UNIT 2

FISHER INFORMATION AND CRAMÉR-RAO INEQUALITY

Structure

2.1 INTRODUCTION

In the previous unit, you have studied some basic concepts which help us to find the best estimator (UMVUE) of a parameter. An estimator T of the parameter θ is said to be a uniformly minimum variance unbiased estimator of θ if and only if

(i) $E(T) = \theta$, that is, the estimator T is an unbiased estimator of the parameter θ; and

(ii) $Var(T) \leq Var(T')$ where T' is any other unbiased estimator of parameter θ.

The above definition implies that an estimator is a UMVUE if and only if the estimator is unbiased and if there is no other unbiased estimator that has a smaller variance for any value of θ. The practical question which then arises is how can we find such an estimator. The one way of seeking a UMVUE is facilitated by the Cramér-Rao inequality. It provides a lower bound on the variance of unbiased estimators of a parameter using the Fisher information. This unit is devoted to describing it.

Tools You Will Need

The following terms are considered essential background material for this Unit. If you doubt your knowledge of any of these terms, you should review the appropriate Unit or section before proceeding:

- Sampling distributions (Units 2,3, 4 and 5).
- Probability distributions (MST-012).

This unit is divided into eight sections. Section 2.1 is introductory in nature, which describes the need and role of Cramér-Rao inequality in estimation. The Cramér-Rao inequality uses the Fisher information, therefore, we discuss it in

Unit Writer- Dr. Prabhat Kumar Sangal, School of Sciences, IGNOU, New Delhi

Sections 2.2 and 2.3 of single and more parameters, respectively. Section 2.4 is devoted to explaining the Cramér-Rao inequality and how to apply it in real-life situations. The limitations of Cramér-Rao inequality are explained in Section 2.5. The unit ends by providing a summary of what we have discussed in this unit in Section 2.6. The terminal questions and the solution of the SAQs/TQs are given in Sections 2.7 and 2.8, respectively.

In the next unit, we shall discuss the second characteristic of a good estimator, that is, consistency.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ describe the concept of Fisher Information and its role in measuring the amount of information a random variable contains about an unknown parameter;
- ❖ apply Fisher Information to evaluate the efficiency of an estimator;
- ❖ explain the Cramér-Rao lower bound as a theoretical limit on the variance of unbiased estimators of a parameter;
- ❖ recognize the conditions under which the Cramér-Rao lower bound is valid;
- ❖ evaluate the efficiency of an estimator by comparing its variance to the Cramér-Rao lower bound; and
- ❖ recognize the role of Fisher Information and the Cramér-Rao Inequality in statistical theory, particularly in achieving optimality in estimation.

2.2 FISHER INFORMATION

In most of the real-life problems, the population parameter (characteristic of the population) is not known, and someone is interested in obtaining the value of the parameter. But, if

- the whole population is too large to study,
- the units of the population are destructive in nature,
- there are limited resources and manpower available, etc.

then it is not practically convenient to examine each unit of the population to find the value of the parameter. For example, as you know, many of us use Facebook, and you are interested to know the average age of the people who use Facebook. However, the true value (average age) of Facebook users is not known. The only way to know the true average age of Facebook users is to survey each and every person in the world who uses Facebook. But it is not possible to survey everyone in the world. In such a situation, one can select randomly some people who use Facebook and note their age.

Now, the question may arise, how do we know the amount of information that a random sample carries about an unknown parameter θ. Fisher information helps us to measure such information.

Fisher Information is a fundamental concept in statistics and information theory that quantifies the amount of information a random variable carries about an unknown parameter of a probability distribution. It is widely used in parameter estimation in classical as well as in Bayesian inference.

The Fisher information measures the amount of information that a random sample carries about an unknown parameter θ.

