stats

MDPI

Article

Distribution of Distances between Elements in a Compact Set

Solal Lellouche [1] and Marc Souris [2,*]

[1] UFR Physique, Université Paris-Diderot, 75013 Paris, France; lellouchesolal@gmail.com
[2] UMR Unité des Virus Emergents (UVE Aix-Marseille Univ-IRD 190-Inserm 1207-IHU Méditerranée Infection), 13005 Marseille, France
* Correspondence: marc.souris@ird.fr

Received: 11 September 2019; Accepted: 21 December 2019; Published: 26 December 2019

Abstract: In this article, we propose a review of studies evaluating the distribution of distances between elements of a random set independently and uniformly distributed over a region of space in a normed $\mathbb{R}$-vector space (for example, point events generated by a homogeneous Poisson process in a compact set). The distribution of distances between individuals is present in many situations when interaction depends on distance and concerns many disciplines, such as statistical physics, biology, ecology, geography, networking, etc. After reviewing the solutions proposed in the literature, we present a modern, general and unified resolution method using convolution of random vectors. We apply this method to typical compact sets: segments, rectangles, disks, spheres and hyperspheres. We show, for example, that in a hypersphere the distribution of distances has a typical shape and is polynomial for odd dimensions. We also present various applications of these results and we show, for example, that variance of distances in a hypersphere tends to zero when space dimension increases.

Keywords: distance distribution; convolution; random vector; spatial analysis; geostatistics; autocorrelation; hypersphere

1. Introduction

The distribution of distances between elements in a set of points is present in many problems, particularly in spatial analysis, and in various fields of application: ecology, epidemiology, forestry, biology, astronomy, economics, particle physics, network applications, etc. [1]. For example, given two points randomly selected in a set of points independently and uniformly distributed in space, we aim to know the probability of the distance between these two points inside the set of distances between all the pairs of points (Figure 1).

This question is important when trying to evaluate or model spatial interactions between elements, such as clustering of objects, spatial autocorrelation of a variable across a set of locations, or neighbor relationships and connectivity [2]. Indeed, in nature many problems involve distance-based interactions between events or elements. For example, most methods used to measure spatial autocorrelation or to model spatial interactions are based on a weighted average of a variable between pairs of elements in a disk [2–5]. If such an index measures a phenomenon related to the distance between the elements, the index may favor the pairs of elements of the most likely distances. In order to avoid such bias, it is necessary to know the distribution of distances and consider the relationship between the distance and the phenomenon independently of the distribution of distances between all pairs of elements [2].

The distribution of distances between two randomly selected points in a compact has been studied for a long time. However, the results are fragmented because they are presented in different articles, with different methods of resolution, depending on the dimension of the space and the type of compact studied. In this article, we first present a literature review of these results. We then propose a unified

method of resolution which uses only standard mathematical objects, and which is generalizable to any type of compact set in any dimension. We will describe this general approach and use it to calculate distributions of Euclidian distances between two randomly chosen points, for compact sets as lines, rectangles, disks, cubes, and specific results for hyperspheres of any dimension.

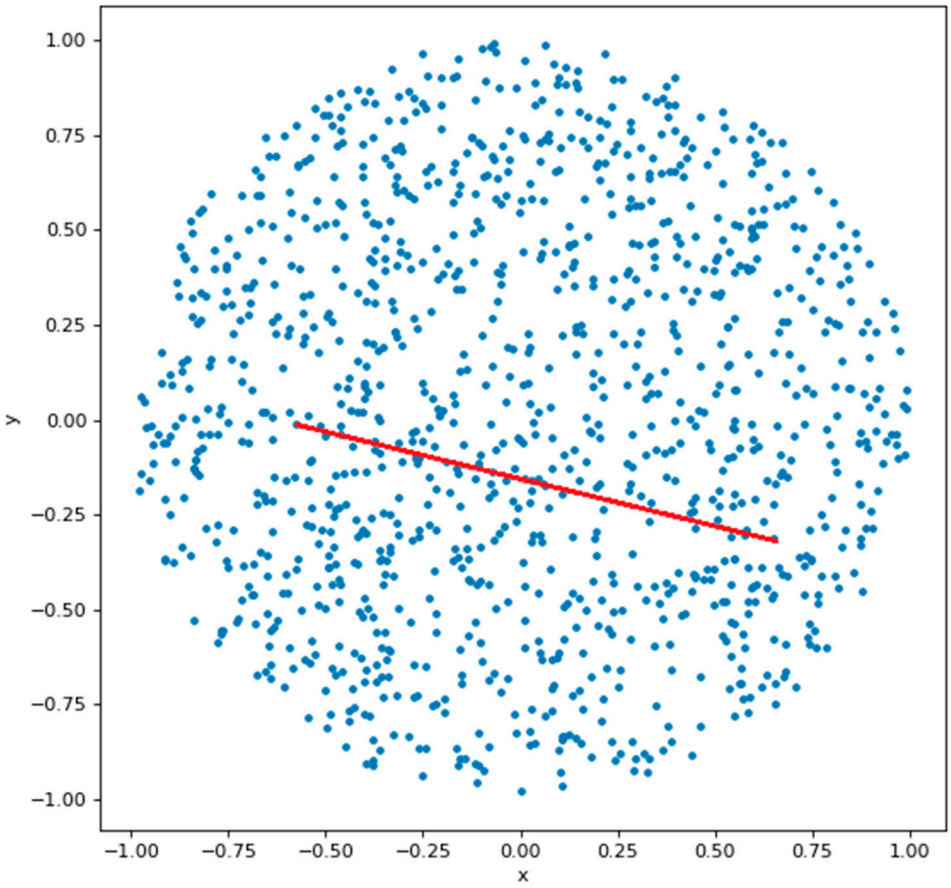


Figure 1. Distance between two randomly chosen points in a set of points independently and uniformly distributed in a disk (as generated by a homogeneous Poisson process). Among all possible distances between points, how likely would this distance be? Would it be below or above average?

2. Literature Review

The task of distance distribution estimation between points is related to stochastic geometry [6].

2.1. Rayleigh Distribution

Famous Nobel prize physicist Lord Rayleigh (1842–1919) solved a slightly simpler problem than the one studied in this article: he modelled the distribution (known as Rayleigh distribution) of Euclidean distances between a central point and a set of points normally distributed around this central point in a real vector space of dimension two [7].

By positioning the central point to the origin, the problem addressed by Rayleigh was to evaluate the distribution of $\| \vec{v} \|$, i.e., $\sqrt{{x_1}^2 + {x_2}^2}$ in cartesian coordinates for Euclidian distance, the vector $\vec{v}$ corresponding to the realization of two real random variables x_1 and x_2 independent but generated by the same density function:

$$\forall t \in \mathbb{R},\ f_{X_i}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right)\ i = 1, 2$$

where σ, the standard deviation of this normal distribution, allows one to set the concentration of points around the center.

The distribution of distances to the center can then be easily obtained by using the independence of the two random variables and by using polar coordinates:

$$\begin{aligned}\forall r \in \mathbb{R}^+, \quad & p(\| \vec{V} \| \leq r) \\ & = p(\vec{V} \in B(\vec{0}, r)) \textit{where } B(\vec{0}, r) \textit{ is the closed ball of raduis } r \\ & = \int_{B(\vec{0},r)} f_X(x,y) d(x,y) \\ & = \int_{B(\vec{0},r)} f_{X_1}(x) f_{X_2}(y) dxdy \textit{ from coordinates independence} \\ & = \frac{1}{2\pi\sigma^2} \int_{B(\vec{0},r)} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) dxdy \textit{ from normal distribution} \\ & = \frac{1}{2\pi\sigma^2} \int_0^{2\pi} \int_0^r \mathrm{t} \exp\left(-\frac{t^2}{2\sigma^2}\right) dtd\theta \textit{ with polar coordinates} \\ & = \int_0^r \frac{t}{\sigma^2} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt\end{aligned}$$

Many phenomena in various fields such as image processing, signal processing, particle physics, etc., follow a Rayleigh distribution.

