

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 ϵ_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 α_0 does not exist
- (B) The maximum likelihood estimator of α_1 does not exist
- (C) The least squares estimator of α_0 exists and is unique
- (D) The least squares estimator of α_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 α_0 and α_1 which will maximise the value of L_1 i.e. the value of α_0 and α_1 which will minimise the value of $\sum_{k=1}^n |y_k - \alpha_0 - \alpha_1 \log_e k|$

a) With respect to α_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 α_0 matches for both cases of modulus

\therefore The maximum likelihood estimator of α_0 exist

b) With respect to α_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 α_1 matches for both cases of modulus

\therefore The maximum likelihood estimator of α_1 exist

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

c) Least square estimator

The least square estimator of α_0 and α_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 α_0 and α_1 exists and are unique

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