alpha} + M^k e^{-c_p^{-1} V_m \gamma^m (M-k-1)}.
$$

The proof is finished by setting

$$
R = R^{(a)} + R^{(e)} + R^{(f)}.
$$

□

5.2 Polytopes

Proof. (Proof of Theorem 1 for polytopes) The proof for polytopes is simpler than that for more general sets as the boundary is piecewise linear. As there is no need for a detailed derivation, a sketch is given here.

Again, we need to examine $d_k^\alpha g_\gamma(d_k)$ as $\lambda(B(x,r)) \geq cr^m$ for a constant $c > 0$ and all $x \in \mathcal{C}$ validating Lemma 10. We only need to examine the term I_1 defined in the first part of the previous proof as I_2 takes a similar form as before. Moreover, it can be shown that if $x \in A(y, \gamma)$ for some $y \in \partial\mathcal{C}$, then

$$
B(x,\gamma) \cap \mathcal{C} = B((\|x - y\|, 0, \ldots, 0), \gamma) \cap \mathcal{H} \qquad (33)
$$

assuming that x does not belong to a set of volume $O(\gamma^2)$. The points where Equation (33) does not hold are contained close to the $m-2$ dimensional planes where two faces intersect. Thus Lemma 9 is valid except in a neglible set and also

$$
\begin{aligned}
\int_{\partial_\gamma \mathcal{C}} f(x) dx &= \int_{\cup_{y \in \partial\mathcal{C}} A(y,\gamma)} f(x) dx \\
&= \int_{x \in \partial\mathcal{C}} \int_0^\gamma f(x - rn(x)) dr dx + O(\gamma^2)
\end{aligned}
$$

for any bounded measurable function f including $E[g(d_k)d_k^\alpha|X_1 = x]$. Notice that here the sets $A(x,r)$ and $A(y,r)$ cannot be assumed to be disjoint or contained in $\mathcal{C}$; however, the overlap is neglible as again problems arise only close to intersections of the faces. We obtain

$$\begin{aligned}
I_1 &= V_m^{-\alpha/m}\frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)}\gamma\int_{\partial\mathcal{C}} p(x)^{1-\alpha/m}dS \\
&\quad +(D - V_m^{-\alpha/m-1/m})\frac{\Gamma(k+\alpha/m+1/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m+1/m)}\int_{\partial\mathcal{C}} p(x)^{1-\alpha/m-1/m}dS \\
&\quad +R^{(a)} \\
&= V_m^{-\alpha/m}\frac{\Gamma(k+\alpha/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m)}\int_{\partial_\gamma\mathcal{C}} p(x)^{1-\alpha/m}dS \\
&\quad +(D - V_m^{-\alpha/m-1/m})\frac{\Gamma(k+\alpha/m+1/m)\Gamma(M)}{\Gamma(k)\Gamma(M+\alpha/m+1/m)}\int_{\partial\mathcal{C}} p(x)^{1-\alpha/m-1/m}dS \\
&\quad +R^{(b)}
\end{aligned}$$

for a remainder term $R^{(b)}$ of the same order as in the first part. □

References

[1] R. L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theoretical of Probability and Its Applications*, 15:458–486, 1970.

[2] D. Evans, A. J. Jones, and W. M. Schmidt. Asymptotic moments of near neighbour distance distributions. *Proceedings of the Royal Society of London A*, 458:2839–2849, 2002.

[3] D. Evans, A. J. Jones, and W. M. Schmidt. A proof of the Gamma test. *Proceedings of the Royal Society of London A*, 458:2759–2799, 2002.

[4] A. O. Hero, B. Ma, O. Michel, and J. Gorman. Applications of entropic spanning graphs. *IEEE Signal Processing Magazine (Special Issue on Mathematics in Imaging)*, 19(5):85–95, 2002.

[5] L. F. Kozachenko and N. N. Leonenko. On statistical estimation of entropy of random vector. *Problems of Information Transmission (in Russian)*, 23(9), 1987.

[6] P. Massart. *Concentration Inequalities and Model Selection: Ecole d'Eté de Probabilités de Saint-Flour XXXIII - 2003*, volume 1896 of *Lecture Notes in Mathematics / Ecole d'Eté Probabilités de Saint-Flour*. Springer, 2007.

[7] M. D. Penrose. Normal approximation in geometric probability. *Electronic Journal of Probability*, 12:989–1035, 2007.

[8] M. D. Penrose and J. E. Yukich. Weak laws of large numbers in geometric probability. *Annals of Applied Probability*, 13(1):277–204, 2003.

[9] A. G. Percus and O. C. Martin. Scaling universalities of kth nearest neighbor distances on closed manifolds. *Advances in Applied Mathematics*, 21(3):424–436, 1998.

[10] A. Tewari and A. M. Gokhale. Nearest-neighbor distributions on thin films, sheets and plates. *Acta Materialia*, 54(7):1957–1963, 2006.

[11] S. Torquato. Mean nearest-neighbor distance in random packings of hard d-dimensional spheres. *Physical Review Letters*, 74(12):2156–2159, 1995.

[12] S. Torquato. Nearest-neighbor statistics for packings of hard spheres and disks. *Physical Review E*, 51(4):3170–3182, 1995.

[13] O. Vasicek. A test for normality based on sample entropy. *Journal of the Royal Statistical Society Series B*, 38(1):54–59, 1976.

[14] J. Yukich and C. Redmond. Limit theorems and rates of convergence for subadditive euclidean functionals. *Annals of Applied Probability*, 4(4):1057–1073, 1994.

TKK REPORTS IN INFORMATION AND COMPUTER SCIENCE

ISBN 978-951-22-9610-1 (Print)
ISBN 978-951-22-9611-8 (Online)
ISSN 1797-5034 (Print)
ISSN 1797-5042 (Online)