2.2. Distance Distribution between Two Random Points Iud in a Region of $\mathbb{R}^n$

More recently, in the 20th century, spatial points' processes in one or two dimensions and related spatial properties, as void or contact distribution or Euclidian distance distribution between k-neighbors, started to receive special attention [8]. Nevertheless, "the research on distributions of distances in point processes of dimensions higher than one have never been an issue of systematic research and have been performed in rather ad hoc way in the past" ([1], p. 2). The problem was not addressed holistically, but depending on the field of application and the geometric form considered, in two or three dimensions, essentially circles and spheres or rectangles and cubes.

Distribution of Euclidian distances between two random points for *iud* set of points in a circle, sphere or hypersphere has been addressed many times in the literature in different fields and with different techniques (geometry, differential equations): in mathematics and statistics [9–13], in chromosome analysis [14], in geography [15], in demography [16–20], in network analysis [1], and in physics [21,22].

For example, in $\mathbb{R}^2$, for two random points $\vec{U}$ and $\vec{V}$ in a circle of radius R, geometric resolution of the distribution of $\| \vec{U} - \vec{V} \|$ described in [1] use the Croften's fixed-point theorem and the mean value theorem [23], and the result has been known since the end of the 19th century:

$$\forall t \in]\,0; 2R], \; f_{\|\vec{U}-\vec{V}\|}(t) = \frac{4t}{\pi R^2}\left(\arccos\left(\frac{t}{2R}\right) - \frac{t}{2R}\sqrt{1-\left(\frac{t}{2R}\right)^2}\right)$$

Again in $\mathbb{R}^2$, distributions of Euclidian distances between two random points in a rectangle has long been addressed [24], and an analytical resolution is presented in [25]. More recently, other studies have focused on polygons [26].

Distribution of Euclidian distances for a cube has also been addressed [27–30], but without general formulation for any dimensions. The distribution function for this random variable seems not to be known before 1978 [27]. Robbins's constant [31] was defined as the mean Euclidian distance between two random points in a unit cube.

Results for cubes have been compiled by EW Weisstein and presented at http://mathworld.wolfram.com/CubeLinePicking.html [32].

As one of the many random quantities studied in Geometric Probability, results were extended to the 4th and 5th dimensions [33] but for higher dimensions the increase of algebraic complexity associated with derivation procedures was a strong limiting factor. These results can have practical applications in multidimensional analysis and data mining.

3. A Unified Method for the Evaluation of Euclidian Distance Distributions between Two Randomly Chosen Points

We present now an original approach using a unified method generalizable to any type of compact set in any dimensions. Mathematical formalization and resolution will use only well-known objects and methods such as random variables and density functions, convolution, marginal distribution, and some standard functions (Gamma, Beta). We will use this approach to calculate distributions of Euclidian distances between two randomly chosen points for hypersphere and hypercube of any dimensions, and therefore confirm results already known in the literature as mentioned before.

3.1. Mathematical Formalization

Let $\vec{u}$ be a vector in a normed $\mathbb{R}$-vector space E of dimension n. Its coordinates are noted $(x_1, x_2, \ldots, x_n)$ in a orthonormal coordinate system, $\| \vec{u} \|$ the norm of $\vec{u}$, d the distance associated to the norm ($d(\vec{u}, \vec{v}) = \| \vec{u} - \vec{v} \|$. In this article, the Euclidean norm will be considered: $\| \vec{u} \| = \sqrt{{x_1}^2 + {x_2}^2 + \ldots + {x_n}^2}$). $B(\vec{u}, r)$ is the closed ball of center $\vec{u}$ and radius r, which corresponds to all the vectors $\vec{v}$ *de* E whose distance to $\vec{u}$ is less than or equal to r:

$$\forall r \in \mathbb{R}^+, \ \forall \vec{u} \in E, \ \vec{v} \in B(\vec{u}, r) \Leftrightarrow \| \vec{u} - \vec{v} \| \leq r$$

Let $\vec{U}$ and $\vec{V}$ denote two random vectors in E independent and identically distributed with the same probability distribution corresponding to a homogeneous Poisson point process $\mathcal{H}$ (a completely spatial random process). Random vectors $\vec{U}$ and $\vec{V}$ allow us to simulate the pairs $(\vec{u}, \vec{v})$ of elements of a subset F of E, and to evaluate the distribution of their distances $d = \| \vec{u} - \vec{v} \|$. Let D be the random variable in $\mathbb{R}^+$ such that: $D = \| \vec{U} - \vec{V} \|$. D represents the distance between two vectors obtained randomly in E. Our problem is to determine the probability density of D from the process $\mathcal{H}$, i.e., to determine the function f_D such that

$$\forall r \in \mathbb{R}^+, \ p(D \leq r) = \int_0^r f_D(t)dt$$

3.2. Using Convolution of Density Functions

Considering that $p(\| \vec{U} - \vec{V} \| \leq t) = p(\vec{U} - \vec{V} \in B(\vec{0}, t))$, finding the density function $f_{\vec{U} - \vec{V}}$ using $f_{\vec{U}}$ and $f_{\vec{V}}$ would lead to the expected distribution of D.

When E is uni-dimensional, $\vec{U}$ and $\vec{V}$ are simply independent $\mathbb{R}$-random variables U and V. Convolution can be used to find the density f_{U+V} from the density functions f_U and f_V [34],

$$\forall x \in \mathbb{R}, \ f_{U+V}(x) = (f_U * f_V)(x) = \int_{\mathbb{R}} f_U(y) f_V(\mathrm{x} - y)dy$$

Therefore, f_{U-V} can be seen as the convolution of f_U and f_{-V},

$$\forall x \in \mathbb{R}, \ f_{U-V}(x) = \int_{\mathbb{R}} f_U(y) f_V(y - x)dy$$

given that $\forall x \in \mathbb{R}, \ f_{-V}(x) = f_V(-x)$, which stays true even in higher dimensions.

When $\vec{U}$ and $\vec{V}$ are two independent n-dimensional random vectors in a vector space E, $(\vec{U}, \vec{V})$ is a $2n$-dimensional vector. Let A be the $(2n \times 2n)$ matrix such that $A \times (\vec{U}, \vec{V}) = (\vec{U} + \vec{V}, \vec{V})$,

$$A = \begin{pmatrix} I_n & I_n \\ 0 & I_n \end{pmatrix}$$

where I_n is the identity matrix in E.

We have

$$A^{-1} = \begin{pmatrix} I_n & -I_n \\ 0 & I_n \end{pmatrix} \text{and } \det(A) = \det(A^{-1}) = 1.$$

Therefore,

$$\begin{aligned} \forall (x,y) \in E^2,\ f_{(\vec{U}+\vec{V},\vec{V})}(x,y) &= \tfrac{1}{|\det(A)|} f_{(\vec{U},\vec{V})}(A^{-1}\dot{(}x,y)) \\ &= f_{(\vec{U},\vec{V})}(x-y,y) \\ &= f_{\vec{U}}(x-y) f_{\vec{V}}(y)\ (\textit{from independence of } \vec{\mathrm{U}} \textit{ and } \vec{\mathrm{V}}) \end{aligned} .$$

As a way of consequence, $f_{\vec{U}+\vec{V}}$ is the marginal distribution of $f_{(\vec{U}+\vec{V},\vec{V})}$:

$$\forall x \in E,\ f_{\vec{U}+\vec{V}}(x) = \int_E f_{\vec{U}}(y) f_{\vec{V}}(x-y)dy$$

Given that $\forall x \in \mathrm{E},\ f_{-V}(x) = f_V(-x)$ thus leads to

$$\forall x \in E,\ f_{\vec{U}-\vec{V}}(x) = \int_E f_{\vec{U}}(y) f_{\vec{V}}(y-x)dy$$

As such we can obtain the distribution of $\| \vec{U} - \vec{V} \|$ from $\vec{U}$ and $\vec{V}$ distributions only.

