

Solution of GATE-ST 2023 Q58

SUJAL GUPTA - EE22BTECH11052

Consider the following regression model

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \epsilon_t, \quad t = 1, 2, \dots, n \quad (1)$$

where α_0 , α_1 and α_2 are unknown parameters and ϵ_t 's are independent and identically distributed random variables each having $\mathcal{N}(\mu, 1)$ distribution with μ unknown. Then which of the following statements is/are true?

- 1) There exists an unbiased estimator of α_1
- 2) There exists an unbiased estimator of α_2
- 3) There exists an unbiased estimator of α_0
- 4) There exists an unbiased estimator of μ

Solution: Assuming that the model is

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \epsilon_t \quad (2)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & A_{11} & A_{12} \\ 1 & A_{21} & A_{22} \\ \vdots & & \\ 1 & A_{n1} & A_{n2} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (3)$$

Finding mean of the y,

$$\mathbf{y} = \mathbf{A}\mathbf{a} + \boldsymbol{\epsilon} \quad (4)$$

$$E(\mathbf{y}) = E(\mathbf{A}\mathbf{a} + \boldsymbol{\epsilon}) \quad (5)$$

$$= \mathbf{A}\mathbf{a} + E(\boldsymbol{\epsilon}) \quad (6)$$

$$E(\boldsymbol{\epsilon}) = \begin{bmatrix} E(\epsilon_1) \\ E(\epsilon_2) \\ \vdots \\ E(\epsilon_n) \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} \quad (7)$$

$$\boldsymbol{\mu}_\epsilon = \mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (8)$$

Let \mathbf{A}_i represent the i^{th} row of \mathbf{A}

$$\text{Cov}(y_i, y_j) = E\left((y_i - E(y_i))(y_j - E(y_j))\right) \quad (9)$$

$$= E\left((\mathbf{A}_i \mathbf{a} + \epsilon_i - E(\mathbf{A}_i \mathbf{a} + \epsilon_i))(\mathbf{A}_j \mathbf{a} + \epsilon_j - E(\mathbf{A}_j \mathbf{a} + \epsilon_j))\right) \quad (10)$$

$$= E\left((\epsilon_i - E(\epsilon_i))(\epsilon_j - E(\epsilon_j))\right) \quad (11)$$

$$= \text{Cov}(\epsilon_i, \epsilon_j) \quad (12)$$

$$\mathbf{C}_y = \mathbf{C}_\epsilon \quad (13)$$

Since the ϵ_i 's are independent and identical vectors,

$$\text{Cov}(\epsilon_i, \epsilon_j) = 0, \quad \forall i \neq j \quad (14)$$

$$\text{Var}(\epsilon_i) = 1, \quad \forall 1 \leq i \leq n \quad (15)$$

$$\mathbf{C}_y = I_{n \times n} \quad (16)$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mathbf{a} + \boldsymbol{\mu}_\epsilon, I) \quad (17)$$

$$p_y = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C}_y)}} \exp \frac{-(\mathbf{y} - \boldsymbol{\mu}_\epsilon - \mathbf{A}\mathbf{a})^T \mathbf{C}_y^{-1} (\mathbf{y} - \boldsymbol{\mu}_\epsilon - \mathbf{A}\mathbf{a})}{2} \quad (18)$$

where \mathbf{C}_y is the covariance matrix for \mathbf{y}

The maximum likelihood function can be written as:

$$L(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\frac{-(\mathbf{y} - \boldsymbol{\mu}_\epsilon - \mathbf{A}\mathbf{a})^T (\mathbf{y} - \boldsymbol{\mu}_\epsilon - \mathbf{A}\mathbf{a})}{2}} \quad (19)$$

$$\ln L(\mathbf{y}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_\epsilon - \mathbf{A}\mathbf{a})^T (\mathbf{y} - \boldsymbol{\mu}_\epsilon - \mathbf{A}\mathbf{a}) \quad (20)$$

$$\frac{\partial \ln L(\mathbf{y})}{\partial \mathbf{a}} = \frac{\partial (-\mathbf{y}^T \mathbf{A}\mathbf{a} - \mathbf{a}^T \mathbf{A}^T \mathbf{y} + \mathbf{a}^T \mathbf{A}^T \mathbf{A}\mathbf{a} + \mathbf{a}^T \mathbf{A}^T \boldsymbol{\mu}_\epsilon + \boldsymbol{\mu}_\epsilon^T \mathbf{A}\mathbf{a})}{\partial \mathbf{a}} \quad (21)$$

$$= -2\mathbf{A}^T \mathbf{y} + 2\mathbf{A}^T \mathbf{A}\mathbf{a} + 2\mathbf{A}^T \boldsymbol{\mu}_\epsilon \quad (22)$$

