stichting
mathematisch
centrum

MC

AFDELING ZUIVERE WISKUNDE
(DEPARTMENT OF PURE MATHEMATICS)

ZN 84/78 JULI

J. VAN DE LUNE

AVERAGE DISTANCES IN/ON CERTAIN n-DIMENSIONAL BODIES

2e boerhaavestraat 49 amsterdam

Average Distances in/on certain n-dimensional bodies

by

J. van de Lune

ABSTRACT

This note contains some theorems and conjectures concerning the average distance between two independent random shots in/on n-dimensional cubes and spheres.

KEY WORDS AND PHRASES: *average distances, Monte Carlo methods, concavity:*

1. RANDOM DISTANCES IN THE N-DIMENSIONAL UNIT CUBE.

For $n = 1,2,3,\ldots$ let $\rho(n)$ be the mathematical expectation of the distance between two independent (homogeneously distributed) random points in the n-dimensional unit cube $I^n := [0,1]^n$, i.e.

$$\rho(n) := \int_{I^{2n}} \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \prod_{i=1}^{n} dx_i dy_i .$$

For example

$$\rho(1) = \int_0^1 \int_0^1 \sqrt{(x-y)^2}\, dx\, dy = 2 \int_0^1 dx \int_0^x (x-y)\, dy = \frac{1}{3} .$$

Numerical computations indicate that, for example,

$$\rho(2) \cong .52,\ \rho(3) \cong .66 \text{ and } \rho(4) \cong .77.$$

From the definition of $\rho(n)$ it is easily seen that

$$\rho(1) < \rho(2) < \rho(3) < \ldots$$

and

$$\rho(n) < \sqrt{n}.$$

Defining

$$q(n) := \frac{\rho(n)}{\sqrt{n}}$$

we thus have

$$q(n) < 1.$$

Computational work indicates that

$$q(1) < q(2) < q(3) < \ldots$$

In addition, Monte Carlo estimates of $\rho(n)$ indicate that the sequence $\{q(n)\}_{n=1}^{\infty}$ is even *concave*. This leads us to the following

CONJECTURE. The sequence $\{q(n)\}_{n=1}^{\infty}$ is (i) increasing and (ii) concave.

Since the sequence under consideration is bounded it is clear that (ii) implies (i). However, we were unable to prove any part of the above conjecture.

More positively we have the following

THEOREM 1.1. $\lim_{n\to\infty} q(n) = \frac{1}{\sqrt{6}}$.

PROOF. Let $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$ be mutually independent stochastic variables all of them being homogeneously distributed on the interval $I := [0,1]$. Let the stochastic variable s_n be defined by

$$s_n := \sum_{i=1}^{n} (x_i - y_i)^2,$$

so that

$$q(n) = E\left(\sqrt{\frac{s_n}{n}}\right).$$

Observing that

$$\mu := E(x_i - y_i)^2 = E(x_i^2 + y_i^2 - 2x_i y_i) =$$

$$= 2\,E(x_i^2) - 2\,E^2(x_i) = \frac{2}{3} - \frac{2}{4} = \frac{1}{6}$$

and

$$\sigma^2 := \mathrm{Var}\,(x_i - y_i)^2 = E((x_i - y_i)^2 - \mu)^2 =$$

$$= E(x_i - y_i)^4 - \frac{1}{36} = \frac{7}{180}$$

we find that

$$\mu_n := E(s_n) = n\mu = \frac{n}{6}$$

and

$$\sigma_n^2 := \mathrm{Var}(s_n) = n\sigma^2 = \frac{7n}{180} .$$

In order to complete our proof we make use of Chebychef's inequality, saying: If x is a real random variable with cumulative probability distribution F(x) and

$$\mu := E(x) := \int_{-\infty}^{\infty} x \, dF(x)$$

(the integral being absolutely convergent)

and

$$\sigma^2 := E(x-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \, dF(x) > 0,$$

then for any $a > 0$ we have

$$\int_{|x-\mu| \geq a\sigma} dF(x) \leq \frac{1}{a^2} .$$

Denoting the cumulative probability distribution of s_n by $F_n(s)$ we have

$$q(n) = E\left(\sqrt{\frac{s_n}{n}}\right) = \int_{-\infty}^{\infty} \sqrt{\frac{s}{n}} \, dF_n(s).$$

