# CS648 : Project 7

***Team : BiasedCoin***

**Divij Singla (210350)**

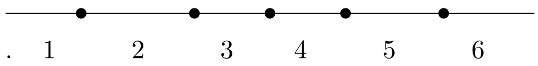
## Problem

Suppose we pick n points randomly uniformly and independently from the $[0, 1]$ line segment. This will create $n + 1$ intervals. What is the expected length of the smallest interval among these intervals?

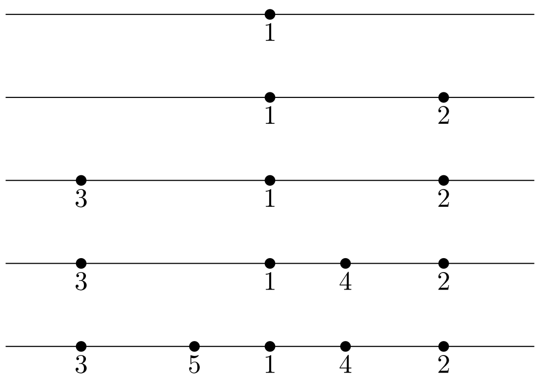
## Definitions

### Indexing

- Starting from origin, intervals are marked 1 to $n + 1$

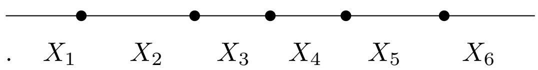



- Points are given label in order of sampling (which is Uniform Random Sampling)




### Random Variables

$X_i$ : Length of $i^{th}$ interval




$Y_i$ : Distance of $i^{th}$ point from the origin

$Z_i$ : Ordered Statistics of $(Y_1, Y_2, ..., Y_n)$

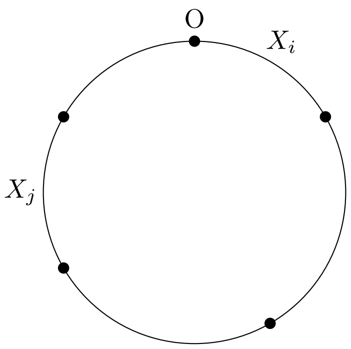
$E_i$ : Independent Identically Distributed (IID) Exponential Random Variables

## Notations

- $P_{1,2,3,..k}(E) = P(\text{E} \mid \text{point 1, 2,..., k appear in order as we move left to right from origin})$
- [min($X_i$)≥ c ∀ i ∈(1,n+1)] is denoted as min(X)
- $F_A(x)$ denotes CDF (Cumulative Distribution Function) of random variable A
- $f_A(x)$ denotes PDF (Probability Density Function) of random variable A
- $\bar{X} : (X_1, X_2, ..., X_{n+1})$, $\bar{Y} : (Y_1, Y_2, ..., Y_n)$, $\bar{Z} : (Z_1, Z_2, ..., Z_n)$, $\bar{E} : (E_0, E_1, ..., E_n)$
- $X \sim Exp(\lambda)$ : X is a random variable having exponential distribution with parameter $\lambda$

# Observation

1. $X_i$ and $X_j$



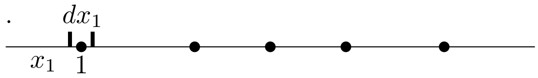

Assume a circle with $n + 1$ points randomly uniformly distributed
By symmetry, $X_i$ and $X_j$ would be distributed symmetrically, i.e. $f_{X_i} = f_{X_j}$ (where f is PDF (Probability Density Function))
Now we can choose any point (say O) and break the circle at point O to a straight line and remove the redundant point in start (or end)



. $X_i$ $X_j$
Doing the operations on circle does not change the distribution of length of intervals and hence, we can conclude: $f_{X_i} = f_{X_j}$ for $[0, 1]$ line.

