

# Solution to Gate ST 2023 Q 26

Mayank Gupta

Question : Consider the following regression model

$$y_k = \alpha_0 + \alpha_1 \log_e k + \epsilon_k, \quad k = 1, 2, \dots, n,$$

where  $\epsilon_k$ 's are independent and identically distributed random variables each having probability density function  $f(x) = \frac{1}{2}e^{-|x|}, x \in \mathbb{R}$ . Then which one of the following statements is true?

- (A) The maximum likelihood estimator of  $\alpha_0$  does not exist
- (B) The maximum likelihood estimator of  $\alpha_1$  does not exist
- (C) The least squares estimator of  $\alpha_0$  exists and is unique
- (D) The least squares estimator of  $\alpha_1$  exists, but it is not unique

**Solution:**

$$f(\epsilon_k) = \frac{1}{2}e^{-|\epsilon_k|} \quad (1)$$

$$\text{Likelihood function : } f(\epsilon_1 \epsilon_2 \dots \epsilon_n) = \prod_{k=1}^n f(\epsilon_k) \quad (2)$$

$$L = \prod_{k=1}^n \frac{1}{2}e^{-|\epsilon_k|} \quad (3)$$

$$L_1 = \ln L = \ln \left( \prod_{k=1}^n \frac{1}{2}e^{-|\epsilon_k|} \right) \quad (4)$$

$$= \sum_{k=1}^n \ln \left( \frac{1}{2}e^{-|\epsilon_k|} \right) \quad (5)$$

$$= \sum_{k=1}^n (-\ln 2 - |y_k - \alpha_0 - \alpha_1 \log_e k|) \quad (6)$$

$$= -n \ln 2 - \sum_{k=1}^n (|y_k - \alpha_0 - \alpha_1 \log_e k|) \quad (7)$$

$$L_1 = \text{function of } \alpha_0, \alpha_1 \quad (8)$$

1) Maximum likelihood estimator

We need to find the value of  $\alpha_0$  and  $\alpha_1$  which will maximise the value of  $L_1$  i.e. the value of  $\alpha_0$  and  $\alpha_1$  which will minimise the value of  $\sum_{k=1}^n |y_k - \alpha_0 - \alpha_1 \log_e k|$

a) With respect to  $\alpha_0$

i) For  $y_k - \alpha_0 - \alpha_1 \log_e k > 0$

$$\min_{\alpha_0} y_k - \alpha_0 - \alpha_1 \log_e k$$

$$\text{s.t. } \alpha_0 \leq y_k - \alpha_1 \log_e k$$

Using Lagrange multiplier method

$$L(\lambda) = y_k - \alpha_0 - \alpha_1 \log_e k - \lambda(\alpha_0 - y_k + \alpha_1 \log_e k) \quad (9)$$

$$\frac{\partial L}{\partial \alpha_0} = -1 - \lambda = 0 \quad (10)$$

$$\frac{\partial L}{\partial \lambda} = y_k - \alpha_0 - \alpha_1 \log_e k = 0 \quad (11)$$

$$\lambda = -1 \quad (12)$$

$$\alpha_0 = y_k - \alpha_1 \log_e k \quad (13)$$

ii) For  $y_k - \alpha_0 - \alpha_1 \log_e k < 0$

$$\begin{aligned} \min_{\alpha_0} \quad & -(y_k - \alpha_0 - \alpha_1 \log_e k) \\ \text{s.t.} \quad & \alpha_0 \geq y_k - \alpha_1 \log_e k \end{aligned}$$

Using Lagrange multiplier method

$$L(\lambda) = -(y_k - \alpha_0 - \alpha_1 \log_e k) - \lambda(\alpha_0 - y_k + \alpha_1 \log_e k) \quad (14)$$

$$\frac{\partial L}{\partial \alpha_0} = 1 - \lambda = 0 \quad (15)$$

$$\frac{\partial L}{\partial \lambda} = y_k - \alpha_0 - \alpha_1 \log_e k = 0 \quad (16)$$

$$\lambda = 1 \quad (17)$$

$$\alpha_0 = y_k - \alpha_1 \log_e k \quad (18)$$

As value of  $\alpha_0$  matches for both cases of modulus

$\therefore$  The maximum likelihood estimator of  $\alpha_0$  exist

b) With respect to  $\alpha_1$

i) For  $y_k - \alpha_0 - \alpha_1 \log_e k > 0$

$$\begin{aligned} \min_{\alpha_1} \quad & y_k - \alpha_0 - \alpha_1 \log_e k \\ \text{s.t.} \quad & \alpha_1 \leq \frac{y_k - \alpha_0}{\log_e k} \end{aligned}$$

Using Lagrange multiplier method

$$L(\lambda) = y_k - \alpha_0 - \alpha_1 \log_e k - \lambda \left( \alpha_1 - \frac{y_k - \alpha_0}{\log_e k} \right) \quad (19)$$

$$\frac{\partial L}{\partial \alpha_1} = -\log_e k - \lambda = 0 \quad (20)$$

$$\frac{\partial L}{\partial \lambda} = - \left( \alpha_1 - \frac{y_k - \alpha_0}{\log_e k} \right) = 0 \quad (21)$$

$$\lambda = -\log_e k \quad (22)$$

$$\alpha_1 = \frac{y_k - \alpha_0}{\log_e k} \quad (23)$$

ii) For  $y_k - \alpha_0 - \alpha_1 \log_e k < 0$

$$\begin{aligned} \min_{\alpha_1} \quad & -(y_k - \alpha_0 - \alpha_1 \log_e k) \\ \text{s.t.} \quad & \alpha_1 \geq \frac{y_k - \alpha_0}{\log_e k} \end{aligned}$$