We now formally define the Fisher information as follows:

If $X_1, X_2, ..., X_n$ is a random sample of size n taken from a population whose probability density (mass) function is f(x; θ) where θ is the population parameter, then the Fisher information measures the amount of information that the random sample carries about an unknown parameter θ. We can calculate the Fisher information that a sample carries as follows:

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log L(\theta)\right)^2\right]$$

where,

- $\log L(\theta)$ is called the **log-likelihood** of the natural logarithm of the likelihood function.
- $\frac{\partial}{\partial\theta}\log L(\theta)$ the partial derivative with respect to θ of the natural logarithm of the likelihood function is called the **score**. It measures the gradient of the log-likelihood with respect to the parameter θ.

Since sample values $X_1, X_2, ..., X_n$ are independent and follow the same distribution, therefore, $L(\theta) = f(x_1;\theta).f(x_2;\theta) ... f(x_n;\theta) = nf(x;\theta)$. Thus, we can write the expression of the Fisher information as

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log L(\theta)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]$$

If log L(θ) is twice partially differentiable with respect to θ, and under certain additional regularity conditions (discussed in Section 2.4), then the Fisher information may also be written as

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right] = -nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]$$

Both definitions are equivalent under regularity conditions that allow interchanging differentiation and expectation.

Interpretation

Since the Fisher information measures the amount of information that a random sample provides about an unknown parameter θ. Therefore, if a sample has high Fisher information means:

- The estimate of the unknown parameter θ based on the sample data is more precise.
- The variance of the estimator for the unknown parameter θ (using the Cramér-Rao Lower Bound) is smaller, implying a more precise estimate.

To judge whether the Fisher information is high or not, we may find the variance of the estimator using the Cramér-Rao Lower Bound, which you will study in Section 2.4 as

$$\mathrm{Var}(T) > \frac{1}{I(\theta)}$$

If the variance is as low as 0.01, 0.05, etc. then we assume that the Fisher information is high.

Applications of Fisher Information

The Fisher information is widely used in parameter estimation in classical as well as in Bayesian inference. Some of them are given as follows:

- With the help of Fisher information, we can judge whether the estimate of the unknown parameter θ based on the sample data is more precise or not.
- The Fisher information is used to calculate the variance and covariance matrices associated with maximum-likelihood estimates.
- It is used to establish the lower bound on the variance of unbiased estimators using the Cramér-Rao bound.
- In Bayesian statistics, the Fisher information plays a role in the derivation of non-informative prior distributions according to Jeffreys' rule (you will study the Bayesian Inference in the 3rd block of this course).

To calculate the Fisher information, you can use either the probability density (mass) function or the likelihood function of the sample information with a single or double partial derivative. If we use a double derivative $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)$, then generally we get the constant expression whose expectation we can easily obtain.

After understanding the Fisher information with its applications and interpretation, let us discuss the process of finding it by taking some examples.

Example 1: Suppose a hospital monitors the time (in hours) until the failure of a particular type of medical device (e.g., a ventilator). The failure times are believed to follow an exponential distribution whose probability density function is given as follows:

$$f(x;\theta) = \theta e^{-\theta x}; \quad x > 0,\ \theta > 0$$

where θ represents the failure rate of the device.

The hospital administration collects data on the failure times of n devices. Calculate the Fisher Information about θ. If the hospital recorded n = 25 failure times and estimated the failure rate θ is 0.05, then calculate Fisher information and interpret it.

Solution: The probability density function of the exponential distribution with parameter θ is given as follows:

$$f(x;\theta) = \theta e^{-\theta x}; \quad x > 0,\ \theta > 0$$

For calculating Fisher information, there are different formulas. Here, we use the following formula to calculate it:

$$I(\theta) = -nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]$$

Therefore, we first take the natural logarithmic on both sides of the above expression, then we get

$$\log f(x;\theta) = \log(\theta) - \theta x$$

We partially differentiate the above expression with respect to the parameter θ, we get

$$\frac{\partial}{\partial\theta}\log f(x;\theta) = \frac{1}{\theta} - x \quad \left[\because \frac{\partial}{\partial x}(\log x) = \frac{1}{x}\right]$$

We now again partially differentiate the above expression with respect to the parameter θ to get the second derivative as follows:

$$\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) = -\frac{1}{\theta^2} \quad \left[\because \frac{\partial}{\partial x}\left(\frac{1}{x}\right) = -\frac{1}{x^2}\right]$$

