Theory Part

Exercise 1:

Assume that $x_1, \dots, x_n$ are the ordered samples from a uniform distribution $U[0,\theta]$ with unknown endpoints θ. The cumulative distribution function is:

$$F(x; 0, \theta) = \frac{x-0}{\theta-0} = \frac{x}{\theta}$$

Therefore, individual spacing are given by

$$D_1 = \frac{x_1 - x_0}{\theta - 0} = \frac{x_1 - 0}{\theta - 0} = \frac{x_1}{\theta}$$

$$D_i = \frac{x_i - x_{i-1}}{\theta}, \text{ where } i = 2, 3, \dots, n$$

$$D_{n+1} = \frac{x_{n+1} - x_{n+1-1}}{\theta - 0} = \frac{\theta - x_n}{\theta}$$

$$[\because x_{n+1} = \theta]$$

Calculate the geometric mean

$$GM = \sqrt[n+1]{D_1, D_i \cdot D_{n+1}}$$

$$= \left(\frac{x_1}{\theta} \cdot \sum_{i=2}^{n} \frac{x_i - x_{i-1}}{\theta} \cdot \frac{\theta - x_n}{\theta} \right)^{1/n+1}$$

$$= \left[\frac{1}{\theta^{n+1}} \left(x_1 \cdot \sum_{i=2}^{n} (x_i - x_{i-1}) \cdot (\theta - x_n) \right) \right]^{\frac{1}{n+1}}$$

$$= \frac{1}{\theta} \left(x_1 \cdot \sum_{i=2}^{n} (x_i - x_{i-1}) \cdot (\theta - x_n) \right)^{\frac{1}{n+1}}$$

Take the logarithm of GM we find.

$$S_n(0, \theta) = \log\left(\frac{1}{\theta}\right) + \frac{1}{n+1} \log x_1 + \frac{1}{n+1} \log \sum_{i=2}^{n} (x_i - x_{i-1}) + \frac{1}{n+1} \log(\theta - x_n)$$

$$= \frac{1}{n+1} \log x_1 + \frac{1}{n+1} \sum_{i=2}^{n} \log(x_i - x_{i-1}) + \frac{1}{n+1} \log(\theta - x_n) - \log(\theta)$$

Differentiating w.r.to θ.

$$\frac{\partial (S_n)}{\partial \theta} = 0 + 0 + \frac{1}{n+1} \cdot \frac{1}{\theta - x_n} (1) - \frac{1}{\theta}$$

$$= \frac{1}{(n+1)(\theta - x_n)} - \frac{1}{\theta}$$

Setting the value to 0

$$-\frac{1}{\theta} + \frac{1}{(n+1)(\theta - x_n)} = 0$$

$$\Rightarrow \frac{-(n+1)(\theta - x_n) + \theta}{\theta(n+1)(\theta - x_n)} = 0$$

$$\Rightarrow -n\theta + nx_n - \cancel{\theta} + x_n + \cancel{\theta} = 0$$

$$\Rightarrow -n\theta + nx_n + x_n = 0$$

$$\Rightarrow n\theta = (n+1)x_n$$

$$\therefore \hat{\theta}_{MS} = \frac{(n+1)x_n}{n}$$

Second Derivative to check global maximum,

$$\frac{\partial^2 (S_n)}{\partial \theta^2} = \frac{1}{\theta^2} - \frac{1}{(n+1)(\theta - x_n)^2}$$

$$= \frac{1}{\frac{(n+1)^2 x_n^2}{n^2}} - \frac{1}{(n+1)\left(\frac{(n+1)x_n}{n} - x_n\right)^2}$$

[put the value of $\hat{\theta}_{MS}$ in θ]

$$= \frac{n^2}{(n+1)^2 x_n^2} - \frac{n^2}{(n+1)((n+1)x_n - nx_n)^2}$$

$$= \frac{n^2}{(n+1)^2 x_n^2} - \frac{n^2}{(n+1) x_n^2}$$

$$= \frac{n^2}{(n+1) x_n^2} \left(\frac{1}{n+1} - 1 \right)$$

$$= \frac{n^2}{(n+1) x_n^2} \left(\frac{1 - n - 1}{n+1} \right)$$

$$= \frac{n^2}{(n+1) x_n^2} \left(\frac{-n}{n+1} \right)$$

$$= -\frac{n^3}{(n+1)^2 x_n^2}$$

Which is < 0, because n is always greater than 0, and the value of x_n start greater than 0.

So, $\boxed{\hat{\theta}_{MS} = \frac{(n+1)x_n}{n}}$

Programming Part

In [51]:

```python
import numpy as np
```

In [53]:

```python
for i in range(1, 4):
    file = 'sampleset_'+str(i)+'_problemsheet4_ex1.txt'
    #print(file)
    with open(file) as f:
        num_list = [line.rstrip() for line in f]
        num_list_int = [float(x) for x in num_list]
        sort_value = sorted(num_list_int)
        n = len(sort_value)
        x_n = sort_value[n-1]
        theta_ms = ((n + 1) * x_n) / n
        print("For dataset " + file + " the maximum spacing estimator is: "+ str(theta_ms) + "\n")
```

```
For dataset sampleset_1_problemsheet4_ex1.txt the maximum spacing estimator is: 4.011813333333335

For dataset sampleset_2_problemsheet4_ex1.txt the maximum spacing estimator is: 3.91578

For dataset sampleset_3_problemsheet4_ex1.txt the maximum spacing estimator is: 4.1274
```