THE DISTANCE BETWEEN TWO RANDOM POINTS IN A 4- AND 5-CUBE.

JOHAN PHILIP

Abstract. We determine exact expressions for the probability distribution and the average of the distance between two random points in a 4-cube and the average distance between two random points in a 5-cube. The obtained expressions contain unsolvable integrals. The two averages are known before. We use a new method resulting in new expressions.

1. Introduction

We consider the random variable V = the distance between two random points in a unit cube in dimensions four and five. V is one of the many random quantities studied in Geometric Probability, see eg. Santalo, [6]. The probability distributions for the distance between two random points in a square and in an ordinary 3-cube are given in Philip, [5]. That paper also gives the distribution of the distance between two random points in a rectangle and in a box with unequal side lengths.

The similar problem for the distance between two random points on the surface of a cube has been considered by Borwain *et.al.* [2] and Philip, [4]. See also Bailey *et.al.* [3].

A survey of the results in this area are collected by Weisstein on the web site [8]. Exact expressions for the average distance in four and five dimensions are given in Bailey *et.al.* [3], but they suffer from misprints. Correct versions of these expressions are given in [8]. Another expression for the four-dimensional average is given in A103983 of Sloane,[7].

The calculations of this paper rely on the results of [5]. The distribution for the 4-cube will be obtained by convolving the distribution for a square by itself. Convolving the distribution for the square with that of an ordinary 3-cube would give the distribution for a 5-cube. We will not calculate the distribution in 5-D but only the average. This is done by reversing the order of the convolution integration and the averaging integration.

1991 *Mathematics Subject Classification.* Primary: 60D05: Secondary: 52A22.
Key words and phrases. Distribution of distance in a 4- and 5-cube.

The numerical value of the average in n dimensions can be found by numerical evaluation of the integral

$$E(n) = \int_0^1 ... \int_0^1 \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}\; dx_1.....dy_n \tag{1}$$

The values $E(4) \approx .7776656535$ and $E(5) \approx .8785309152$ can be found in [8] or as A103983 and A103984 in [7]. We have checked these values to five digits accuracy by Monte Carlo tests.

Our results are long expressions and we rely heavily on the use of a formula manipulating program, in our case *Maple 10*.

2. Notation and formulation.

The four or five coordinates of the two random points are assumed to be evenly distributed on the unit interval and are assumed to be independent.

The calculations will be done for the square U of the distance V between the two random points. When the distribution function $H(u)$ for U is found, the wanted distribution $K(v)$ for the distance V is obtained as $K(v) = H(v^2)$. The corresponding densities satisfy $k(v) = 2\,v\,h(v^2)$.

3. The four-dimensional distribution.

We shall calculate the four-dimensional density $h(u)$ as the convolution of the density $g(u)$ by itself, where $g(u)$ is the density for the square of the distance in a (2-dimensional) square.

$$h(u) = \int g(u-s)\,g(s)\,ds, \tag{2}$$

The density $g(s)$ taken from [5] is as follows:

(3)
$$g(s) = \begin{cases} g_1(s) = \pi - 4\sqrt{s} + s, & 0 < s \le 1\ ; \\ g_2(s) = -2 + 4\sqrt{s-1} - s - 2\arcsin(1 - 2/s), & 1 < s \le 2. \end{cases}$$

Since g is supported by $(0,2)$, h will be supported by $(0,4)$. The analytical expressions for h will be different in the four unit intervals that form its support. We get

$$h_1(u) = \int_0^u g_1(u-s)\,g_1(s)\,ds, \qquad 0 < u \le 1\ , \tag{4}$$

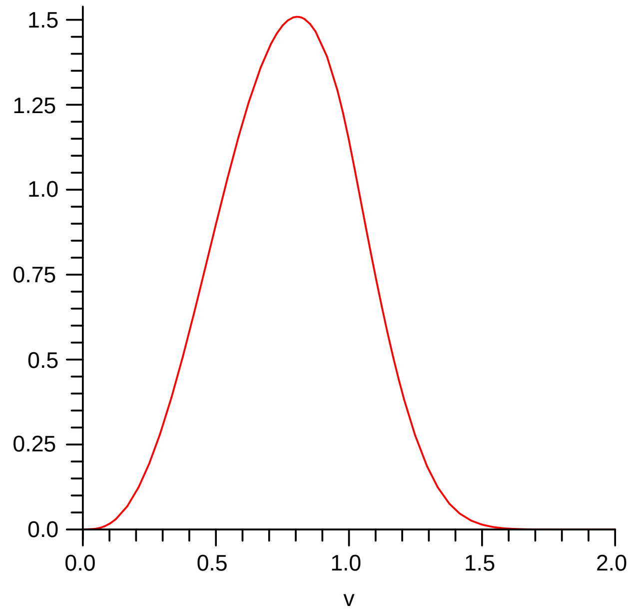

FIGURE 1. Density function $k(v)$ for the distance between two random points in a four-dimensional cube.