# Solution

Assume point 1 lies closest to origin at distance $x_1$ in a small $dx_1$ locality,




To assure point 1 is closest to origin, rest of the points should lie at a distance $x \geq x_1$, this happens with Probability $= (1 - x_1)^{n-1}$

Hence,

$$P(X_1 \geq c \mid \text{point 1 is closest to origin}) = \int_c^1 (dx_1)(1 - x_1)^{n-1}$$

$$P_1(X_1 \geq c) = \int_c^1 (dx_1)(1 - x_1)^{n-1} \text{ refer notations defined above}$$

Similarly,

$$P_{1,2...,k}(X_1 \geq c, X_2 \geq c, ..., X_k \geq c) = \int_c^1 \int_c^1 ... \int_c^1 (dx_1)(dx_2)...(dx_k)(1 - \sum_{i=1}^{k} x_i)^{n-k}$$

$$= \frac{(1 - kc)^n}{n(n-1)...(n-k+1)}$$

As points are randomly uniformly distributed, all permutations of k points are equally likely to be the first k points (in order) w.r.t the distance from origin, i.e.

$$P(X_1 \geq c, X_2 \geq c, ..., X_k \geq c) = k!\binom{n}{k} P_{1,2...,k}(X_1 \geq c, X_2 \geq c, ..., X_k \geq c) = (1 - kc)^n$$

for $k = n + 1$ (All points)

$$P(X_1 \geq c, X_2 \geq c, ..., X_{n+1} \geq c) = (1 - (n+1)c)^n$$

$$P(min(X_i) \geq \text{c} \ \forall \text{ i } \in (1, n+1)) = (1 - (n+1)c)^n$$

$$P(min(X) \geq \text{c}) = (1 - (n+1)c)^n$$

$$P(min(X) \leq \text{c}) = 1 - (1 - (n+1)c)^n = F_{min(X)}(c)$$

$$\frac{d}{dx} F_{min(X)}(x) = f_{min(X)}(x)$$

$$f_{min(X)}(x) = n(n+1)(1 - (n+1)x)^{n-1}$$

$$E(min(X)) = \int_0^{\frac{1}{n+1}} x f_{min(X)}(x) dx$$

$$E(min(X)) = n(n+1) \int_0^{\frac{1}{n+1}} x(1 - (n+1)x)^{n-1} dx$$

$$E(min(X)) = n(n+1)\frac{1}{n(n+1)^3}$$

$$\boxed{E(min(X)) = \frac{1}{(n+1)^2}}$$

$\implies$ Expected length of the smallest interval is $\dfrac{1}{(n+1)^2}$

Notable Results for $k = 1$

$$P(X_1 \geq c) = (1-c)^n$$

$$P(X_1 \leq c) = F_{X_1}(c) = (1-c)^n$$

$$f_{X_1}(x) = n(1-x)^{n-1} \tag{1}$$

We already know,

$$f_{X_1}(x) = f_{X_2}(x) = ... = f_{X_{n+1}}(x)$$

## Empirical Analysis

The following experiment was simulated on computer using python

1. $n$ points were generated randomly uniformly and independently in $[0, 1]$
2. Smallest interval length was noted
3. For same n, experiment was carried out multiple times, and average of smallest interval length was computed
4. Same procedure repeated for all $n \in [1000, 10000]$

