Mini Project 3 Report

Jiadao Zou
jxz172230

Q1

(8 points) Suppose we would like to estimate the parameter θ (> 0) of a Uniform (0, θ) population based on a random sample X1, . . . , Xn from the population. In the class, we have discussed two estimators for θ — the maximum likelihood estimator, θˆ1 = X(n), where X(n) is the maximum of the sample, and the method of moments estimator, θˆ2 = 2X, where X is the sample mean. The goal of this exercise is to compare the mean squared errors of the two estimators to determine which estimator is better. Recall that the mean squared error of an estimator θˆ of a parameter θ is defined as E{(θˆ— θ)2}. For the comparison, we will focus on n = 1, 2, 3, 5, 10, 30 and θ = 1,5,50,100.

1. (a) Explain how you will compute the mean squared error of an estimator using Monte Carlo simulation.
   I would computing the mean squared errors ($E(\theta-\widehat{\theta})^2$) according to both the maximum likelihood estimator ($\widehat{\theta}=X_{max}$) and momentum estimator ($\widehat{\theta}=2X_{mean}$). Since one time simulation has little confidence, I would use Monte Carlo method to generate samples which are large enough to have a high confidence.
2. (b) For a given combination of (n, θ), compute the mean squared errors of both θˆ1 and θˆ2 using Monte Carlo simulation with N = 1000 replications. Be sure to compute both estimates from the same data.
   Please see below.

3. (c) Repeat (b) for the remaining combinations of (n,θ). Summarize your results graphically.

```
MSE of MLE Estimator:
        theta = 1  theta = 5  theta = 50  theta = 100
n = 1    0.339281   8.313311  793.683456  3365.048649
n = 2    0.167847   4.036426  422.035022  1541.175043
n = 3    0.105671   2.440826  263.047794  1003.078964
n = 5    0.046911   1.228923  117.231859   460.556421
n = 10   0.015067   0.366818   40.298466   137.281992
n = 30   0.001740   0.050091    5.238040    18.221776

MSE of Moment Estimator:
        theta = 1  theta = 5  theta = 50  theta = 100
n = 1    0.324174   8.721065  799.884406  3352.898568
n = 2    0.175618   4.301449  421.263214  1666.594559
n = 3    0.118644   2.627964  280.486726  1169.054005
n = 5    0.068588   1.759420  156.627211   654.598378
n = 10   0.035196   0.733607   83.593162   314.092408
n = 30   0.011664   0.272113   29.174323   112.253773
```

For details, please see the code section.

4. (d) Based on (c), which estimator is better? Does the answer depend on n or θ? Explain. Provide justification for all your conclusions.
   – Generally speaking, with n get bigger, the MLE estimator would beat over Moment estimator. As we could see from the table, if the n is pretty small, the sample may tend to be biased and in some case, it would show the contrary result. However, it could be take as reasonable variance.
   – The n and theta would influence the Mean Square Error, With n getting larger or theta getting smaller, the MSE would become smaller and smaller.

Q2

(12 points) Suppose the lifetime, in years, of an electronic component can be modeled by a continuous random variable with probability density function

$$f(x) = \begin{cases} \frac{\theta}{x^{\theta+1}} & x \geq 1, \\ 0, & x < 1, \end{cases}$$

where θ > 0 is an unknown parameter. Let X1, . . . , Xn be a random sample of size n from this population.

- (a) Derive an expression for maximum likelihood estimator of θ.

$$likelihood: \quad L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \frac{\theta^n}{(\prod_{i=1}^{n} X_i)^{\theta+1}}$$

$$log - likelohood: \quad \ln L(\theta) = n\ln(\theta) - (\theta+1)\sum_{i=1}^{n} \ln X_i$$

$$take\ Derivative: \quad \frac{d\ln L(\theta)}{d(\theta)} = \frac{n}{\theta} - \sum_{i=1}^{n} \ln X_i$$

$$make\ derivative = 0: \quad \hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln X_i}$$

- (b) Suppose n = 5 and the sample values are x1 = 21.72, x2 = 14.65, x3 = 50.42, x4 = 28.78, x5 = 11.23. Use the expression in (a) to provide the maximum likelihood estimate for θ based on these data.
  0.32338742. For details, please see the code section

- (c) Even though we know the maximum likelihood estimate from (b), use the data in (b) to obtain the estimate by numerically maximizing the log–likelihood function using optim function in R. Do your answers match?
  – Since I use Python's library SciPy here, I am not sure about the inner technique's difference between it and R, such as step size, iteration number, minimizing algorithms, etc. My answer is 52.15, give the initial position is $\theta = 1$.
  – Comparing answers of (b) and (c), the answers don't match.

```
t:  [0.32338742]  and function minimum value is  [4.59591893e-12]
t:  [52.15]  and function minimum value is  [0.]
```

- d) Use the output of numerical maximization in (c) to provide an approximate standard error of the maximum likelihood estimate and an approximate 95% confidence interval for θ. Are these approximations going to be good? Justify your answer.

$$Standard\,Error = \frac{Standard\,Deviation}{\sqrt{N}} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{n}(X_i - \bar{X})^2}}{\sqrt{N}}$$

$$CI: \quad \bar{X} \pm t_{\alpha/2,4} * SE \in [-2.30 * e - 55, \quad 5.402 * e - 55]$$

These approximations sadly are not good cause sample space is too small, also we don't know the variance of true distribution but only simulated it by computing sample variance.