$$h_2(u) = \int_0^{u-1} g_2(u-s)\, g_1(s)\, ds + \int_{u-1}^{1} g_1(u-s)\, g_1(s)\, ds + \int_1^u g_1(u-s)\, g_2(s)\, ds = \int_{u-1}^{1} g_1(u-s)\, g_1(s)\, ds + 2\int_1^u g_1(u-s)\, g_2(s)\, ds, \qquad 1 < u \le 2\ , \tag{5}$$

$$h_3(u) = \int_{u-2}^{1} g_2(u-s)\, g_1(s)\, ds + \int_1^{u-1} g_2(u-s)\, g_2(s)\, ds + \int_{u-1}^{2} g_1(u-s)\, g_2(s)\, ds = \int_1^{u-1} g_2(u-s)\, g_2(s)\, ds + 2\int_{u-1}^{2} g_1(u-s)\, g_2(s)\, ds, \qquad 2 < u \le 3\ , \tag{6}$$

$$h_4(u) = \int_{u-2}^{2} g_2(u-s)\, g_2(s)\, ds, \qquad 3 < u \le 4. \tag{7}$$

The practical integrations in (4) to (7) are made with *Maple 10*. The *Maple* worksheet `h4D.mw` is available at `www.math.kth.se/~johanph`.

We give the result here as function of u, which is the square of distance between two random points in a four-dimensional cube.

(8)

$$h(u) = \begin{cases} \frac{1}{6}u^3 - \frac{32}{15}u^{5/2} + 3\pi u^2 - \frac{16}{3}\pi u^{3/2} + \pi^2 u, & 0 < u \le 1, \\ -\frac{1}{2}u^3 - (2+3\pi)u^2 + \frac{32}{3}u^{3/2} + (\frac{2}{3} - 12\pi + \pi^2)u \\ +\frac{4}{15}\left(24u^2 + (27+20\pi)u - 6 + 10\pi\right)\sqrt{u-1} \\ +2\pi - \frac{2}{15} - (12u^2 + 8\pi u)\operatorname{arcsec}(\sqrt{u}), & 1 < u \le 2, \\ \frac{1}{2}u^3 + (4-3\pi)u^2 + \frac{32}{3}\pi u^{3/2} + (\frac{14}{3} - \pi^2)u \\ -10\pi - \frac{26}{15} - \frac{32}{3}(u-1)^{3/2} - \frac{1}{5}(32u^2 + 72u + 8)\sqrt{u-2} \\ -32u^{3/2}\operatorname{arcsec}(u-1) + 8(3u^2 + 6u + 10)\operatorname{arcsec}(\sqrt{u-1}) \\ +16\int_1^{u-1} \operatorname{arcsec}(\sqrt{s})\, \operatorname{arcsec}(\sqrt{u-s})\, ds, & 2 < u \le 3, \\ -\frac{1}{6}u^3 + (-2+3\pi)u^2 + (-\frac{16}{3} + 12\pi - \pi^2)u - \frac{32}{15} - 12\pi \\ +\left(-\frac{32}{3}\pi + 32\operatorname{arcsec}(u-2)\right)\, (u-1)^{3/2} \\ +\frac{1}{15}(32u^2 + 108u + 48 - 120\pi)\sqrt{u-3} \\ +\left(-12u^2 + (-48+8\pi)\,u + 80 - 16\pi\right)\operatorname{arcsec}(\sqrt{u-2}) \\ +16\int_{u-2}^{2} \operatorname{arcsec}(\sqrt{s})\, \operatorname{arcsec}(\sqrt{u-s})\, ds, & 3 < u \le 4. \end{cases}$$

We see no way of solving the integral in this expression. The density $k(v)$ for the distance $v = \sqrt{u}$ is $k(v) = 2\,v\,h(v^2)$ and is shown in Figure 1.