Put the value of $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)$ in the expression of the Fisher information, we get

$$I(\theta) = -nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right] = -nE\left[-\frac{1}{\theta^2}\right]$$

$$= -n\left(-\frac{1}{\theta^2}\right) = \frac{n}{\theta^2} \qquad \text{[since E(a) = a]}$$

We now calculate the Fisher information for n = 25 and θ = 0.05 as

$$I(\theta) = \frac{n}{\theta^2} = \frac{25}{(0.12)^2} = 1736.11$$

We can calculate the variance of the estimator as

$$Var(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

For $I(\theta)$ = 1736.11

$$Var(\hat{\theta}) \geq \frac{1}{1736.11} = 0.000576$$

A very small variance indicates that the sample provides a high degree of information about the parameter θ and we get a reliable and consistent estimate of θ.

This high precision is beneficial for:

- Making accurate maintenance schedules for devices.
- Minimizing unexpected failures and downtime.
- Optimizing resource allocation in the healthcare facility.

I think you understood what the Fisher information is and how we obtain it. You have the curiosity to find it yourself. For that, you can try the following Self Assessment Question.

SAQ 1

Suppose the programme coordinator of the MSCAST programme randomly selected n learners and surveyed whether they have their own laptops or not. The $X_1, X_2, \ldots, X_n$ represent the output where

$X_i = 0$ if a randomly selected learner does not have his/her laptop and

$X_i = 1$ if a randomly selected learner has his/her laptop.

Assuming that the X_i are independent Bernoulli random variables with unknown and equal probability p that a learner has their own laptop then

(i) Find the Fisher information about the unknown probability p.

(ii) If the programme coordinator collected the information about the laptop of 16 randomly selected learners and obtained as 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1 then estimate the probability p that a learner has own laptop.

(iii) Calculate the Fisher information and interpret it.

The above formulas for calculating the Fisher information are used when we have to find it for a single unknown parameter. But in real life, we have to deal with more than one unknown parameter. In such a situation, we use the **Fisher Information Matrix,** which is described in the next section.

2.3 FISHER INFORMATION MATRIX

In the previous sections, you have studied how to calculate the Fisher information when we deal with a single parameter. The **Fisher information matrix** is a generalization of the Fisher Information for problems involving multiple parameters. It quantifies the amount of information that a set of observations provides about the parameters of a statistical model. When there are k parameters, then the Fisher information takes the form of an k × k matrix. Therefore, it is called the Fisher information matrix.

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population whose probability density (mass) function is $f(x, \theta_1, \theta_2, \ldots, \theta_k)$ where $\theta_1, \theta_2, \ldots, \theta_k$ are the population parameter, then the Fisher information matrix is defined as follows:

$$I(\theta) = nE\begin{bmatrix} \left(\frac{\partial}{\partial\theta_1}\log f\right)^2 & \left(\frac{\partial}{\partial\theta_1}\log f\right)\left(\frac{\partial}{\partial\theta_2}\log f\right) & \ldots & \left(\frac{\partial}{\partial\theta_1}\log f\right)\left(\frac{\partial}{\partial\theta_k}\log f\right) \\ \left(\frac{\partial}{\partial\theta_2}\log f\right)\left(\frac{\partial}{\partial\theta_1}\log f\right) & \left(\frac{\partial}{\partial\theta_2}\log f\right)^2 & \ldots & \left(\frac{\partial}{\partial\theta_2}\log f\right)\left(\frac{\partial}{\partial\theta_k}\log f\right) \\ \ldots & \ldots & \ldots & \ldots \\ \left(\frac{\partial}{\partial\theta_k}\log f\right)\left(\frac{\partial}{\partial\theta_1}\log f\right) & \left(\frac{\partial}{\partial\theta_k}\log f\right)\left(\frac{\partial}{\partial\theta_2}\log f\right) & \ldots & \left(\frac{\partial}{\partial\theta_k}\log f\right)^2 \end{bmatrix}$$

The Fisher Information Matrix I(θ) is a symmetric matrix. The **diagonal elements** of the Fisher information matrix represent the amount of Fisher information about individual parameters, whereas the **off-diagonal elements** represent the interaction or correlation between parameters.

