Expected Distance of a Random Interior Point in an N-Dimensional Hypercube to its Nearest Facet

Aiden Ament

July 2023

1 Introduction

I first was exposed to this problem on my 170A final. The question is: Select a point at random from within the unit square. Let the random variable X denote the distance from this point to the nearest edge. What is the expected value of X. This question was the hardest question on the final and stumped me at the time, but I eventually got around to solving not only the 2 and 3 dimensional cases, but also the more general n dimensional case.

2 2 Dimensions

In the 2D case, we can divide the unit square into 4 symmetric sections which contain points closest to their respective edge.

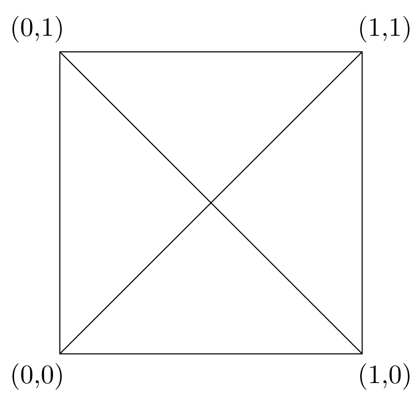

Figure 1: Regions divide the square into sections with the same closest edge.

By definition,

$$\mathbb{E}[X] = \int_0^1 \int_0^1 X \cdot P(X)\, dy\, dx$$

Here our points are sampled uniformly, so $P(X) = 1$. Furthermore, the value of X is: $x,\ 1-x,\ y,\ 1-y$ depending on the region the random point is within. Due to the symmetry of the square, the value of the integral over each region remains constant which yields

$$\mathbb{E}[X] = 4 \cdot \int_R X \ dR$$

Then, picking the bottom region arbitrarily, we can further split it into two distinct pieces: when $x < 0.5$ and when $x > 0.5$. Again due to symmetry, the integral over both of these pieces will be identical. However, now each region is easily parameterized.

$$\int_R X \ dR = 2 \cdot \int_{x=0}^{x=0.5} \int_{y=0}^{y=x} y \ dy \ dx$$

$$\int_R X \ dR = 2 \cdot \frac{1}{6} x^3 |_0^{\frac{1}{2}} = 2 \cdot \frac{1}{6} \cdot \frac{1}{2^3}$$

Putting all of this together we get that

$$\mathbb{E}[X] = 4 \cdot \int_R X \ dR = 4 \cdot 2 \cdot \frac{1}{6} \cdot \frac{1}{2^3} = \frac{1}{6}$$

So the expected distance of a random point selected inside a unit square to its closest edge is $\frac{1}{6}$

3 3 Dimensions

Now we analyze a similar question. Select a point within the unit cube at random. Let the random variable X represent the distance to the closest face. What is the expected value of X.

Again, we can start by dividing the cube into regions which contain points that have the same closest face. As there are 6 faces of a cube, we will split the cube into 6 symmetric regions.

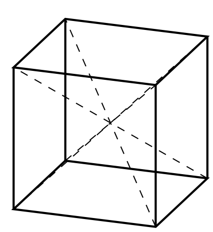

Figure 2: Regions divide the cube into sections with the same closest face.

Similarly to the 2D case, we find that

$$\mathbb{E}[X] = 6 \cdot \int_R X \ dR$$

To calculate $\int_R X \ dR$, similarly to the two dimensional case, we further subdivide R into 8 symmetrical pieces as shown below.

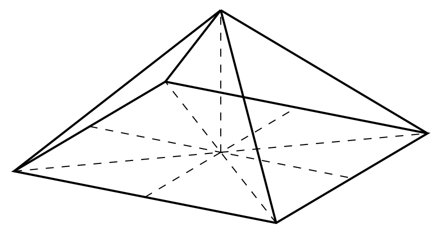

Figure 3: Bottom region of cube after subdivision.

Again, due to the symmetry of the cube, the integral over each one of these pieces is identical, so we get that

$$\int_R X \ dR = 8 \cdot \int_{x=0}^{x=0.5} \int_{y=0}^{y=x} \int_{z=0}^{z=y} z \ dz \ dy \ dx$$

$$\int_R X \ dR = 8 \cdot \frac{1}{4!} x^4 |_0^{\frac{1}{2}} = 8 \cdot \frac{1}{24} \cdot \frac{1}{2^4}$$

Finally, putting all of this together, we get that

$$\mathbb{E}[X] = 6 \cdot \int_R X \ dR = 6 \cdot 8 \cdot \frac{1}{24} \cdot \frac{1}{2^4} = \frac{1}{8}$$

So the expected distance of a random point selected inside a unit cube to its closest face is $\frac{1}{8}$