4. The average distance between two random points in a 4-cube.

The exact value of the average is given by

$$E(4) = \int \sqrt{u}\, h(u)\, du = \int \sqrt{u} \int g(u-s)\, g(s)\, ds\, du. \tag{9}$$

One can evaluate the first integral in (9), but we choose to use the second double integral and start by calculating

(10)

$$m(s) = \int \sqrt{u}\, g(u-s)\, du = \int_0^1 \sqrt{s+t}\; g_1(t)\, dt + \int_1^2 \sqrt{s+t}\; g_2(t)\, dt.$$

We get

(11)
$$m(s) = \frac{4}{15}s^{5/2} - \frac{1}{15}(8s^2-9s-2)\sqrt{1+s} + \frac{1}{15}(4s^2-9s+1)\sqrt{2+s}$$
$$- \frac{4}{3}s^{3/2}\arcsin\left(\frac{1}{1+s}\right) - \frac{1}{2}s^2\log(s) + (s^2-2s-\frac{1}{3})\log(1+s)$$
$$+ s^2\log(1+\sqrt{1+s}) - (s^2-2s-\frac{1}{3})\log(1+\sqrt{2+s}).$$

The function $m(s)$ is well defined and increasing on the interval $0 \leq s < \infty$ starting from $m(0) = E(2) = \frac{1}{3}\log(1+\sqrt{2}) + \frac{1}{15}(2+\sqrt{2}) \approx .521405$.

Having $m(s)$, we get

$$E(4) = \int_0^1 m(s)\, g_1(s)\, ds + \int_1^2 m(s)\, g_2(s)\, ds. \tag{12}$$

The calculation of $E(4)$ is described in the *Maple* worksheet `E4D.mw`, which are available at `www.math.kth.se/~johanph`. The result is

(13)
$$E(4) = \frac{1}{5}\left[-\frac{23}{27} - \frac{16\pi}{63} + \left(\frac{73}{126} - \frac{34\pi}{21} + (\frac{68}{21}+\frac{\pi}{6})\arccos(\frac{1}{3})\right)\sqrt{2}\right.$$
$$+\frac{8}{21}\sqrt{3} - \frac{52}{21}\log(2) + \left(\frac{197}{84}+\frac{\pi}{6}\right)\log(3)$$
$$+\left(\frac{5}{14}-\frac{2\pi}{3}\right)\log(1+\sqrt{2}) + \frac{104}{21}\log(1+\sqrt{3})$$
$$\left.-\frac{2}{3}\int_1^{\sqrt{2}} \operatorname{arcsec}(1+t^2)\ \operatorname{arcsec}(t)\, dt\right].$$

We see no way of solving the integral of this expression. A numerical evaluation of $E(4)$ gives the value .77766... given in the introduction.

The second moment of the distance is easily calculated in any number of dimensions. The density function for the square of the distance in the unit interval is

$$f(u) = 1/\sqrt{u} - 1, \qquad 0 < u < 1. \tag{14}$$

The average of U is $\alpha_{2,1} = 1/6$. The squared distance in n dimensions is the sum of n such squared distances and we have $\alpha_{2,n} = n/6$.

5. The average distance between two random points in a 5-cube.

We have not tried to derive the five-dimensional density function but have used the same method as in the former section to calculate the average $E(5)$. The density $h_5(u)$ of the 5-cube is the convolution of the two-dimensional density $g(u)$ in (14) and the three-dimensional density $h_3(u)$ derived in [5]. We get

$$
\begin{aligned}
E(5) &= \int \sqrt{u}\, h_5(u)\, du = \int \sqrt{u} \int g(u-s)\, h_3(s)\, ds\, du \\
&= \int_0^3 m(s)\, h_3(s)\, ds,
\end{aligned} \tag{15}
$$

where $m(s)$ is given in (11).

Our calculation of the exact value of $E(5)$ is described in the *Maple* worksheets `E5D0.mw, E5D1.mw, and E5D2.mw` , which are available at `www.math.kth.se/~johanph`. The result is

(16)

$$
\begin{aligned}
E(5) = \frac{1}{7}\Bigg[&-\frac{449}{495} - \frac{4\pi}{27} + \frac{4\pi^2}{45} + \left(\frac{3239}{8910} - \frac{73}{9}\arcsin(\frac{1}{3})\right)\sqrt{2} \\
&+ \left(\frac{568}{495} - \frac{8}{3}\arcsin(\frac{1}{4})\right)\sqrt{3} - \frac{380}{891}\sqrt{5} \\
&- \left(\frac{124}{27} + \frac{223\pi}{90}\right)\log(2) + \left(\frac{295}{36} + \frac{\pi}{4}\right)\log(3) \\
&+ \left(\frac{7}{54} - \frac{\pi}{2}\right)\log(1+\sqrt{2}) + \left(\frac{40}{9} - \frac{47\pi}{15}\right)\log(1+\sqrt{3}) \\
&+ \left(\frac{64}{27} + \frac{52\pi}{15}\right)\log(1+\sqrt{5}) + \frac{52\pi}{45}\log(\sqrt{3}+\sqrt{5}) \\
&+ \int_1^2 \left(\frac{1}{(1+u)\sqrt{2+u}} + \frac{104}{15(2+u)\sqrt{3+u}}\right)\operatorname{arcsec}(\sqrt{u})\, du \\
&+ \frac{1}{4}\left(\int_{-3}^{-2} k_1(u)\, du - \int_1^2 k_1(u)\, du\right) \\
&+ \frac{26}{15}\left(\int_{-4}^{-3} k_2(u)\, du - \int_1^2 k_2(u)\, du\right)\Bigg],
\end{aligned}
$$