Using Lagrange multiplier method

$$L(\lambda) = -(y_k - \alpha_0 - \alpha_1 \log_e k) - \lambda \left( \alpha_1 - \frac{y_k - \alpha_0}{\log_e k} \right) \quad (24)$$

$$\frac{\partial L}{\partial \alpha_1} = \log_e k - \lambda = 0 \quad (25)$$

$$\frac{\partial L}{\partial \lambda} = - \left( \alpha_1 - \frac{y_k - \alpha_0}{\log_e k} \right) = 0 \quad (26)$$

$$\lambda = \log_e k \quad (27)$$

$$\alpha_1 = \frac{y_k - \alpha_0}{\log_e k} \quad (28)$$

As value of  $\alpha_1$  matches for both cases of modulus

$\therefore$  The maximum likelihood estimator of  $\alpha_1$  exist

$\therefore$  Option (A) and (B) are incorrect

c) Least square estimator

The least square estimator of  $\alpha_0$  and  $\alpha_1$  is  $\tilde{\alpha}_0$  and  $\tilde{\alpha}_1$  which will minimise

parameter	value	description
$\bar{y}$	$\frac{1}{n} \sum_{k=1}^n y_k$	Average value of $y_k$
$\bar{x}$	$\frac{1}{n} \sum_{k=1}^n \log_e k$	Average value of $\log_e k$

TABLE 1

VARIABLES USED

$$Q(\alpha_0, \alpha_1) = \sum_{k=1}^n (y_k - \alpha_0 - \alpha_1 \log_e k)^2 \quad (29)$$

$$\frac{\partial Q}{\partial \alpha_0} = -2 \sum_{k=1}^n (y_k - \alpha_0 - \alpha_1 \log_e k) = 0 \quad (30)$$

$$\sum_{k=1}^n (y_k - \alpha_0 - \alpha_1 \log_e k) = 0 \quad (31)$$

$$n\bar{y} - n\alpha_0 - \alpha_1 n\bar{x} = 0 \quad (32)$$

$$\implies \tilde{\alpha}_0 = \bar{y} - \tilde{\alpha}_1 \bar{x} \quad (33)$$

$$\frac{\partial Q}{\partial \alpha_1} = -2 \sum_{k=1}^n (y_k - \alpha_0 - \alpha_1 \log_e k) \log_e k = 0 \quad (34)$$

$$\implies \tilde{\alpha}_1 = \frac{\sum_{k=1}^n (\log_e k - \bar{x})(y_k - \bar{y})}{\sum_{k=1}^n (\log_e k - \bar{x})^2} \quad (35)$$

$\therefore$  Least square estimator of  $\alpha_0$  and  $\alpha_1$  exists and are unique

$\therefore$  Option (C) is correct and (D) is incorrect

d) Steps for simulation the given distribution whose probability density function is  $f(x) = \frac{1}{2}e^{-|x|}$

i) Write a function cdf for calculating the cdf of any random variable

$$p_X(x) = \begin{cases} \frac{1}{2}e^x & x \leq 0 \\ \frac{1}{2}e^{-x} & x > 0 \end{cases} \quad (36)$$

$$F_X(x) = \begin{cases} \int_{-\infty}^x \left(\frac{1}{2}e^x\right) dx & x \leq 0 \\ \int_{-\infty}^0 \left(\frac{1}{2}e^x\right) dx + \int_0^x \left(\frac{1}{2}e^{-x}\right) dx & x > 0 \end{cases} \quad (37)$$

$$F_X(x) = \begin{cases} \frac{1}{2}e^x & x \leq 0 \\ \frac{1}{2}(2 - e^{-x}) & x > 0 \end{cases} \quad (38)$$

- ii) Declare a function inverse cdf ( $I(u)$ ) such that its input is any random number and output is random variable whose cdf equals that of the given distribution

For  $x \leq 0$

$$u = \frac{1}{2}e^x \quad (39)$$

$$e^x = 2u \quad (40)$$

$$x = \ln 2u \quad (41)$$

$$\because x \leq 0 \quad (42)$$

$$u \leq 0.5 \quad (43)$$

For  $x > 0$

$$u = \frac{1}{2}(2 - e^{-x}) \quad (44)$$

$$2 - e^{-x} = 2u \quad (45)$$

$$e^{-x} = 2 - 2u \quad (46)$$

$$x = -\ln(2 - 2u) \quad (47)$$

$$\because x > 0 \quad (48)$$

$$u > 0.5 \quad (49)$$

$$I(u) = \begin{cases} \ln(2u) & u \leq 0.5 \\ -\ln(2 - 2u) & u > 0.5 \end{cases} \quad (50)$$

- iii) Define three arrays `random_vars` , `cdf_values` , `theoretical_cdf_values` to store random variables, simulated cdf values and theoretical cdf values
- iv) Generate random numbers using `rand()` and calling inverse cdf function to generate our random variable
- v) Calling cdf function to calculate the cdf of the generated random variable
- vi) Storing the random variable,theoretical cdf and generated cdf into their respective arrays
- vii) Storing the data of these three array into a .dat file
- viii) Plotting these .dat file in python