3.3. Distribution with Random Set of Points Iud in a Compact Set

Compact sets are convenient to model any type of spatial region of any shape and with finite size (which is always the case in reality). We assume in the following that the set F is a set of elements *iud* with uniform density ρ in a compact K, corresponding to a homogeneous Poisson process of density ρ on K. So, we have:

$$f_{\vec{U}} = f_{\vec{V}} = \frac{1}{\lambda(K)} 1_K \tag{1}$$

where 1_K is the indicator function of K, and λ the Lebesgue measure in E.

Using previously presented tools,

$$\begin{aligned} \forall \vec{x} \in E,\ f_{\vec{U}-\vec{V}}(\vec{x}) &= \textstyle\int_E f_{\vec{U}}(\vec{y}) f_{\vec{V}}(\vec{y}-\vec{x}) d\vec{y} \\ &= \textstyle\int_E \frac{1}{\lambda(K)^2} 1_K(\vec{y}) \cdot 1_K(\vec{y}-\vec{x}) d\vec{y} \\ &= \textstyle\frac{1}{\lambda(K)^2} \int_E 1_K(\vec{y}) \cdot 1_{K+x}(\vec{y}) d\vec{y} \\ &= \textstyle\frac{1}{\lambda(K)^2} \int_E 1_{K \cap (K+\vec{x})}(\vec{y}) d\vec{y} \\ &= \textstyle\frac{\lambda(K \cap (K+\vec{x}))}{\lambda(K)^2} \end{aligned} \tag{2}$$

where $K + \vec{x} = \{\vec{k} + \vec{x} | \vec{k} \in k\}$.

The distribution of $D = \| \vec{U} - \vec{V} \|$ follows:

$$\begin{aligned} \forall r \in \mathbb{R}^+,\ p(D \le r) &= p(\vec{U} - \vec{V} \in B(\vec{0}, r)) \\ &= \textstyle\frac{1}{\lambda(K)^2} \int_{B(\vec{0},r)} \lambda(\mathrm{K} \cap (\mathrm{K} + \vec{\mathrm{x}})) d\vec{x} \end{aligned} \tag{3}$$

Because λ is defined as a measure with translational invariance, one can be sure that $p(D \le r)$ is not affected by the position of K inside E but only by the "shape" of K and $K \cap (K + \vec{x})$. This

translational invariance is very intuitive; distances inside a spatial area are never affected by the global position of the area:

$$\forall(\vec{x},\vec{u})E^2,\ \lambda([K+\vec{u}]\cap[K+\vec{x}+\vec{u}])=\lambda([K\cap K+\vec{x}]+\vec{u})=\lambda(K\cap(K+\vec{x}))$$

4. Using Equation (3) for Typical Compact Sets

We will apply this general resolution formula for typical compact sets, especially to hypercubes and hyperspheres of any dimension.

4.1. K is a Segment in a 1-Dimension Space

In dimension 1, the compact K is a segment $[a,b]$ $(b>a)$ in $\mathbb{R}$. The set F stands for a set of random values uniformly distributed in $[a,b]$. $\vec{U}$ and $\vec{V}$ are simply independent $\mathbb{R}$-random variables U and V. We can easily determine the well-known density function [9] (Figure 2). We have

$$f_U=f_V=\frac{1}{\mathrm{b}-\mathrm{a}}1_{[a,b]}$$

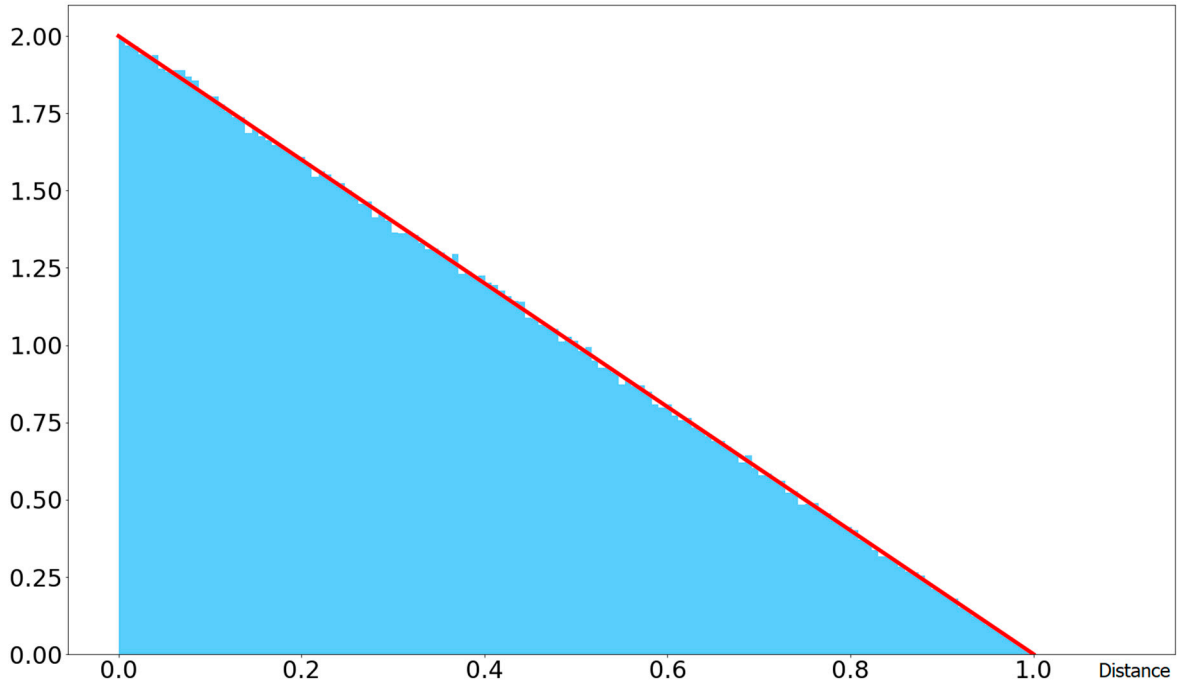


Figure 2. Distribution of distances from values generated by a homogeneous Poisson process in [0,1] with density $\rho=1500$. In blue is the distribution of observed distances and in red is the theorical distribution $f_{\mathrm{D}}(t)=2(1-\mathrm{t})$ in [0,1].