Now observe that for $a > 0$

$$E\left(\sqrt{\frac{s_n}{n}}\right) = \int_{|s-\mu_n| < a\sigma_n} \sqrt{\frac{s}{n}} \, dF_n(s) + \int_{|s-\mu_n| \geq a\sigma_n} \sqrt{\frac{s}{n}} \, dF_n(s) \leq$$

(since $0 \leq s \leq n$ and $\sqrt{\frac{s}{n}}$ is increasing in s for $s \geq 0$)

$$\leq \int_{|s-\mu_n| \geq a\sigma_n} dF_n(s) + \int_{|s-\mu_n| < a\sigma_n} \sqrt{\frac{\mu_n + a\sigma_n}{n}}\, dF_n(s) \leq$$

$$\leq \frac{1}{a^2} + \sqrt{\frac{\mu_n + a\sigma_n}{n}} \int_{|s-\mu_n| < a\sigma_n} dF_n(s) \leq \frac{1}{a^2} + \sqrt{\frac{1}{6} + \frac{a\sigma}{\sqrt{n}}}\,,$$

from which it is clear that

$$\limsup_{n\to\infty} E\left(\sqrt{\frac{s_n}{n}}\right) \leq \frac{1}{\sqrt{6}}\,.$$

On the other hand we have for any fixed $a > 0$ (and n large enough)

$$E\left(\sqrt{\frac{s_n}{n}}\right) = \left\{ \int_{|s-\mu_n| < a\sigma_n} + \int_{|s-\mu_n| \geq a\sigma_n} \right\} \sqrt{\frac{s}{n}}\, dF_n(s) \geq$$

$$\geq \int_{|s-\mu_n| < a\sigma_n} \sqrt{\frac{s}{n}}\, dF_n(s) \geq \int_{|s-\mu_n| < a\sigma_n} \sqrt{\frac{\mu_n - a\sigma_n}{n}}\, dF_n(s) =$$

$$= \sqrt{\frac{\mu_n - a\sigma_n}{n}} \left\{ 1 - \int_{|s-\mu_n| \geq a\sigma_n} dF_n(s) \right\} \geq$$

$$\geq \sqrt{\frac{1}{6} - \frac{a\sigma}{\sqrt{n}}} \left\{ 1 - \frac{1}{a^2} \right\},$$

from which it is clear that

$$\liminf_{n\to\infty} E\left(\sqrt{\frac{s_n}{n}}\right) \geq \frac{1}{\sqrt{6}}\,,$$

completing our proof.

The reader will find no difficulties to construct more general versions of the above theorem.

Similarly one might investigate the average distance between two independent

random shots on the boundary of the n-dimensional unit cube.

We conclude this section by stating some related observations.
In order to compute the multiple integrals $\rho(n)$ one may use some discretization process (or the Monte Carlo method).
In relation to $\rho(1)$ we observe that

$$d_1(m) := \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left| \frac{i}{m} - \frac{j}{m} \right| = \frac{1}{3} \left(1 - \frac{1}{m^2} \right)$$

which converges concavely to its limit $\rho(1)$ as $m \to \infty$
In the 2-dimensional case we have to deal with the sequence $\{d_2(m)\}_{m=1}^{\infty}$ where

$$d_2(m) := \frac{1}{m^5} \sum_{i1=1}^{m} \sum_{i2=1}^{m} \sum_{j1=1}^{m} \sum_{j2=1}^{m} \sqrt{(i1-j1)^2 + (i2-j2)^2} ,$$

and numerical computations suggest that this sequence is also tending concavely (and hence increasingly) to its limit $\rho(2)$.
Similar observations were made in the 3,4 and 5-dimensional cases, the higher dimensional cases being too much time consuming for any reasonable numerical verification.