The normal equation is

$$\frac{\partial \ln L(\mathbf{y})}{\partial \mathbf{a}} = 0 \quad (23)$$

$$\mathbf{a} = (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{y} - \mathbf{A}^T \boldsymbol{\mu}_\epsilon) \quad (24)$$

For unbiased estimator,

$$E(\mathbf{a}) = \mathbf{a} \quad (25)$$

$$E(\mathbf{a}) = E\left((\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{y} - \mathbf{A}^T \boldsymbol{\mu}_\epsilon)\right) \quad (26)$$

$$= (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T E(\mathbf{y}) - \mathbf{A}^T \boldsymbol{\mu}_\epsilon) \quad (27)$$

$$= (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T (\mathbf{A}\mathbf{a} + \boldsymbol{\mu}_\epsilon) - \mathbf{A}^T \boldsymbol{\mu}_\epsilon) \quad (28)$$

$$= (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{A}) \mathbf{a} \quad (29)$$

$$= \mathbf{a} \quad (30)$$

Hence there exist unbiased estimator for $\alpha_0, \alpha_1, \alpha_2$

For Maximum Likelihood Estimator of μ

$$\frac{\partial \ln L(\mathbf{y})}{\partial \boldsymbol{\mu}_\epsilon} = \frac{\partial (-\mathbf{y}^T \boldsymbol{\mu}_\epsilon - \mathbf{a}^T \mathbf{A}^T \boldsymbol{\mu}_\epsilon + \boldsymbol{\mu}_\epsilon^T \mathbf{y} + \boldsymbol{\mu}_\epsilon^T \mathbf{A}\mathbf{x} + \boldsymbol{\mu}_\epsilon^T \boldsymbol{\mu}_\epsilon)}{\partial \boldsymbol{\mu}_\epsilon} \quad (31)$$

$$= -2\mathbf{y} + 2\mathbf{A}\mathbf{a} + 2\boldsymbol{\mu}_\epsilon \quad (32)$$

The normal equation is

$$\frac{\partial \ln L(\mathbf{y})}{\partial \boldsymbol{\mu}_\epsilon} = 0 \quad (33)$$

$$\boldsymbol{\mu}_\epsilon = \mathbf{y} - \mathbf{A}\mathbf{a} \quad (34)$$

$$E(\boldsymbol{\mu}_\epsilon) = E(\mathbf{y} - \mathbf{A}\mathbf{a}) \quad (35)$$

$$= E(\mathbf{y}) - \mathbf{A}\mathbf{a} \quad (36)$$

$$= E(\boldsymbol{\epsilon}) \quad (37)$$

$$= \boldsymbol{\mu}_\epsilon \quad (38)$$

Since,

$$\boldsymbol{\mu}_\epsilon = \mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (39)$$

Hence there exists an unbiased estimator for μ as well.

Simulation in C

- 1) Defining some arbitrary values for $\alpha_0, \alpha_1, \alpha_2$ and μ
- 2) Generating dependent variable y_i using $y_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \epsilon_i, \quad i = 1, 2, \dots, n$
- 3) Storing the generated data into a "generated_data.dat" file.
- 4) ϵ_i is generated as a gaussian variable
- 5) Defining 4 variables *beta0hat*, *beta1hat*, *beta2hat* and *mu - hat* for storing the estimated values
- 6) Defining 4 variables *beta0bias*, *beta1bias*, *beta2bias* and *mu - bias* for keeping track of the deviation of the estimated values from the true values
- 7) Using nested for loops to ensure that average bias for each coefficient is calculated over multiple simulations
- 8) Dividing the total bias by the number of simulations to get the average bias per simulation
- 9) We get the following values:
 Average Bias for Alpha0 : 0.016775
 Average Bias for Alpha1 : -0.003000
 Average Bias for Alpha2 : -0.000500
 Average Bias for mu : -0.000001