Due to translational invariance, $[a,b]$ can be replaced with $[0,\ell]$ *with* $\ell=b-a$.
From (3) we have:

$$\forall x\in\mathbb{R},\ f_{U-V}(x)=\frac{1}{\ell^2}\lambda([0,\ell]\cap[x,\mathrm{x}+\ell])$$

Therefore, if $|x|\geq\ell$, $f_{U-V}=0$. Otherwise, $f_{U-V}(x)=\frac{1}{\ell}(1-\frac{|x|}{\ell})$. Hence,

$$\forall x\in\mathbb{R},\ f_{U-V}(x)=\frac{1}{\ell}(1-\frac{|x|}{\ell})1_{[-\ell,\ell]}(x)$$

Then,

$$p(|U-V|\leq t)=\frac{1}{\ell}\int_{-t}^{t}(1-\frac{|x|}{\ell})1_{[-\ell,\ell]}(x)dx=\frac{2}{\ell}\int_{0}^{t}(1-\frac{\mathrm{x}}{\ell})1_{[0,\ell]}(x)dx$$

Hence,

$$\forall t\mathbb{R}^+,\ f_{|U-V|}(t) = \frac{2}{\ell}\left(1-\frac{t}{\ell}\right)1_{[0,\ell]}(t) \tag{4}$$

As the density is linear, it is easy to calculate the mean, variance and median (called m):

$$\mathbb{E}(|U-V|) = \frac{\ell}{3}$$

$$\mathrm{Var}(|U-V|) = \frac{\ell^2}{18}$$

$$m = \left(1-\frac{\sqrt{2}}{2}\right)\ell\ \ (\sim\ 0.293\ \ell)$$

4.2. K Is a Rectangle in a 2-Dimensional Space

The two-dimensional rectangle displays a highly convenient property, namely that the x-axis and y-axis are statistically independent.

Let us introduce parameters for our rectangle: $K = [\alpha;\alpha+a]\times[\beta;\beta+b]$ where $(a,b)\in\mathbb{R}^{+2}$ and $(\alpha,\beta)\in\mathbb{R}^2$. Thanks to the remark previously made about translational invariance, (α,β) can be replaced by $(0,0)$ without changing the final result. Here, Equation (1) gives

$$\begin{aligned}\forall(x,y)\in\mathbb{R}^2,\ f_{\vec{U}}(x,y) = f_{\vec{V}}(x,y) &= \tfrac{1}{ab}1_{[0;a]\times[0;b]}(x,y)\\ &= \tfrac{1}{ab}1_{[0;a]}(x)\cdot 1_{[0;b]}(y)\end{aligned}$$

Then, we have

$$\begin{aligned}&\forall(x_1,x_2)\in\mathbb{R}^2,\\ &f_{\vec{U}-\vec{V}}(x_1,x_2) = \tfrac{1}{(ab)^2}\int_{\mathbb{R}_2}1_{[0;a]}(y_1)\cdot 1_{[0;b]}(y_2)\cdot 1_{[x_1;a+x_1]}(y_1)\cdot 1_{[x_2;b+x_2]}(y_2)d(y_1,y_2)\\ &= \tfrac{1}{ab}\int_{\mathbb{R}_2}1_{[0;1]\cap[\frac{x_1}{a};1+\frac{x_1}{a}]}(y_1)\cdot 1_{[0;1]\cap[\frac{x_2}{b};1+\frac{x_2}{b}]}(y_2)d(y_1,y_2)\\ &= \tfrac{1}{ab}1_{[-a;a]}(x_1)\cdot 1_{[-b;b]}(x_2)\int_{\max(0;\frac{x_1}{a})}^{\min(1;1+\frac{x_1}{a})}dy_1\cdot\int_{\max(0;\frac{x_2}{b})}^{\min(1;1+\frac{x_2}{b})}dy_2\\ &= \tfrac{1}{ab}1_{[-a;a]}(x_1)\cdot 1_{[-b;b]}(x_2)(\min(1;1+\tfrac{x_1}{a})-\max(0;\tfrac{x_1}{a}))\cdot(\min(1;1+\tfrac{x_2}{b})-\max(0;\tfrac{x_2}{b}))\\ &= \tfrac{1}{ab}\cdot(1-\tfrac{|x_1|}{a})\cdot(1-\tfrac{|x_2|}{b})\end{aligned}$$

$f_{\vec{U}-\vec{V}}$ must be integrated on $B(\vec{0},t)$ in order to get the distribution $p(\|\vec{U}-\vec{V}\|\leq t)$ for $t\in\mathbb{R}^+$. By noticing that the longest distance inside K is $\sqrt{a^2+b^2}$, the distribution function becomes

$$\begin{aligned}&\forall t\in\left[0;\sqrt{a^2+b^2}\right],\\ &p(\|\vec{U}-\vec{V}\|\leq t) = \tfrac{1}{ab}\int_{B(\vec{0},r)}(1-\tfrac{|x_1|}{a})\cdot(1-\tfrac{|x_2|}{b})d(x_1,x_2)\\ &= \tfrac{1}{ab}\int_{-t}^{t}1_{[-b;b]}(x_2)\cdot(1-\tfrac{|x_2|}{b})\int_{-\sqrt{t^2-x_2^2}}^{\sqrt{t^2-x_2^2}}1_{[-a;a]}(x_1)\cdot(1-\tfrac{|x_1|}{a})d(x_1,x_2)\\ &= \tfrac{4}{ab}\int_0^t 1_{[0;b]}(x_2)\cdot(1-\tfrac{|x_2|}{b})\int_0^{\sqrt{t^2-x_2^2}}1_{[0;a]}(x_1)\cdot(1-\tfrac{|x_1|}{a})d\lambda(x_1,x_2)\\ &= \tfrac{4}{ab}\int_0^{\min(b;t)}(1-\tfrac{|x_2|}{b})\int_0^{\min(a;\sqrt{t^2-x_2^2})}(1-\tfrac{|x_1|}{a})d(x_1,x_2)\\ &= \tfrac{4}{ab}\int_0^{\min(b;t)}(1-\tfrac{x}{b})\cdot(1-\tfrac{1}{2a}\min(a;\sqrt{t^2-x^2}))\min(a;\sqrt{t^2-x^2})dx\end{aligned}$$

Clearly, we need to separate different cases:

- When $t \in [0; \min(a,b)]$, the calculation's results in a polynomial density function for $\| \vec{U} - \vec{V} \|$:

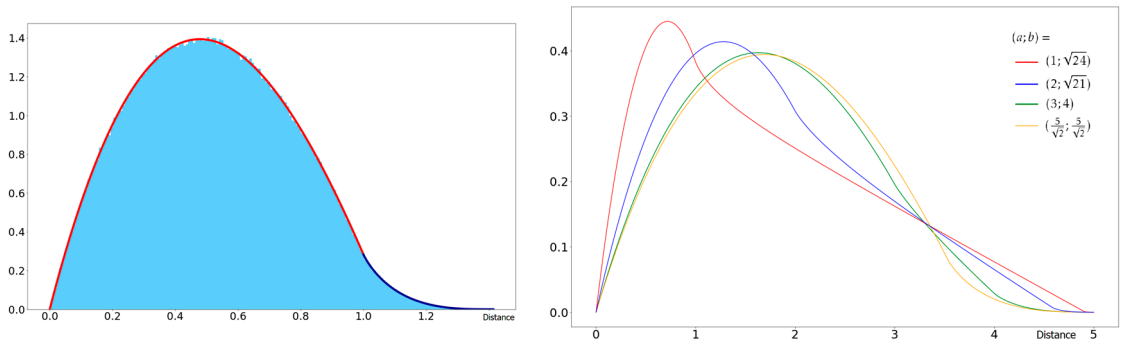
$$\forall x \in [0; \min(a;b)],\ f_{\|\vec{U}-\vec{V}\|}(x) = \frac{2x}{(ab)^2}(x^2 - 2(a+b)x + \pi ab) \tag{5}$$

- It is possible to calculate explicitly $p(\| \vec{U} - \vec{V} \| \leq t)$ for values of t in $[\min(a;b); \sqrt{a^2+b^2}]$. Nevertheless, the expression is not polynomial anymore and ends up being much less simple than the integral form.
- If $a < b$, the proportion of the sample that follows a polynomial distribution is given by

$$\frac{a}{b} \cdot (\pi - \frac{4}{3} - \frac{5a}{6b})$$

In the particular case where $a = b$ (i.e., K is a square), the polynomial function describes the first $\pi - 2 - \frac{1}{6} (\sim\ 97\%)$ of all distances.