TABLE of $d_n(m)$

n / m	2	3	4	5
1	.000000	.000000	.000000	.000000
2	.426777	.560918	.669171	.760917
3	.484437	.622396	.734025	.829977
4	.502004	.640685	.753834	.851741
5	.509513	.648578	.762597	
6	.513386	.652712	.767263	
7	.515638	.655153		
8	.517060	.656716		
9	.518015	.657779		
10	.518687	.658534		
11	.519178	.659090		
12	.519546	.659511		
13	.519831			
14	.520055			
15	.520234			
16	.520380			
17	.520500			
18	.520600			
19	.520685			
20	.520757			
21	.520818			
22	.520871			
23	.520918			
24	.520958			
25	.520994			

However, so far we were unable to give a theoretical explanation for the numerically observed concavity of the sequences $\{d_n(m)\}_{m=1}^{\infty}$, where n is a fixed positive integer greater than 1.
Similarly one might investigate the set of lattice points on the boundary of the cube $[1,m]^n$.

2. Random distances in/on n-dimensional spheres.

Let B_1 and B_2 be two closed solid spheres, with radius R_1 resp. R_2, in the n-dimensional space $\mathbb{R}^n$ and let the distance between the centers of B_1 and B_2 be d. Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be two independent (homogeneously distributed) random points in B_1 resp. B_2. Denoting the mathematical expectation of the distance between x and y by $\rho_n(R_1, d, R_2)$ we have the following

THEOREM 2.1.

$$\lim_{n\to\infty} \rho_n(R_1,d,R_2) = (R_1^2 + d^2 + R_2^2)^{\frac{1}{2}}.$$

In order to prove this we make some preparatory observations.
It is well known that the volume $V_n(r)$ of a sphere with radius r in n-dimensional space is given by the formula

$$V_n(r) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}\, r^n.$$

If $x = (x_1, \ldots x_n)$ is a random point in the unit sphere in n-dimensional space then:

a. $$E(r^k) = \int_0^1 r^k \, d\,\frac{V_n(r)}{V_n(1)} = \int_0^1 r^k n r^{n-1}\, dr = \frac{n}{n+k}$$

for every $k > -1$, where $r := \| x \|_2 := (x_1^2 + \ldots x_n^2)^{\frac{1}{2}}$

b. $$E(x_i) = 0 \quad \text{for every } i \in \{1,2,\ldots,n\}$$

c. $$E(x_i^2) = \frac{1}{n} E(r^2) = \frac{1}{n+2} \quad \text{for every } i \in \{1,2,\ldots,n\}$$

d. $E(x_i x_j) = 0$ for $i \neq j$

e. $E(\| x \|_2 \quad x_i) = 0$ for every $i \in \{1,2,\ldots,n\}$.

If $y = (y_1,\ldots,y_n)$ is also a random point in the n-dimensional unit sphere and if x and y are independent then

f. $$E(x.y) := E(x_1y_1 + \ldots + x_ny_n) = \sum_{k=1}^{n} E(x_k)\, E(y_k) = 0$$

g. $$\mathrm{Var}(x.y) = E(x.y)^2 - E^2(x.y) = E(x.y)^2 = E(x_1y_1+\ldots+x_ny_n)^2 =$$

$$= E(x_1^2y_1^2+\ldots+x_n^2y_n^2+2\sum_{i<j} x_iy_ix_jy_j) = \sum_{k=1}^{n} E(x_k^2)\, E(y_k^2) +$$

$$+ 2\sum_{k=1}^{n} E(x_ix_j)\, E(y_iy_j) = \frac{n}{(n+2)^2}.$$

h. $E(\|x\|_2^2 \quad (x.y)) = 0$

i. $E(x_i \quad (x.y)) = 0$ for every $i \in \{1,2,\ldots,n\}$,

the last three properties being proved by means of arguments based on conditional probabilities.