where

$$
k_1(u) = \frac{\operatorname{arcsec}(u)}{\sqrt{1+u}\sqrt{2+u}} \quad \text{and} \quad k_2(u) = \frac{\operatorname{arcsec}(u)}{\sqrt{1+u}\sqrt{3+u}}\ .
$$

6. Comparison of results.

The expression for $E(4)$ given by Bailey, Borwein and Crandall [3] has, after correction of misprints, the following form

(17)
$$
\begin{aligned}
E_{BBC}(4) = {} & \frac{136}{105}\sqrt{2}\arctan\left(\frac{1}{2}\sqrt{2}\right) - \frac{34}{105}\pi\sqrt{2} + \frac{8}{105}\sqrt{3} + \frac{73}{630}\sqrt{2} \\
& - \frac{4}{5}\mathrm{Cl}_2(\alpha) + \frac{4}{5}\mathrm{Cl}_2\left(\alpha - \frac{1}{2}\pi\right) + \frac{197}{420}\log(3) + \frac{1}{14}\log(1+\sqrt{2}) \\
& + \frac{4}{5}\alpha\log(1+\sqrt{2}) - \frac{1}{5}\pi\log(1+\sqrt{2}) + \frac{52}{105}\log(2+\sqrt{3}) \\
& - \frac{23}{135} - \frac{16}{315}\pi + \frac{26}{15}\,\mathrm{G}\ ,
\end{aligned}
$$

where G is Catalan's number, $\alpha = \arcsin\left(\frac{2}{3} - \frac{\sqrt{2}}{6}\right)$ and $\mathrm{Cl}_2(z)$ is the Clausen function from [1] :

$$\mathrm{Cl}_2(z) = -\int_0^z \log\left(|2\sin(t/2)|\right) dt. \tag{18}$$

Noting that $\alpha = \frac{\pi}{4} - \arcsin\left(\frac{1}{3}\right)$, one can write

$$\mathrm{Cl}_2(\alpha) - \mathrm{Cl}_2\left(\alpha - \frac{1}{2}\pi\right) = -\int_{\arcsin\left(\frac{1}{3}\right)-\frac{\pi}{4}}^{\arcsin\left(\frac{1}{3}\right)+\frac{\pi}{4}} \log\left(|2\sin(t/2)|\right) dt. \tag{19}$$

Equating the expressions for $E(4)$ given in (13) and (17), we get the following relation between the insolvable integrals of the two expressions. For reasons that will become apparent, we name this expression A. Notice that $E_{BBC}(4)$ contains $\frac{4}{5}A$.

(20)
$$
\begin{aligned}
A &= \int_{\arcsin\left(\frac{1}{3}\right)-\frac{\pi}{4}}^{\arcsin\left(\frac{1}{3}\right)+\frac{\pi}{4}} \log\left(|2\sin(t/2)|\right) dt + \frac{13}{6}G + \left(\alpha - \frac{\pi}{4}\right)\log(1+\sqrt{2}) \\
&= \frac{\pi}{24}\left(\log(3) - 4\log(1+\sqrt{2}) + \sqrt{2}\arccos\left(\frac{1}{3}\right)\right) \\
&- \frac{1}{6}\int_1^{\sqrt{2}} \mathrm{arcsec}(1+t^2)\ \mathrm{arcsec}(t)\, dt.
\end{aligned}
$$

The integrand of the first integral in (20) has a singularity at $t = 0$ while the second integrand is bounded and continuous . We see no way of deducing this relation directly.

