Solution to GATE ST 2023.40

Devansh Jain - EE22BTECH11018

Question: Let $X_1, X_2, ..., X_{10}$ be a random sample of size 10 from a population having $\mathcal{N}\left(0, \theta^2\right)$ distribution, where $\theta > 0$ is an unknown parameter. Let $T = \frac{1}{10}\sum_{i=1}^{10} X_i^2$. If the mean square error of $cT\,(c > 0)$ as an estimator of θ^2, is minimized at $c = c_0$, then the value of c_0 equals

(i) $\frac{5}{6}$
(ii) $\frac{2}{3}$
(iii) $\frac{3}{5}$
(iv) $\frac{1}{2}$

(GATE ST 2023)

Solution: The mean of T is given by,

$$E(T) = E\left(\frac{1}{10}\sum_{i=1}^{10} X_i^2\right) \tag{1}$$

$$= \frac{1}{10}\sum_{i=1}^{10} E\left(X_i^2\right) \tag{2}$$

Since

$$E\left(X^2\right) = V(X) + (E(X))^2 \tag{3}$$

Using (3)

$$E(T) = \frac{1}{10}\sum_{i=1}^{10} V(X_i) + E(X_i)^2 \tag{4}$$

$$= \frac{1}{10}\left(10\theta^2\right) \tag{5}$$

$$= \theta^2 \tag{6}$$

Similarly, using (3), the variance of T is given by,

$$V(T) = E\left(T^2\right) - E(T)^2 \tag{7}$$

$$= \frac{1}{100}\left(E\left(\left(\sum_{i=1}^{10} X_i^2\right)^2\right) - \left(E\left(\sum_{i=1}^{10} X_i^2\right)\right)^2\right) \tag{8}$$

But

$$\mathrm{E}\left(\left(\sum_{i=1}^{10} X_i^2\right)^2\right) = \mathrm{E}\left(\sum_{i=1}^{10}\sum_{j=1}^{10} X_i^2 X_j^2\right) \tag{9}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\left(X_i^2 X_j^2\right) \tag{10}$$

and

$$\left(\mathrm{E}\left(\sum_{i=1}^{10} X_i\right)\right)^2 = \left(\sum_{i=1}^{10} \mathrm{E}\left(X_i\right)\right)^2 \tag{11}$$

$$= \sum_{i=1}^{10} \sum_{j=1}^{10} \mathrm{E}\left(X_i\right) \mathrm{E}\left(X_j\right) \tag{12}$$

Using (10) , (12), and the definition of covariance,

$$V(T) = \frac{1}{100}\left(\sum_{i=1}^{10} \sum_{j=1}^{10} \left(E\left[X_i^2 X_j^2\right] - E\left(X_i^2\right) E\left(X_j^2\right)\right)\right) \tag{13}$$

$$= \frac{1}{100}\left(\sum_{i=1}^{10} \sum_{j=1}^{10} \operatorname{cov}\left(X_i^2, X_j^2\right)\right) \tag{14}$$

As all the variables are i.i.d's and are thus uncorrelated,

$$\operatorname{cov}\left(X_i^2, X_j^2\right) = \begin{cases} 0 & \text{if } i \neq j \\ V\left(X_i^2\right) & \text{if } i = j \end{cases} \tag{15}$$

Now,

$$V(T) = \frac{1}{100} \sum_{i=1}^{10} \operatorname{cov}\left(X_i^2, X_i^2\right) \tag{16}$$

$$= \frac{1}{100} \sum_{i=1}^{10} V\left(X_i^2\right) \tag{17}$$

Using (3) to find $V\left(X_i^2\right)$,

$$V\left(X_i^2\right) = E\left(X_i^4\right) - E\left(X_i^2\right)^2 \tag{18}$$

Using moment generating function to find $E\left(X_i^4\right)$ and $E\left(X_i^2\right)$,

$$M_X(t) = E\left(e^{tX}\right) \tag{19}$$

$$= e^{\frac{1}{2}\theta^2 t^2} \tag{20}$$

Differentiating it with respect to t,

$$\frac{dM_X(t)}{dt} = E\left(Xe^{tX}\right) \tag{21}$$

$$\frac{d^n M_X(t)}{dt^n} = E\left(X^n e^{tX}\right) \tag{22}$$

$$\frac{dM_X(t)}{dt} = \theta^2 t e^{\frac{1}{2}\theta^2 t^2} \tag{23}$$

$$\frac{d^2 M_X(t)}{dt^2} = \theta^2 e^{\frac{1}{2}\theta^2 t^2} + \theta^4 t^2 e^{\frac{1}{2}\theta^2 t^2} \tag{24}$$