PROOF OF THEOREM 2.1.

Without loss of generality we may assume that B_1 is the sphere with radius R_1 around the origin and that B_2 is the sphere with radius R_2 around the point $(d,0,0,\ldots,0)$. From now on a point in B_2 will be denoted by $y = (y_1+d, y_2, \ldots, y_n)$.

Let $S_n := \| x-y \|_2^2 = (x_1-y_1-d)^2 + \sum_{k=2}^{n} (x_k-y_k)^2 = \sum_{k=1}^{n} (x_k-y_k)^2 - 2d(x_1-y_1)+d^2 =$

$= \|x\|_2^2 + \|y\|_2^2 - 2(x.y) - 2d(x_1-y_1) + d^2$, so that by our introductory remarks

$$E(S_n) = \frac{nR_1^2}{n+2} + \frac{nR_2^2}{n+2} + d^2,$$

and hence

$$\lim_{n\to\infty} E(S_n) = R_1^2 + d^2 + R_2^2$$

so that our solution will be complete (as a consequence of Chebychef's inequality) if we can show that the variation of S_n tends to zero as $n \to \infty$.

Since

$$\mathrm{Var}\ (S_n) = \mathrm{Var}(S_n - d^2) = \mathrm{Var}\ (\|x\|_2^2 + \|y\|_2^2 - 2(x.y) -$$

$$- 2d(x_1 - y_1)) = E(\|x\|_2^4 + 4(x.y)^2 + 4d^2(x_1^2 - 2x_1y_1^2) +$$

$$+ 2\|x\|_2^2 \|y\|_2^2 + - 4\|x\|_2^2 (x.y) - 4d\|x\|_2^2 (x_1 - y_1) -$$

$$- 4\|y\|_2^2 (x.y) - 4\|y\|_2^2 d(x_1 - y_1) + 8d(x.y)(x_1 - y_1)) +$$

$$- E^2(\|x\|_2^2 + \|y\|_2^2 - 2(x.y) - 2d(x_1 - y_1)),$$

it follows from our introductory remarks that

$$\mathrm{Var}(S_n) = E(\|x\|_2^4 + \|y\|_2^4 + 4(x.y)^2 + 4d^2(x_1^2 + y_1^2) +$$

$$+ 2\|x\|_2^2 \|y\|_2^2) - E^2(\|x\|_2^2 + \|y\|_2^2) = \frac{n}{n+4} R_1^4 + \frac{n}{n+4} R_2^4 +$$

$$+ 4\frac{n}{(n+2)^2} R_1^2 R_2^2 + 4d^2 \frac{2}{n+2} + 2\frac{n}{n+2} R_1^2 \frac{n}{n+2} R_2^2) - (\frac{n}{n+2} R_1^2 + \frac{n}{n+2} R_2^2)^2$$

so that

$$\lim_{n\to\infty} \mathrm{Var}\ (S_n) = R_1^4 + R_2^4 + 2R_1^2R_2^2 - (R_1^2 + R_2^2)^2 = 0,$$

completing the proof.

THEOREM 2.2. *If* $x = (x_1,\ldots,x_n)$ *resp.* $y = (y_1,\ldots,y_n)$ *are two independent random points on the boundaries of the spheres* B_1 *resp.* B_2, *then the*

mathematical expectation of the distance between x *and* y *tends to* $(R_1^2 + d^2 + R_2^2)^{\frac{1}{2}}$ *as* $n \to \infty$.