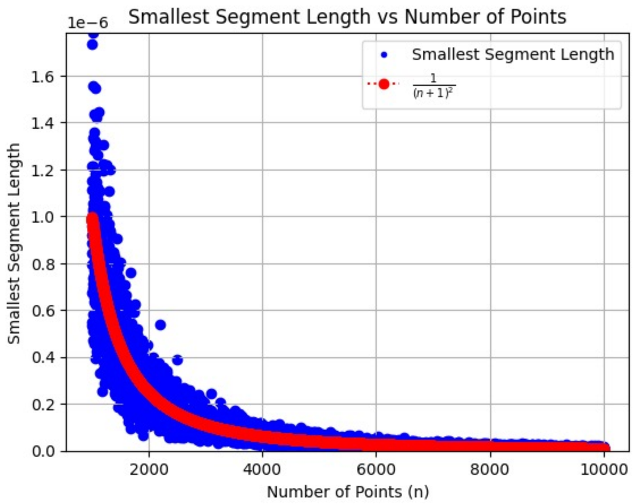



Figure 1: Simulated expt 10 times per $n$ and took average

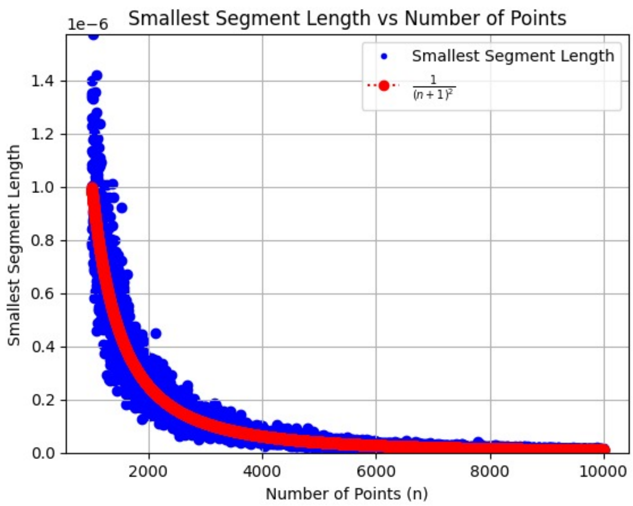



Figure 2: Simulated expt 20 times per $n$ and took average

## Curve Fit with $\dfrac{1}{(n+1)^k}$

Simulated data achieved a curve-fit with $\frac{1}{(n+1)^k}$ for $k = 1.998850203988063$

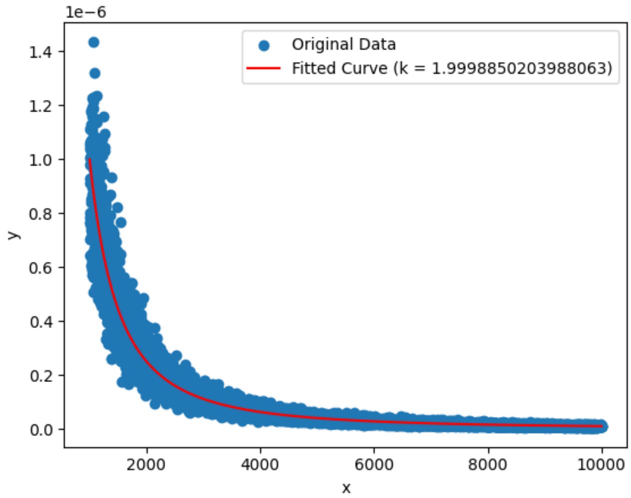



Figure 3: Curve Fit with 20 times per n data

# Joint Density Function

As the points are distributed randomly, uniformly, and independently, $Y_1, Y_2, ..., Y_n$ are uniform and independent random variables.

We sort them in ascending order and call the corresponding random variables (in order): $Z_1, Z_2, ..., Z_n$.

$X_i = Z_i - Z_{i-1}$ where $Z_0 = 0$ and $Z_{n+1} = 1$.

$$f_{\bar{Y}}(y_1, y_2, ..., y_n) = 1 \text{ where } 0 \leq y_i \leq 1$$

$$f_{\bar{Z}}(z_1, z_2, ..., z_n) = n! \text{ where } 0 \leq z_0 \leq z_1 \text{ ... } \leq z_n \leq 1$$

$$f_{\bar{Z}}(Z_1 = z_1, Z_2 = z_2, ..., Z_n = z_n) = n! \text{ where } 0 \leq z_0 \leq z_1 \text{ ... } \leq z_n \leq 1$$