As an example, Figure 3 shows the graph for density on a square (a) and inside some rectangles (b).



(**a**) (**b**)

Figure 3. (**a**) The distribution of distance on a square. The graph for density is in red when polynomial (5). The light blue bars show a simulation from values generated by a homogeneous Poisson process in $[0,1] \times [0,1]$ with density $\rho = 1500$. (**b**) Densities in rectangles verifying $\sqrt{a^2 + b^2} = 5$ (this condition allows the curves to share the same domain of definition). It appears that the more symmetrical the rectangle is, the rarer the longest and shortest distances are.

4.3. K Is a Disk in a 2-Dimensional Space

K is a disk with radius R,

$$\forall \vec{x} \in \mathbb{R}^2,\ f_{\vec{U}}(\vec{x}) = f_{\vec{V}}(\vec{x}) = \frac{1}{\pi R^2} 1_{B(\vec{c},R)}(\vec{x})$$

where $\vec{c}$ is the center of the circle and R its radius. It is obviously possible to limit our study to the case where $\vec{c} = \vec{0}$.

From (3) we have

$$\begin{aligned}
\forall \vec{x} \in \mathbb{R}^2,\ f_{\vec{U}-\vec{V}}(\vec{x}) &= \left(\tfrac{1}{\pi R^2}\right)^2 \int_{\mathbb{R}^2} 1_{B(\vec{0},R)}(\vec{y}) 1_{B(\vec{0},R)}(\vec{y} - \vec{x}) d(\vec{y}) \\
&= \left(\tfrac{1}{\pi R^2}\right)^2 \int_{\mathbb{R}^2} 1_{B(\vec{0},R)}(\vec{y}) 1_{B(\vec{x},R)}(\vec{y}) d(\vec{y}) \\
&= \tfrac{R^2}{\pi^2 R^4} \int_{\mathbb{R}^2} 1_{B(\vec{0},1)}(\vec{y}) 1_{B(\frac{\vec{x}}{R},1)}(\vec{y}) d(\vec{y}) \\
&= \tfrac{1}{(\pi R)^2} \lambda(B(\vec{0},1) \cap B(\tfrac{\vec{x}}{R},1))
\end{aligned}$$

To calculate f_D we need to integrate the area $S = \lambda(B(\vec{0},1) \cap B(\frac{\vec{x}(r,\theta)}{R},1))$ on all angles $\theta \in]-\pi;\pi]$.
We can demonstrate the isotropy of S (the area remains the same for any value of θ).
Let $M(\theta)$ be the rotation operator,

$$\begin{aligned}\forall\theta \in]-\pi;\pi],\ \lambda(B(\vec{0},1) \cap B(\tfrac{1}{R}\vec{x}(r,\theta),1)) &= \lambda[M(\theta)\cdot(B(\vec{0},1) \cap B(\tfrac{1}{R}\vec{x}(r,0),1))] \\ &= \left|det(M(\theta))\right|\cdot\lambda(B(\vec{0},1) \cap B(\tfrac{1}{R}\vec{x}(r,0),1)) \\ &= \lambda(B(\vec{0},1) \cap B(\tfrac{1}{R}\vec{x}(r,0),1))\end{aligned}$$

Thus,

$$\begin{aligned}\forall t \in \mathbb{R}^+,\ f_{\|\vec{U}-\vec{V}\|}(t) &= t\int_0^{2\pi} \lambda(K \cap [K + x(t,\theta)])d\theta \\ &= \tfrac{2\pi t}{(\pi R)^2}\lambda\Big(B(\vec{0},1) \cap B(\tfrac{1}{R}\vec{x}(t,0),1)\Big) = \tfrac{2\pi t}{(\pi R)^2}S\end{aligned}$$

To calculate the surface S, let $l(u)$ be the length of the chord for the u coordinate on the x-axis [35]:

$$\begin{aligned}\forall r \in]0;2R],\ S &= 2\int_{\frac{r}{2R}}^{1} l(u)du \\ &= 2\int_{\frac{r}{2R}}^{1} \sqrt{1-u^2}du \\ &= 2\int_0^{\arccos\frac{r}{2R}} \sin^2\theta d\theta \\ &= 2\left[\arccos\left(\tfrac{r}{2R}\right) - \tfrac{r}{2R}\sqrt{1-\left(\tfrac{r}{2R}\right)^2}\right]\end{aligned}$$

This finally leads to the already-mentioned result above,

$$\forall t \in]0;2R],\ f_{\|\vec{U}-\vec{V}\|}(t) = \frac{4t}{\pi R^2}\left(\arccos\left(\frac{t}{2R}\right) - \frac{t}{2R}\sqrt{1-\left(\frac{t}{2R}\right)^2}\right) \tag{6}$$

$2R$ being the longest distance possible inside a circle.
Figure 4 shows the graph of distribution of distances inside a circle.

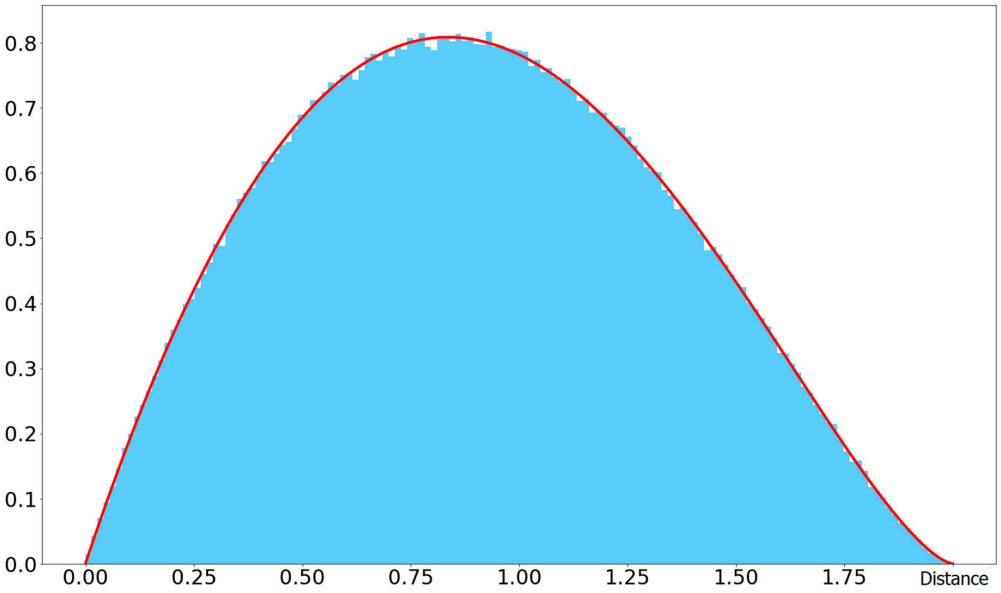


Figure 4. Distribution of distances inside a circle (R = 1). In red the graph for the density (6), in light blue a simulated distribution from values generated by a homogeneous Poisson process with density $\rho = 1500$ inside the circle.