PROOF. If $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ are two independent random points on the boundary of the n-dimensional unit sphere then, similarly as in the proof of theorem 2.1, we have:

a. $E(r^k) = E(1) = 1$, for any $k \in \mathbb{R}$, where $r := \|x\|_2$ $(= 1)$

b. $E(x_i) = 0$ for every $i \in \{1,2,\ldots,n\}$

c. $E(x_i^2) = \frac{1}{n} E(r^2) = \frac{1}{n}$ for every $i \in \{1,2,\ldots,n\}$

d. $E(x_i y_j) = 0$ if $i \neq j$

e. $E(\|x\|_2^2 \; x_i) = o$ for every $i \in \{1,2,\ldots,n\}$

f. $E(x.y) = 0$

g. $\mathrm{Var}(x.y) = \frac{1}{n}$

h. $E(\|x\|_2^2 \; (x.y)) = o$

i. $E(x_i(x.y)) =$ for every $i \in \{1,2,\ldots,n\}$

The remaining part of the proof is very similar to that of theorem 2.1.

We conclude this section by stating a related observation.
Let $\rho^*(n)$ be the mathematical expectation of the distance between the two independent random points x and y on the boundary of the n-dimensional unit sphere. Then, as one may verify,

$$\rho^*(1) = 1,\ \rho^*(2) = \frac{4}{\pi} \text{ and } \rho^*(3) = \frac{4}{3}$$

and Monte Carlo estimates of $\rho^*(n)$ for $n \geq 4$ indicate that the sequence $\{\rho^*(n)\}_{n=1}^{\infty}$ is increasing. However, we were unable to prove this.

3. GENERATING RANDOM SHOTS IN/ON N-DIMENSIONAL SPHERES

Virtually all modern electronic computers are equipped with a "random generator" returning (homogeneously distributed) values in the interval (0,1). Since successive values produced by this generator are stochastically independent it is clear that we may generate "random shots" *in* the n-dimensional unit cube $[0,1]^n$. By a simple transformation we obtain random shots in the cube $[-1,1]^n$.

The Monte Carlo method, in its most simple form, is based on the principle: "Shoot and tally".
However, applying this procedure to the n-dimensional unit sphere as a subset of the cube $[-1,1]^n$ one will experience that if n is quite large ($n \geq 15$, say) practically speaking none of the shots will hit the sphere. This observation is easily explained by observing that in high dimensional spaces the volume of the unit sphere is very small:

$$V_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cong \pi^{\frac{n}{2}} \left(\frac{n}{2}\right)^{-\frac{n}{2}} e^{\frac{n}{2}} \left(\frac{\pi n}{2}\right)^{-\frac{1}{2}}.$$

Below we will describe simple procedures which transform every random shot in the cube $[-1,1]^n$ into a weighted shot in/on the corresponding unit sphere. Let $x = (x_1,\ldots,x_n)$ be a random shot in the cube $[-1,1]^n$ and define $\tilde{x} := x / \|x\|_\infty$, where $\|x\|_\infty := \max_i |x_i|$, so that $\tilde{x}$ is the projection of x (from the origin) onto the boundary of the cube.
Defining $x^* := x/\|\tilde{x}\|_2$ (in the unit sphere) with weight $\|\tilde{x}\|_2^{-n}$, we may obtain a Monte Carlo estimate of $\rho(n)$ by means of the formula

$$\frac{\sum \|x^* - y^*\|_2 \, \|\tilde{x}\|_2^{-n} \, \|\tilde{y}\|_2^{-n}}{\sum \|\tilde{x}\|_2^{-n} \, \|\tilde{y}\|_2^{-n}}.$$

Defining $x^* := x/\|x\|_2$ (on the boundary of the unit sphere) the last formula may also serve to give a Monte Carlo estimate of $\rho^*(n)$.
All Monte Carlo estimates referred to in the previous sections were based

on the procedures just described.

ACKNOWLEDGEMENT. The author expresses his gratitude to L.G.L.T. Meertens and M. Voorhoeve for their useful comments on the subject.