Let $(x_1, ..., x_n) \in (0, \infty)^n$ be a realisation of independent on $[0, \theta]$ uniformly distributed random variables $X_1, ..., X_n$. What is Maximum Spacing Estimator in this case? Using the data set provide on Moodle computer the unknown parameter θ via the Maximum spacing estimator for the three different sets of samples (note that they are of different sizes).

Solution:

Assume that $x_{(1)}, ..., x_{(n)}$ are the ordered samples from a uniform distribution $U[0, \theta]$ with unknown endpoints θ. The cumulative distribution function is [1]:

$$F(x;\ 0,\theta) = \frac{x-0}{\theta-0} = \frac{x}{\theta} \quad \text{………………………………………………...…} (1)$$

Therefore, individual spacings are given by

$$D_1 = \frac{x_{(1)} - x_{(0)}}{\theta - 0} = \frac{x_{(1)} - 0}{\theta} = \frac{x_{(1)}}{\theta} \quad \text{………………………………………………...…} (2)$$

$$D_i = \frac{x_{(i)} - x_{(i-1)}}{\theta} \; where\ i = 2, 3, \ldots\ldots, n \quad \text{…………………………………………} (3)$$

$$D_{n+1} = \frac{x_{(n+1)} - x_{(n+1-1)}}{\theta - 0} = \frac{\theta - x_{(n)}}{\theta} \quad \text{………………………..……………………} (4)$$

When the geometric mean* is calculated and then the logarithm is taken and then the S_n will be:

$$S_n(0,\theta) = \frac{1}{n+1} ln(x_{(1)}) + \sum_{i=2}^{n} ln\,(x_{(i)} - x_{(i-1)}) + \frac{1}{n+1}\, ln(\theta - x_{(n)}) - ln(\theta) \quad \text{…....} (5)$$

In equation (5) only the third term depends on the parameters θ. Differentiating with respect to θ, we got

$$\frac{d}{d\theta}\left[\frac{1}{n+1}\, ln(\theta - x_{(n)}) - ln(\theta)\right] = \frac{1}{n+1} * \frac{d}{d\theta}\left[ln(\theta - x_{(n)})\right] - \frac{d}{d\theta}\left[ln(\theta)\right]$$

$$= \frac{\frac{1}{\theta+1} * \frac{d}{d\theta}[\theta - x_{(n)}]}{n+1} - \frac{1}{\theta}$$

$$= \frac{\frac{d}{d\theta}[\theta] + \frac{d}{d\theta}[-x_{(n)}]}{(n+1)(\theta - x_{(n)})} - \frac{1}{\theta}$$

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

$$= \frac{1}{(n+1)(\theta - x_{(n)})} - \frac{1}{\theta} \quad \text{……………………………} (6)$$

*The geometric mean is defined as the n^{th} root of the product of n numbers, i.e., for a set of numbers $x_1, x_2, ..., x_n$, the geometric mean is defined as [3]:

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \sqrt[n]{x_1 x_2 \cdots x_n}$$

Solving $\frac{d}{d\theta} S_{(n)} = 0$, the maximum spacing estimates estimators of θ is:

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

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

$$\Rightarrow -n\theta + nx_{(n)} + x_{(n)} = 0$$

$$\Rightarrow \theta = \frac{x_{(n)} * (n+1)}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

Now, for “sampleset_1_problemsheet4_ex1.txt”, if we sort all the *x* then $x_{(n)} = 3.8824$ $n = 30$ so according to equation (7)

$$\theta = \frac{3.8824 * (30 + 1)}{30} = 4.011813333$$

for “sampleset_2_problemsheet4_ex1.txt”, if we sort all the *x* then $x_{(n)} = 3.839, n = 50$ so according to equation (7)

$$\theta = \frac{3.839 * (50 + 1)}{50} = 3.91578$$

for “sampleset_2_problemsheet4_ex1.txt”, if we sort all the *x* then $x_{(n)} = 3.6688, n = 8$ so according to equation (7)

$$\theta = \frac{3.6688 * (8 + 1)}{8} = 4.1274$$

Reference:

[1]. Maximum spacing estimation, https://en.wikipedia.org/wiki/Maximum_spacing_estimation

[2]. Cheng, R. C. H., & Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. Journal of the Royal Statistical Society: Series B (Methodological), 45(3), 394-403.

[3]. Geometric mean, https://en.wikipedia.org/wiki/Geometric_mean