We can calculate the mean and variance of this distribution:

$$\begin{gathered}\forall R \in \mathbb{R}_+^*,\ \mathbb{E}(\| \vec{U} - \vec{V} \|) = \int_0^{2R} t f_{\|\vec{U}-\vec{V}\|}(t)dt \\ = \frac{2^5}{\pi} R \int_0^{\frac{\pi}{2}} \sin(\theta)\cos^2(\theta)(\theta - \sin^2(\theta))d\theta \\ = \frac{2^7}{45\pi} R (\simeq 0.905{\cdot}R)\end{gathered}$$

If m stands for the median, solving $\int_0^m f_{\|\vec{U}-\vec{V}\|}(t)dt = \frac{1}{2}$ for m turns out to be unsolvable analytically. Nevertheless, we can show that

$$\forall R \in \mathbb{R}_+^*,\ \int_0^{\mu} f_{\|\vec{U}-\vec{V}\|}(t)dt \simeq 0.511$$

where μ is the mean of the distribution. The median is only very slightly lower than mean.

To calculate the variance, we calculate first $\mathbb{E}(\| \vec{U} - \vec{V} \|^2)$,

$$\begin{gathered}\forall R \in \mathbb{R}_+^*,\ \mathbb{E}(\| \vec{U} - \vec{V} \|^2) = \int_0^{2R} t^2 f_{\|\vec{U}-\vec{V}\|}(t)dt \\ = \frac{2^6}{\pi} R^2 \int_0^{\frac{\pi}{2}} \sin(\theta)\cos^3(\theta)(\theta - \cos(\theta)\sin(\theta))d\theta \\ = R^2\end{gathered}$$

which is quite remarkable. Therefore,

$$\mathrm{Var}(\| \vec{U} - \vec{V} \|) = \left(1 - \left(\frac{2^7}{45\pi}\right)^2\right){\cdot}R^2 \simeq 0.180{\cdot}R^2$$

Let us check, as an exercise, that the density is indeed normalized:

$$\begin{gathered}\forall R \in \mathbb{R}_+^*,\ \frac{4}{\pi R^2} \int_0^{2R} t\left(\arccos\left(\frac{t}{2R}\right) - \frac{t}{2R}\sqrt{1 - \left(\frac{t}{2R}\right)^2}\right)dt \\ = \frac{16}{\pi} \int_0^{\frac{\pi}{2}} \sin(\theta)\cos(\theta)(\theta - \cos(\theta)\sin(\theta))d\theta \\ = \frac{16}{\pi}\left(\int_0^{\frac{\pi}{2}} \theta{\cdot}\partial_\theta\left(\frac{1}{2}\sin^2(\theta)\right)d\theta - \int_0^{\frac{\pi}{2}} \frac{1}{4}\sin^2(2\theta)d\theta\right) \\ = \frac{16}{\pi}\left(\frac{\pi}{4} - \frac{1}{4}\int_0^{\frac{\pi}{2}} (1 - \cos(2\theta))d\theta - \frac{1}{8}\int_0^{\frac{\pi}{2}} (1 - \cos(4\theta))d\theta\right) \\ = \frac{16}{\pi}\left(\frac{\pi}{4} - \frac{\pi}{8} - \frac{\pi}{16}\right) = 1\end{gathered}$$

4.4. K Is a Sphere in a Three-Dimensional Space

When using Formula (2), it is clear that when a space's dimension is equal to three or more, the issue is simply to calculate a volume (or hypervolume). Calculations are quite similar to those used for the circle. Firstly, let us give random vectors $\vec{U}$ and $\vec{V}$'s density,

$$\forall \vec{x} \in \mathbb{R}^3,\ f_{\vec{U}}(\vec{x}) = f_{\vec{V}}(\vec{x}) = \frac{1}{\frac{4}{3}\pi R^3} 1_{B(\vec{c},R)}(\vec{x})$$

where $B(\vec{c}, R)$ is the 3D sphere centered at $\vec{c}$ with a radius R. Just like previously, replacing $\vec{c}$ with $\vec{0}$ will not change any results. We have:

$$\begin{gathered}\forall \vec{x} \in \mathbb{R}^3,\quad f_{\vec{U}-\vec{V}}(\vec{x}) = \left(\frac{1}{\frac{4}{3}\pi R^3}\right)^2 \int_{\mathbb{R}^3} 1_{B(\vec{0},R)}(\vec{y}) 1_{B(\vec{0},R)}(\vec{y} - \vec{x})d(\vec{y}) \\ = \left(\frac{1}{\frac{4}{3}\pi R^3}\right)^2 \int_{\mathbb{R}^3} 1_{B(\vec{0},R)}(\vec{y}) 1_{B(\vec{x},R)}(\vec{y})d(\vec{y}) \\ = \left(\frac{3}{4\pi R^3}\right)^2 R^3 \lambda(B(\vec{0}, 1) \cap B(\tfrac{\vec{x}}{R}, 1))\end{gathered}$$

Using spherical coordinates (r, θ, ϕ) gives us

$$\begin{aligned}\forall t \in \mathbb{R}^+, \mathbb{P}(\| \vec{U} - \vec{V} \| \leq t) &= \frac{9}{16\pi^2 R^3} \int_{B(\vec{0},t)} \lambda(B(\vec{0},1) \cap B(\frac{\vec{x}}{R},1))dt \\ &= \frac{9}{16\pi^2 R^3} \int_0^t t^2 \Big(\int_0^{2\pi} \int_0^{\pi} \sin(\phi) \lambda(B(\vec{0},1) \cap B(\frac{\vec{x}}{R},1))d\theta d\phi \Big) dt\end{aligned}$$

giving us directly the density $f_{\|\vec{U}-\vec{V}\|}$. A rotation matrix determinant is still equal to 1. Therefore angles (θ, ϕ) will not change the value of the volume:

$$\forall t \in [0; 2R],\ f_{\|\vec{U}-\vec{V}\|}(t) = \frac{9}{16\pi^2 R^3} \cdot (4\pi t^2) \cdot \lambda \left(B(\vec{0},1) \cap B\left(\frac{\vec{x}(r,0,0)}{R}, 1 \right) \right)$$

The volume of the intersection can be calculated similarly to the previous two-dimensional area. The difference is that for every $l(y)$ in our circle, there is a whole disc for every y. Therefore:

$$\begin{aligned}\lambda(B(\vec{0},1) \cap B\Big(\tfrac{\vec{x}(r,0,0)}{R},1\Big)) &= 2\int_{\frac{r}{2R}}^1 \pi l(u)^2 du \\ &= 2\int_{\frac{r}{2R}}^1 \pi(1-u^2)du \\ &= 2\pi\Big(\tfrac{1}{3}(\tfrac{r}{2R})^3 - \tfrac{r}{2R} + \tfrac{2}{3}\Big)\end{aligned}$$

Finally,

$$\forall t \in [0; 2R],\ f_{\|\vec{U}-\vec{V}\|}(t) = \frac{3}{2R^3} t^2 \left(\left(\frac{t}{2R} \right)^3 - 3\left(\frac{t}{2R} \right) + 2 \right)$$

The density is polynomial in a three-dimensional space. Clearly, third dimension favors the presence of longer distances:

$$\forall R \in \mathbb{R}_+^*,\ \mathbb{E}(\| \vec{U} - \vec{V} \|) = \left(1 + \frac{1}{35} \right) R\ (\simeq 1.029\ R)$$

which is greater than the expected value calculated in a two-dimensional space. Calculating the median implies solving for m a polynomial equation of degree 6:

$$2\left(\frac{m}{2R} \right)^6 - 9\left(\frac{m}{2R} \right)^4 + 8\left(\frac{m}{2R} \right)^3 = \frac{1}{2}$$

Unfortunately it is impossible to give a general solution. Nevertheless, $m_{app} = 1.033 \cdot R$ gives a very good approximation of m. This estimation shows that this time, the median is greater than the mean. It is remarkable that in contrary to 2D, 3D distances are a slightly more likely to be longer than average.