$$f_{\bar{X}}(X_1 = x_1, X_2 = x_2, ..., X_{n+1} = x_{n+1}) = f_{\bar{X}}(Z_1 - 0 = x_1, Z_2 - Z_1 = x_2, ..., Z_{n+1} - Z_n = x_{n+1})$$

$$= f_{\bar{Z}}(Z_1 = x_1, Z_2 = x_1 + x_2, Z_3 = x_1 + x_2 + x_3 ..., 1 = x_1 + x_2 + ...x_{n+1})$$

$$= f_{\bar{Z}}(Z_1 = x_1, Z_2 = x_1 + x_2, Z_3 = x_1 + x_2 + x_3 ..., Z_n = x_1 + x_2 ... + x_n) \text{ where } x_1 + x_2 + ... + x_{n+1} = 1$$

$$= n! \text{ where } x_1 + x_2 + ... + x_{n+1} = 1$$

$$\implies f_{\bar{X}}(x_1, x_2, ..., x_{n+1}) = n! \text{ where } x_1 + x_2 + ... + x_{n+1} = 1 \tag{2}$$

## Exponential Random Variables

Consider Independent Identically Distributed $E_0, E_1, ...E_n \sim Exp(1)$

$$f_{\bar{E}}(e_0, e_1, ..., e_n) = e^{-(e_0+e_1+...+e_n)} = e^{-\sum e_i}$$

$$f_{\bar{E},S}(e_0, e_1, ..., e_n, s) = e^{-\sum e_i} = e^{-s} \text{ where } S = \sum_{i=0}^{n} E_i$$

Some well known results:
1. If $E_0, E_1, ...E_n \sim Exp(1)$ are IIDs, then

$$f_{\sum_{i=0}^{n} E_i}(s) = \frac{e^{-s} s^n}{n!}$$

2. If X and Y are jointly continuous, the conditional PDF of X given Y is given by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Using these results,

$$f_{\bar{E}|S}(e_0, e_1, ..., e_n|s) = \frac{f_{\bar{E},S}(e_0, e_1, ..., e_n, s)}{f_S(s)} = \frac{e^{-s}}{\frac{e^{-s} s^n}{n!}}$$

$$f_{\bar{E}|S}(e_0, e_1, ..., e_n|s) = n! s^{-n}$$

$$f_{\bar{E}|S}(e_0, e_1, ..., e_n|s=1) = n! \text{ which is same as the joint PDF of } \bar{X} \text{ as in (2)}$$

$$f_{E_i|S}(e_i|s) = \frac{f_{E_i,S}(e_i,s)}{f_S(s)} = \frac{f_{E_i,S}(E_i = e_i, S = s)}{f_S(s)}$$

$$= \frac{f_{E_i,S}(E_i = e_i, \sum_{i=0}^{n} E_i = s)}{f_S(s)}$$

$$= \frac{f_{E_i,S}(E_i = e_i, E_0 + E_1 + ... + E_{i-1} + E_{i+1} + ... + E_n = s - e_i)}{f_S(s)}$$

$$= e^{-e_i} \frac{\frac{e^{-(s-e_i)}(s-e_i)^{n-1}}{(n-1)!}}{\frac{e^{-s} s^n}{n!}}$$

$$f_{E_i|S}(e_i|s) = \frac{n}{s}\left(1 - \frac{e_i}{s}\right)^{n-1}$$

$$f_{E_i|S}(e_i|s=1) = n(1-e_i)^{n-1} \text{ which is same as the PDF of } X_i \text{ as in (1)}$$

Both the Joint density and Density of $X_i$ $\forall$ $i$ are same for IID exp RVs conditioned on their sum $= 1$
$\implies$ We can represent $X_1, X_2, ...X_{n+1}$ as IID Exponential Random Variables $E_0, E_2, ...E_n$ with $\lambda = 1$ and impose condition that $S = \sum_{i=0}^{n} E_i = 1$