The expression for $E(5)$ given by Bailey, Borwein and Crandall [3] has, after correction of misprints, the following form

(21)

$$E_{BBC}(5) = \frac{65}{42}G - \frac{449}{3465} - \frac{4}{189}\pi + \frac{4}{315}\pi^2 + \frac{3239}{62370}\sqrt{2} + \frac{568}{3465}\sqrt{3}$$
$$- \frac{380}{6237}\sqrt{5} - \frac{73}{63}\sqrt{2}\arctan\left(\frac{\sqrt{2}}{4}\right) - \frac{8}{21}\sqrt{3}\arctan\left(\frac{1}{\sqrt{15}}\right)$$
$$+ \left(\frac{52}{63}\pi - \frac{184}{189}\right)\log(2) + \frac{295}{252}\log(3) + \frac{64}{189}\log(1+\sqrt{5})$$
$$+ \left(\frac{1}{54} - \frac{5\pi}{28} + \frac{5}{7}\alpha\right)\log(1+\sqrt{2}) + \frac{4}{63}(10-13\pi)\log(\sqrt{2}+\sqrt{6})$$
$$- \frac{5}{7}\mathrm{Cl}_2(\alpha) + \frac{5}{7}\mathrm{Cl}_2(\alpha - \pi/2) + \frac{52}{63}K_3,$$

where G, α and $\mathrm{Cl}_2(z)$ are the same as above and

$$K_3 = \int_3^4 \frac{\mathrm{arcsec}(x)}{\sqrt{x^2 - 4x + 3}}\,dx\ .$$

Notice that $E_{BBC}(5)$ contains $\frac{5}{7}A$. This implies that we can use (20) to replace G, α and Cl_2 in $E_{BBC}(5)$ by the arcsec-integral of $E(4)$. This results in our favourite expression for $E(5)$:

(22)

$$E_F(5) = \frac{1}{21}\Bigg[-\frac{449}{165} - \frac{4\pi}{9} + \frac{4\pi^2}{15} + \left(\frac{568}{165} - 8\arcsin\left(\frac{1}{4}\right)\right)\sqrt{3}$$
$$+ \left(\frac{3239}{2970} + \frac{5}{16}\pi^2 - \left(\frac{73}{3} + \frac{5\pi}{8}\right)\arcsin\left(\frac{1}{3}\right)\right)\sqrt{2} - \frac{380}{297}\sqrt{5}$$
$$+ \left(-\frac{124}{9} + \frac{26\pi}{3}\right)\log(2) + \left(\frac{295}{12} + \frac{5\pi}{8}\right)\log(3)$$
$$+ \left(\frac{7}{18} - \frac{5\pi}{2}\right)\log(1+\sqrt{2}) + \left(\frac{40}{3} - \frac{52\pi}{3}\right)\log(1+\sqrt{3})$$
$$+ \frac{64}{9}\log(1+\sqrt{5})$$
$$+ \frac{52}{3}\int_3^4 \frac{\mathrm{arcsec}(t)}{\sqrt{t^2-4t+3}}\,dt - \frac{5}{2}\int_1^{\sqrt{2}} \mathrm{arcsec}(1+t^2)\mathrm{arcsec}(t)\,dt\Bigg].$$

7. Comment.

Even if *Maple* is very helpful in doing the calculations of this paper, there are several things it doesn't do. The success of the calculations relies on manual simplification of trigonometric expressions, on manual factorization of polynomials and on choosing suitable parts in partial integrations.

References

[1] Abramowitz and Stegun, *Handbook of Mathematical Functions*, Dover Publications, N.Y. 1964.

[2] D.H. Bailey, J.M. Borwein, V Kapoor, and E.W. Weisstein, *Ten Problems in Experimental Mathematics,* http://crd.lbl.gov/~dhbailey/dhbpapers/tenproblems.pdf, March 8, 2006.

[3] D.H. Bailey, J.M. Borwein, R.E. Crandall, *Box integrals,* Journal of Computational and Applied Mathematics Vol. 206, Issue 1, Sept. 2007, pp. 196-208.

[4] J. Philip, *Calculation of Expected Distance on a Unit Cube.*, Jan. 2007, `www.math.kth.se/~johanph/surfacedist.pdf`.

[5] J. Philip, *The Probability Distribution of the Distance between two Random Points in a Box.*, TRITA MAT 07 MA 10, ISSN 1401-2278, Dec. 2007, `www.math.kth.se/~johanph/habc.pdf`.

[6] L.A. Santalo, *Integral Geometry and Geometric Probability*, Encyclopedia of Mathematics and Its Applications, Addison-Wesley, 1976

[7] N.J.A. Sloane, *The Online Encyclopedia of Integer Sequences*, `www.research.att.com/~njas/sequences/`.

[8] E. Weisstein at `www.mathworld.wolfram.com`. Look at Geometry > Computational Geometry > Random Point Picking > .. or at Geometry > Solid Geometry > Polyhedra > .. .

Department of Mathematics, Royal Institute of Technology, S-10044 Stockholm Sweden

E-mail address: `johanph@kth.se`