Finally, let us calculate the variance

$$\forall R \in \mathbb{R}_+^*,\ \mathrm{Var}(\| \vec{U} - \vec{V} \|) = \left(1 - \frac{1}{175} \right) \frac{R^2}{7}\ (\simeq 0.142 \cdot R^2)$$

which is as expected significantly lower than two-dimensional variance.

4.5. K Is a Hypersphere in Higher Dimensions (n > 3)

Random vectors $\vec{U}$ and $\vec{V}$ are now considered independently and uniformly distributed in $B(\vec{c}, R) \subset \mathbb{R}^n$ where $n \in \mathbb{N}, n \geq 4$. A generalization of previous methods allows us to expect $f_{\|\vec{U}-\vec{V}\|}$ to take the following form:

$$\forall t \in [0; 2R], f_{\|\vec{U}-\vec{V}\|}(t) = \gamma_n t^{n-1} \int_{\frac{t}{2R}}^1 \mathrm{Vol}_{n-1}(\sqrt{1-y^2}) dy$$

where γ_n is a constant and Vol_n the volume of hypersphere in n-dimensional space:

$$\mathrm{Vol}_n(R) = \frac{2\pi^{\frac{n}{2}}}{n\Gamma(\frac{n}{2})}R^n \tag{7}$$

Therefore, we have

$$\forall t \in [0;2R],\ f_{\|\vec{U}-\vec{V}\|}(t) = \gamma_n t^{n-1}\int_{\frac{t}{2R}}^{1}(1-y^2)^{\frac{n-1}{2}}dy$$

where γ_n can be evaluated as the normalizing constant. The generalized binomial theorem [36] gives us the way to evaluate the integral,

$$\int_{\frac{t}{2R}}^{1}(1-y^2)^{\frac{n-1}{2}}dy = \sum_{k=0}^{+\infty}\binom{\frac{n-1}{2}}{k}\frac{(-1)^k}{2k+1}(1-\frac{t}{2R}^{2k+1})$$

where $\binom{\frac{n-1}{2}}{k}$ is the generalized binomial factor $\binom{\nu}{k} = \frac{\nu(\nu-1)\ldots(\nu-k+1)}{k!}$ with $k \in \mathbb{N}$ and $\nu \in \mathbb{R}$.

Using the fact that $\int_1^{2R} f_{\|\vec{U}-\vec{V}\|}(t)dt = 1$, we have

$$\begin{aligned}\frac{1}{\gamma_n} &= \frac{(2R)^n}{n}\sum_{k=0}^{+\infty}\binom{\frac{n-1}{2}}{k}\frac{(-1)^k}{2k+1}\int_0^1 y^{n-1}(1-y^{2k+1})dy\\ &= \frac{(2R)^n}{n}\int_0^1 y^n\sum_{k=0}^{+\infty}\binom{\frac{n-1}{2}}{k}(-1)^k y^{2k}dy\\ &= \frac{(2R)^n}{n}\int_0^1 x^{\frac{n-1}{2}}(1-x)^{\frac{n-1}{2}}dx\\ &= \frac{(2R)^n}{n}\beta(\tfrac{n+1}{2},\tfrac{n+1}{2})\end{aligned}$$

where β is Euler's Beta function [37].

This gives the explicit form of density function $f_{\|\vec{U}-\vec{V}\|}$ of a random variable $D_n = \|\vec{U}-\vec{V}\|$ in dimension n (Figure 5a)

Therefore, the mean of D_n can be calculated using

$$\begin{aligned}\mathbb{E}(D_n) &= \frac{\gamma_n}{2}\frac{(2R)^{n+1}}{n+1}\beta(\tfrac{n+2}{2},\tfrac{n+1}{2})\\ &= 2R\frac{n}{n+1}\frac{\Gamma(\frac{n+2}{2})}{\Gamma(\frac{n+1}{2})}\frac{\Gamma(n+1)}{\Gamma(n+\frac{3}{2})}\end{aligned}$$

where Γ is Euler Gamma function.

Using Stirling approximation [35], we have $\lim_{q\to\infty}\mathbb{E}(D_n) = R\sqrt{2}$.

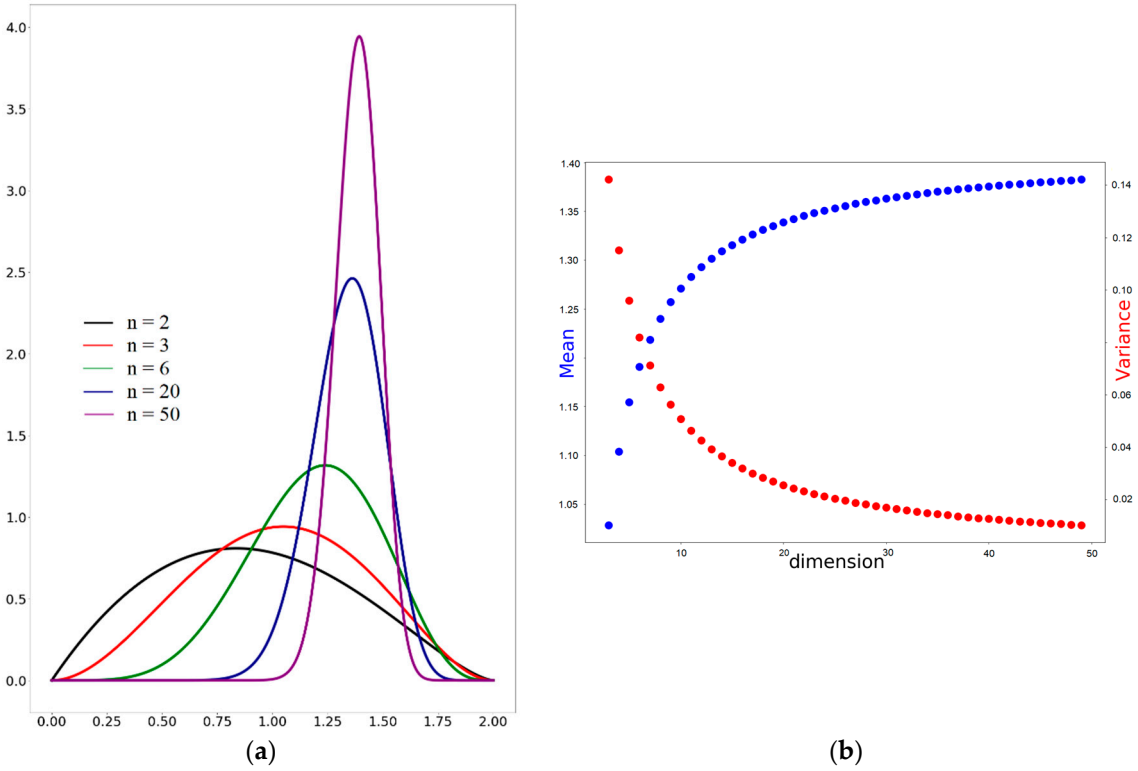
On the other hand,

$$\begin{aligned}\mathbb{E}(D_n^2) &= \frac{\gamma_n}{2}\frac{(2R)^{n+2}}{n+2}\beta(\tfrac{n+3}{2},\tfrac{n+1}{2})\\ &= 2R^2(1-\tfrac{2}{n+2})\end{aligned}$$

We have then $\lim_{n\to\infty}(\mathbb{E}(D_n^2)) = \lim_{n\to\infty}(\mathbb{E}((D_n)^2) = 2R^2$, which implies that $\lim_{n\to\infty}(\mathrm{Var}(D_n)) = 0$ (Figure 5b).