4 n Dimensions

We now formalize the question for n dimensions. Select randomly select a point within the unit n-cube. Let X represent it's distance to the nearest (n-1) face. What is the expected value of X.
Similarly to the previous cases we start by splitting the n-cube into regions based on the closest (n-1) face. In the one dimensional case, there are 2 points. In the two dimensional case, there are 4 lines. In the three dimensional case there are 6 faces. More generally, an n-cube contains 2n (n-1) faces. Consequently, we can write

$$\mathbb{E}[X] = 2n \cdot \int_R X \ dR$$

We now try and decompose the region further. Lets define

$$f(x_1, x_2, \ldots, x_{n-1}) = \min \{x_1,\ 1 - x_1,\ x_2,\ 1 - x_2, \ldots,\ x_{n-1},\ 1 - x_{n-1}\}$$

We can interpret this function as taking a point on an (n-1) face and outputting the maximum distance away which still has the given (n-1) face as it's closest boundary point. We can use this function to define R.

$$\int_R X\ dR = \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1} \cdots \int_{x_{n-1}=0}^{x_{n-1}=1} \int_{x_n=0}^{x_n=f(x_1,\ldots,x_{n-1})} x_n\ dR$$

Here, the first n-1 integrals define an (n-1) face for R, and then the last integral defines how far away from the (n-1) face any given point can be. Here we have that $X = x_n$ as that is the minimal distance to the closest (n-1) face by construction.

Similarly with the lower dimensional cases, we can split this integral up into many symmetric easily integrable regions. In particular, we can notice that f is symmetric with respect to it's arguments. This means we can split the integral into $(n-1)!$ pieces corresponding to different permutations of arguments, and the value over each piece will remain constant. We can also notice that within each of these pieces, there are two 2^{n-1} possible functions from which the min can be reporting from - two for each argument. Ultimately this means we can divide the original region into $2^{n-1}(n-1)!$ pieces, and the integral over each piece will remain constant. While this sounds confusing, looking at the resulting integral should clear things up.

$$\int_{x_1=0}^{x_1=0.5} \int_{x_2=0}^{x_2=x_1} \cdots \int_{x_{n-1}=0}^{x_{n-1}=x_{n-2}} \int_{x_n=0}^{x_n=x_{n-1}} 1\ dR$$

This corresponds specifically to the region where $x_1 < 0.5$ and $x_1 \geq x_2, \geq \ldots \geq x_{n-1}$ and consequently $f(x_1, \ldots, x_{n-1}) = x_{n-1}$. The integral evaluates to $\frac{1}{n!} \cdot \frac{1}{2^n}$, and as we know that the hyper volume of R is $\frac{1}{2n}$, it is easy to deduce that this subdomain is $\frac{1}{2^{n-1} \cdot (n-1)!}$ of the region being evaluated as we'd expect.

The whole purpose for splitting the domain into these pieces is because the symmetry of the problem means that the integral over each piece is constant. So we have

$$\int_R X\ dR = 2^{n-1}(n-1)! \cdot \int_{x_1=0}^{x_1=0.5} \int_{x_2=0}^{x_2=x_1} \cdots \int_{x_{n-1}=0}^{x_{n-1}=x_{n-2}} \int_{x_n=0}^{x_n=x_{n-1}} x_n\ dR$$

$$\int_R X\ dR = 2^{n-1} \cdot (n-1)! \cdot \frac{1}{(n+1)!} x^{n+1}|_0^{\frac{1}{2}} = 2^{n-1} \cdot (n-1)! \cdot \frac{1}{(n+1)!} \cdot \frac{1}{2^{n+1}}$$

$$\int_R X\ dR = \frac{1}{(n+1) \cdot n} \cdot \frac{1}{2^2}$$

Putting this altogether

$$\mathbb{E}[X] = 2n \cdot \int_R X \; dR = 2n \cdot \frac{1}{(n+1) \cdot n} \cdot \frac{1}{2^2} = \frac{1}{2(n+1)}$$

So the expected value of the distance to the boundary of a randomly selected point in an n-cube is $\frac{1}{2(n+1)}$

5 Significance

While the problem itself is neat, it says something rather interesting about how space scales in higher dimensions. In particular, the proportion of space close to the boundary box increases with the dimension. This is because $\mathbb{E}[X^{(n)}] < \mathbb{E}[X^{(n-1)}]$. This states that the expected distance of a random point to the boundary strictly decreases as dimension size increases, which can only happen if there is a higher proportion of space closer to the edges. While this alone is somewhat interesting, it can likely be deduced much more simply by just observing $(1-\epsilon)^n \to 0$ as $n \to \infty$. The expected value tells us something slightly more interesting however. The rate that volume consolidates near the edges of the box seems to be linear. If you want to half the expected distance to the boundary, you need to roughly double the dimension of your box. This probabilistic interpretation is not something I've seen before and I think it provides a new lens to understand how volume scales in higher dimensions.