$$\frac{d^3 M_X(t)}{dt^3} = \theta^4 t e^{\frac{1}{2}\theta^2 t^2} + 2\theta^4 t e^{\frac{1}{2}\theta^2 t^2} + \theta^6 t^3 e^{\frac{1}{2}\theta^2 t^2} \tag{25}$$

$$= 3\theta^4 t e^{\frac{1}{2}\theta^2 t^2} + \theta^6 t^3 e^{\frac{1}{2}\theta^2 t^2} \tag{26}$$

$$\frac{d^4 M_X(t)}{dt^4} = 3\theta^4 e^{\frac{1}{2}\theta^2 t^2} + 3\theta^6 t^2 e^{\frac{1}{2}\theta^2 t^2} + 3\theta^6 t^2 e^{\frac{1}{2}\theta^2 t^2} + \theta^8 t^4 e^{\frac{1}{2}\theta^2 t^2} \tag{27}$$

$$= 3\theta^4 e^{\frac{1}{2}\theta^2 t^2} + 6\theta^6 t^2 e^{\frac{1}{2}\theta^2 t^2} + \theta^8 t^4 e^{\frac{1}{2}\theta^2 t^2} \tag{28}$$

Now,

$$E\left(X^4\right) = \frac{d^4 M_X(t)}{dt^4}|_{t=0} \tag{29}$$
$$= 3\theta^4 \tag{30}$$
$$E\left(X^2\right) = \frac{d^2 M_X(t)}{dt^2}|_{t=0} \tag{31}$$
$$= \theta^2 \tag{32}$$

$$V\left(X_i^2\right) = 3\theta^4 - \left(\theta^2\right)^2 \tag{33}$$
$$= 2\theta^4 \tag{34}$$
$$V(T) = \frac{1}{100}\sum_{i=1}^{10} 2\theta^4 \tag{35}$$
$$= \frac{1}{5}\theta^4 \tag{36}$$

The mean square error of cT as an estimator of θ^2 is given by

$$f = E\left(\left(cT - \theta^2\right)^2\right) \tag{37}$$

Since

$$E\left(X^2\right) = V(X) + (E(X))^2 \tag{38}$$

Now,

$$f = V\left(cT - \theta^2\right) + \left(E\left(cT - \theta^2\right)^2\right) \tag{39}$$
$$= V(cT) + \left(E(cT) - \theta^2\right)^2 \tag{40}$$
$$= c^2 V(T) + \left(cE(T) - \theta^2\right)^2 \tag{41}$$
$$= \frac{c^2}{5}\theta^4 + \theta^4 (c-1)^2 \tag{42}$$

Minimizing the mean square error,

$$\frac{df}{dc} = 0 \tag{43}$$
$$\frac{df}{dc} = \frac{2c}{5}\theta^4 + 2\theta^4 (c-1) \tag{44}$$
$$\frac{c}{5} + c - 1 = 0 \tag{45}$$
$$6c = 5 \tag{46}$$
$$c = \frac{5}{6} \tag{47}$$

To verify that cT is a good estimate for θ, let

$$\theta = 0.5 \tag{48}$$
$$E\left(\left(cT - \theta^2\right)^2\right) = \frac{1}{5}\left(\frac{25}{36}\right)(0.5)^4 + \left(\frac{-1}{6}\right)^2 (0.5)^4 \tag{49}$$
$$= \frac{1}{96} \tag{50}$$
$$= 0.0104 \tag{51}$$

The mean square error is small relative to the value of θ, hence cT is a good estimate.

Simulation procedure:

(i) (double) rand()/ RAND_MAX:
Generates a uniform distribution between 0 and 1.

(ii) sqrt(variance)*(sqrt(-2 * log(u1)) * cos (2 *M_PI * u2)) + mean:
Transforms the uniform distribution into gaussian distribution of desired mean and variance. Ten such random variables are generated.

(iii) The values of the random variables are squared and then added together to generate T.

(iv) The value of c which minimizes the mean square error is found by calculating $E\left(\left(cT - \theta^2\right)^2\right)$ for a range of values of c and choosing that value of c which gives the minimum value for the expression.