As we can see with Equation (7), $\lim_{n\to\infty}\mathrm{Vol}_n(R) = 0$, and it is well-known that hyperspheres become "hollow" when the dimension is high enough [38], and most points in a hypersphere tend to agglomerate towards its hypersurface. This has a consequence on distances that is quite intuitive and explains why the variance of distances tends to zero when the dimension increases, i.e., diversity of geometric configurations is increasingly limited as the dimension in space increases. Our result

shows how fast this phenomenon impacts the distances between points inside the hypersurface when dimension increases.



(**a**) (**b**)

Figure 5. (**a**) Distribution of distances inside a hypersphere ($R = 1$) for different values of space dimension n. (**b**) Mean value of distance tends to $R\sqrt{2}$ while variance tends to 0 when $n \to +\infty$ (n = 50 in the graphs).

5. Conclusions

We have developed a general and unified method to obtain the distribution of distances between two points randomly selected in a *iud* cloud of points in a geometric figure. These distributions are useful, especially in spatial statistics, to know the statistical representativeness (the weight) of a distance between two points. In the case of *iud* set of random points in a hypersphere, the expression of density is given for any dimension, and the variance of these distributions converge to zero when the dimension increases. This result also opens new perspectives in multidimensional analysis and data mining.

Author Contributions: Authors contribute equally to this work (conceptualization: M.S.; Formal analysis: S.L. and M.S.; Methodology: S.L.; Supervision: M.S.). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moltchanov, D. Distance distribution in random networks. *Ad Hoc Netw.* **2012**, *10*, 1146–1166. [CrossRef]
2. Souris, M.; Demoraes, F. Improvement of Spatial Autocorrelation, Kernel Estimation, and Modeling Methods by Spatial Standardization on Distance. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 199. [CrossRef]
3. Fotheringham, S.; Rogerson, P.A. (Eds.) *The Sage Handbook of Spatial Analysis*; Sage: London, UK, 2009.
4. Shabenberger, O.; Gotway, C. *Statistical Methods for Spatial Data Analysis*; Chapman & Hall, CRC Press: Boca Raton, FL, USA, 2005.

5. Souris, M. *Epidemiology and Geography. Principles, Methods and Tools of Spatial Analysis*; Wiley-ISTE: Fort Wayne, IN, USA, 2019.
6. Stoyan, D.; Kendall, W.; Mecke, J. *Stochastic Geometry and Its Applications*, 2nd ed.; Wiley: Hoboken, NJ, USA, 1995.
7. Pearson, K. The problem of the Random Walk. *Nature* **1905**, *72*, 318. [CrossRef]
8. Thompson, H. Distribution of distance to nth neighbour in a population of randomly distributed individuals. *Ecology* **1956**, *37*, 391–394. [CrossRef]
9. Borel, E. *Principes et Formules Classiques du Calcul des Probabilités*; GauthierVillars: Paris, France, 1924.
10. Garwood, F. The variance of the overlap of geometrical figures with reference to a bombing problem. *Biometrika* **1947**, *34*, 1–17. [CrossRef] [PubMed]
11. Hammersley, J.M. The distribution of Distance in a hypersphere. *Ann. Math. Stat.* **1950**, *21*, 447–452. [CrossRef]
12. Lord, R.D. The Distribution of Distance in a Hypersphere. *Ann. Math. Stat.* **1954**, *25*, 794–798. [CrossRef]
13. Alagar, V.S. The distribution of the distance between random points. *J. Appl. Probab.* **1976**, *13*, 558–566. [CrossRef]
14. Barton, D.; David, E.; Fix, F. Random points in a circle and the analysis of chromosome patterns. *Biometrika* **1963**, *50*, 23–29. [CrossRef]
15. Kuiper, J.H.; Paelinck, J.H. Frequency distribution of distances and related concepts. *Geogr. Anal.* **1982**, *14*, 253–259. [CrossRef]
16. Thanh, L.M. Distribution Théorique des Distances Entre deux Points Répartis Uniformément sur le Territoire. In *Les Déplacements Humains, Entretiens de Monaco en Sciences Humaines*; Sutter, J., Ed.; Hachette: Paris, France, 1962; pp. 173–184.
17. Courgeau, D. Migrations et découpages du territoire. *Population* **1973**, *28*, 511–537. [CrossRef]
18. Courgeau, D.; Baccaïni, B. Migrations et distances. *Population* **1989**, *42*, 57–82. [CrossRef]
19. Rogerson, P. Buffon's needle and the estimation of migration distances. *Math. Popul. Stud.* **1990**, *2*, 229–238. [CrossRef] [PubMed]
20. Cohen, J.; Courgeau, D. Modeling distances between humans using Taylors's law and geometric probability. *Math. Popul. Stud.* **2017**, *24*, 197–218. [CrossRef]
21. Tu, S.-J.; Fischbach, E. Random distance distribution for spherical objects: general theory and applications to physics. *J. Phys. A Math. Gener.* **2002**, *35*, 6557–6570. [CrossRef]
22. Parry, M.; Fischbach, E. Probability distribution of distance in a uniform ellipsoid: theory and applications to physics. *J. Math. Phys.* **2000**, *41*, 2417–2433. [CrossRef]
23. Crofton, M. *Probability, in Encyclopaedia Britannica*, 9th ed.; Britannica Inc.: Chicago, IL, USA, 1885.
24. Miller, L. Distribution of link distances in a wireless network. *J. Res. Natl. Inst. Stand Technol.* **2001**, *106*, 401–412. [CrossRef]
25. Kostin, A. Probability distribution of distance between pairs of nearest stations in wireless network. *Electron. Lett.* **2010**, *46*, 1299–1300. [CrossRef]
26. Hsu, A. The Expected Distance between Two Random Points in a Polygon. Master's Thesis, Massachussets Institute of Technology (MIT), Cambridge, MA, USA, 1990.
27. Robbins, D.P.; Bolis, T.S. Solution to problem E2629: Average Distance between Two Points in a Box. *Am. Math. Month.* **1978**, *85*, 277–278. [CrossRef]
28. Mathai, A.M.; Moschopoulos, P.; Pederzoli, G. Random points associated with rectangles. *Rendiconti del Circolo Matematico Di Palermo* **1999**, *48*, 162–190. [CrossRef]
29. Mathai, A.; Moschopoulos, P.; Pederzoli, G. Distance between random points in a cube. *Statistica* **1999**, *59*, 61–81.
30. Philip, J. *The Probability Distribution of the Distance between Two Random Points in a Box*; Department of Mathematics, Royal Institute of Technology: Stockholm, Sweden, 2007.
31. Le Lionnais, F. *Les Nombres Remarquables*; Hermann: Paris, France, 1983.
32. Weisstein, E.W. "Cube Line Picking" From MathWorld—A Wolfram Web Resource. Available online: http://mathworld.wolfram.com/CubeLinePicking.html (accessed on 1 April 2019).
33. Philip, J. *The Distance between Two Random Points in a 4- and 5-Cube*; Department of Mathematics, Royal Institute of Technology: Stockholm, Sweden, 2008.
34. Garet, O.; Kurtzmann, A. *De L'intégration Aux Probabilités*; Ellipses: Paris, France, 2011.

35. Harris, J.W.; Stocker, H. Spherical Zone (Spherical Layer). In *Handbook of Mathematics and Computational Science*; Springer-Verlag: New York, NY, USA, 1998.
36. Graham, R.L.; Knuth, D.E.; Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed.; Addison-Wesley: Reading, MA, USA, 1994.
37. Andrews, G.E.; Askey, R.; Roy, R. *Special Functions*; Cambridge University Press: London, UK, 1999.
38. Berger, M.; Gostiaux, B. *Géométrie Différentielle*; Armand Colin: Paris, France, 1972.