Minimal Arc Problem

By: Osman Akar

August 12, 2020

1 Minimal Arc Problem and Solution

Problem There is a circle with unit circumference. n points are chosen on the circle at uniformly random. Find the expected length of the minimum arc bounded by 2 chosen points.

Simulation It is not hard to simulate problem to find the answer (thanks to *Strong Law of Large Numbers*). The answer is $\frac{1}{n^2}$.

Solution I will modify the problem as follows: $n-1$ points are selected on a unit segment. Say $X_1, X_2, \ldots, X_{n-1}$. Since $P(X_i = X_j) = 0$ for $i \neq j$, we can assume $\{X_i\}_{i=1}^{n-1}$ are pairwise different. Let $0 \leq \tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_{n-1} \leq 1$ be the ordered sequence of $X_1, X_2, \ldots, X_{n-1}$. Define $Z = min\{\tilde{X}_1, \tilde{X}_2 - \tilde{X}_1, \tilde{X}_3 - \tilde{X}_2, \ldots, \tilde{X}_{n-1} - \tilde{X}_{n-2}, 1 - \tilde{X}_{n-1}\}$. Then the answer of the problem is $\mathbb{E}Z$. In other words, we can assume that the first uniformly chosen point is fixed to 0. We can do this assumption because the answer of the problem is invariant under rotations[1].

The idea of the solution is to compute the mass density of Z directly. To compute the density, we will compute cumulative mass density $\mathbb{P}(Z \leq z)$. First, define random vectors

$$\mathbb{X} = (X_1, X_2, \ldots, X_{n-1}), \quad \tilde{\mathbb{X}} = (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{n-1})$$

Since $X_1, X_2, \ldots, X_{n-1}$ are independent and uniformly distributed, the joint density of $\mathbb{X}$

$$f_{\mathbb{X}}(x_1, x_2, \ldots, x_{n-1}) = 1$$

in the unit cube $0 \leq x_1, x_2, \ldots, x_{n-1} \leq 1$. $\tilde{\mathbb{X}} = (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{n-1})$ is the ordered sequence of $(X_1, X_2, \ldots, X_{n-1})$, so the joint density of $\tilde{\mathbb{X}}$ is

$$f_{\tilde{\mathbb{X}}}(x_1, x_2, \ldots, x_{n-1}) = (n-1)!$$

on the set $0 \leq x_1 < x_2 < \cdots < x_{n-1} \leq 1$. We multiplied $f_{\mathbb{X}}(\cdot)$ with $(n-1)!$ to get $f_{\tilde{\mathbb{X}}}(\cdot)$ because all permutations of $(X_1, X_2, \ldots, X_{n-1})$ gives the same ordered sequence $(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{n-1})$ [2].

[1] Actually this is how we simulated the problem

[2] For more details you can read the following MIT Document on discrete probability at p.90-91

Define $Y_1 = X_1$, $Y_i = \tilde{X}_i - \tilde{X}_{i-1}$ for $i = 2, 3, \ldots, n-1$, which represents the length of intervals between successive points (i.e. spacings). This can be thought as a linear transformation between $(n-1)-$dimensional spaces

$$\mathbb{T} : (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{n-1}) \to (\tilde{X}_1, \tilde{X}_2 - \tilde{X}_1, \tilde{X}_3 - \tilde{X}_2, \ldots, \tilde{X}_{n-1} - \tilde{X}_{n-2})$$

or

$$\mathbb{T}(\tilde{\mathbb{X}}) = \mathbb{Y}$$

where $\mathbb{Y} = (Y_1, Y_2, \ldots .Y_{n-1})$. The determinant of the Jacobian of $\mathbb{T}$ is 1. Applying Change of Variables, the density of $\mathbb{Y}$ is

$$\begin{aligned} f_{\mathbb{Y}}(y_1, y_2, \ldots, y_{n-1}) &= f_{\mathbb{Y}}(\mathbb{T}(\tilde{\mathbb{X}})) \\ &= f_{\tilde{\mathbb{X}}}(x_1, x_2, \ldots, x_{n-1}) \cdot |\frac{\partial \mathbb{Y}}{\partial \tilde{\mathbb{X}}}| \\ &= (n-1)! \end{aligned}$$

on the set $\{(y_1, y_2, \ldots, y_{n-1}) \mid 0 \le y_1, y_2, \ldots, y_{n-1}$ and $y_1 + y_2 + \ldots, y_{n-1} \le 1\} \subset \mathbb{R}^{n-1}$.

