Solution to Gate ST 2023 Q 26

Mayank Gupta

Question : Consider the following regression model

$$y_k = \alpha_0 + \alpha_1 \log_e k + \epsilon_k, \qquad k = 1, 2, \ldots, n,$$

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:

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

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

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

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

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

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

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

$$L_1 = \text{function of } \alpha_0, \alpha_1 \tag{8}$$

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 $|y_k - \alpha_0 - \alpha_1 \log_e k|$

$$|y_k - \alpha_0 - \alpha_1 \log_e k| = \begin{cases} y_k - \alpha_0 - \alpha_1 \log_e k, & y_k - \alpha_0 - \alpha_1 \log_e k > 0 \\ -\left(y_k - \alpha_0 - \alpha_1 \log_e k\right), & y_k - \alpha_0 - \alpha_1 \log_e k < 0 \end{cases} \tag{9}$$

To minimize $\sum_{k=1}^{n}\left(|y_k - \alpha_0 - \alpha_1 \log_e k|\right)$, we should choose α_0 and α_1 such that as many terms as possible fall into the second case (negative contribution) and as few terms as possible fall into the second case (positive contribution).

$$y_k - \alpha_1 \log_e k < \alpha_0 \tag{10}$$

$$\therefore \tilde{\alpha_0} = \min_{\left(k|y_k-\alpha_0-\alpha_1 \log_e k<0\right)} \left(y_k - \alpha_1 \log_e k\right) \tag{11}$$

$$y_k - \alpha_0 < \alpha_1 \log_e k \tag{12}$$

$$\therefore \tilde{\alpha_1} = \min_{\left(k|y_k-\alpha_0-\alpha_1 \log_e k<0\right)} \left(\frac{y_k - \alpha_0}{\log_e k}\right) \tag{13}$$

$$\tag{14}$$

$\therefore$ Maximum likelihood estimator of α_0 and α_1 exist
$\therefore$ Option (A) and (B) are incorrect

2) Least square estimator
The least square estimator of α_0 and α_1 is $\tilde{\alpha_0}$ and $\tilde{\alpha_1}$ which will minimise

$$Q\left(\alpha_0, \alpha_1\right) = \sum_{k=1}^{n} \left(y_k - \alpha_0 - \alpha_1 \log_e k\right)^2 \tag{15}$$

$$\frac{\partial Q}{\partial \alpha_0} = -2 \sum_{k=1}^{n} \left(y_k - \alpha_0 - \alpha_1 \log_e k\right) = 0 \tag{16}$$

$$\sum_{k=1}^{n} \left(y_k - \alpha_0 - \alpha_1 \log_e k\right) = 0 \tag{17}$$

$$n\bar{y} - n\alpha_0 - \alpha_1 n\bar{x} = 0, \quad \text{where } \bar{x} = \frac{1}{n}\sum_{k=1}^{n} \log_e k, \quad \bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k \tag{18}$$

$$\implies \tilde{\alpha_0} = \bar{y} - \tilde{\alpha_1}\bar{x} \tag{19}$$

$$\frac{\partial Q}{\partial \alpha_1} = -2 \sum_{k=1}^{n} \left(y_k - \alpha_0 - \alpha_1 \log_e k\right) \log_e k = 0 \tag{20}$$

$$\implies \tilde{\alpha_1} = \frac{\sum_{k=1}^{n}\left(\log_e k - \bar{x}\right)\left(y_k - \bar{y}\right)}{\sum_{k=1}^{n}\left(\log_e k - \bar{x}\right)^2} \tag{21}$$

$\therefore$ Least square estimator of α_0 and α_1 exists and are unique
$\therefore$ Option (C) is correct and (D) is incorrect