Let's come back to the problem. We want to compute the density of $Z = min\{\tilde{X}_1, \tilde{X}_2 - \tilde{X}_1, \tilde{X}_3 - \tilde{X}_2, \ldots, \tilde{X}_{n-1} - \tilde{X}_{n-2}, 1 - \tilde{X}_{n-1}\} = min\{Y_1, Y_2, \ldots, Y_{n-1}, 1 - (Y_1 + Y_2 + \cdots + Y_{n-1})\}$

$$\begin{aligned} \mathbb{P}(Z \le z) &= 1 - \mathbb{P}(Z > z) \\ &= 1 - \mathbb{P}(Y_1 > z, Y_2 > z, \ldots, Y_{n-1} > z, Y_1 + Y_2 + \cdots + Y_{n-1} < 1 - z) \end{aligned}$$

Note that above computations showed that $\mathbb{Y} = (Y_1, Y_2, \ldots, Y_{n-1})$ distributed uniformly on the $(n-1)$-dimensional pyramid $\{(y_1, y_2, \ldots, y_{n-1}) \mid 0 \le y_1, y_2, \ldots, y_{n-1}$ and $y_1 + y_2 + \ldots, y_{n-1} \le 1\}$. To compute $\mathbb{P}(Y_1 > z, Y_2 > z, \ldots, Y_{n-1} > z, Y_1 + Y_2 + \cdots + Y_{n-1} < 1 - z)$ we only need to compute the volume of the section $V_z = \{(y_1, y_2, \ldots, y_{n-1}) \mid y_1 > z, y_2 > z, \ldots, y_{n-1} > z, y_1 + y_2 + \cdots + y_{n-1} < 1 - z\}$. Defining $w_i = y_i - z,\ i = 1, 2, \ldots, n-1$, the volume of V_z is equal to volume of the pyramid $\{(w_1, w_2, \ldots, w_{n-1}) \mid w_1 > 0, w_2 > 0, \ldots, w_{n-1} > 0, w_1 + w_2 + \cdots + w_{n-1} < 1 - nz\}$ which is $\frac{1}{(n-1)!}(1 - nz)^{n-1}$. Thus the cumulative distribution of Z is

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \le z) \\ &= 1 - \mathbb{P}(Z > z) \\ &= 1 - \mathbb{P}(Y_1 > z, Y_2 > z, \ldots, Y_{n-1} > z, Y_1 + Y_2 + \cdots + Y_{n-1} < 1 - z) \\ &= 1 - \int_{V_z} f_{\mathbb{Y}}(y_1, y_2, \ldots, y_{n-1}) d(y_1, y_2, \ldots, y_{n-1}) \\ &= 1 - \int_{V_z} (n-1)! d(y_1, y_2, \ldots, y_{n-1}) \\ &= 1 - (n-1)! Vol(V_z) \\ &= 1 - (1 - nz)^{n-1} \end{aligned}$$

Then the density of Z is

$$f_Z(z) = \frac{\partial}{\partial z} F_Z(z) = n(n-1)(1-nz)^{n-2}$$

Finally we can finally compute

$$\begin{aligned}
EZ &= \int_0^{1/n} z f_Z(z) dz \\
&= \int_0^{1/n} zn(n-1)(1-nz)^{n-2} dz \\
&= (n-1) \int_0^{1/n} (1-nz)^{n-2} - (1-nz)^{n-1} dz \\
&= (n-1) \left[-\frac{(1-nz)^{n-1}}{n(n-1)} + \frac{(1-nz)^n}{n^2} \right]_{z=0}^{1/n} \\
&= (n-1) \left[\frac{1}{n(n-1)} - \frac{1}{n^2} \right] \\
&= \frac{1}{n^2} \qquad \square
\end{aligned}$$