If twice partially differentiable with respect to θ, and under certain additional regularity conditions such as the twice partially differentiate of $\log f(x;\theta)$ with respect to the parameters $\theta_1, \theta_2, \ldots, \theta_k$ exist, then we may write the Fisher information matrix may also be written as follows:

$$I(\theta) = -nE\begin{bmatrix} \frac{\partial^2}{\partial\theta_1^2}\log f & \frac{\partial^2}{\partial\theta_1\partial\theta_2}\log f & \ldots & \frac{\partial^2}{\partial\theta_1\partial\theta_k}\log f \\ \frac{\partial^2}{\partial\theta_2\partial\theta_1}\log f & \frac{\partial^2}{\partial\theta_2^2}\log f & \ldots & \frac{\partial^2}{\partial\theta_2\partial\theta_k}\log f \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial^2}{\partial\theta_k\partial\theta_1}\log f & \frac{\partial^2}{\partial\theta_k\partial\theta_2}\log f & \ldots & \frac{\partial^2}{\partial\theta_k^2}\log f \end{bmatrix}$$

Both definitions are equivalent under regularity conditions that allow interchanging differentiation and expectation.

The process of obtaining the Fisher information matrix in case of more than a single parameter is similar as discussed in the previous section. Here, we find $\left(\frac{\partial}{\partial\theta_i}\log f\right)\left(\frac{\partial}{\partial\theta_j}\log f\right)$ $i, j = 1, 2, \ldots k$ instead of single $\frac{\partial^2}{\partial\theta^2}\log f$.

Finding the Fisher information matrix is very complicated when there are more than two parameters. Do not bother about that, in this course, we consider the probability distribution up to two parameters only.

Let us explain the procedure of finding the Fisher information matrix when there are two parameters with the help of an example.

Example 2: A company produces ball bearings with diameters that follow a normal distribution with mean diameter of the ball bearings μ and variance in the diameter σ^2. To monitor the manufacturing process, the quality control team of the company selected 40 ball bearings from the manufacturing process and observed a standard deviation of 0.1mm.

(i) Calculate the Fisher information matrix for the unknown mean (μ) and variance (σ^2).

(ii) Interpret the Fisher information.

Solution: We know that the probability density function of the normal distribution with mean μ and variance σ^2 given as follows:

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}; \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

We take natural logarithmic on both sides of the above expression, then we get

$$\log f(x;\mu,\sigma^2) = \log f = \log 1 - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$$

There are two parameters μ and σ^2, therefore, to find the fisher information matrix, first, we partially differentiate the above expression with respect to the parameters μ and then σ^2, we get

$$\frac{\partial}{\partial\mu}(\log f) = 0-0-0-\frac{1}{2\sigma^2}2(-1)(x-\mu) = \frac{1}{\sigma^2}(x-\mu)$$

$$\frac{\partial}{\partial\sigma^2}(\log f) = 0-0-\frac{1}{2\sigma^2}-\frac{1}{2(\sigma^2)^2}(x-\mu)^2(-1)$$

$$= -\frac{1}{2\sigma^2}-\frac{1}{2(\sigma^2)^2}(x-\mu)^2$$

We now again partially differentiate the above expression with respect to the parameters μ and σ^2 to get the second derivative as follows:

$$\frac{\partial^2}{\partial\mu^2}(\log f) = \frac{1}{\sigma^2}(0-1) = -\frac{1}{\sigma^2}$$

$$\frac{\partial}{\partial(\sigma^2)^2}(\log f) = -\frac{1}{2(\sigma^2)^2}(-1)+\frac{1}{2(\sigma^2)^3}(-2)(x-\mu)^2$$

$$= \frac{1}{2(\sigma^2)^2}+\frac{1}{(\sigma^2)^3}(x-\mu)^2$$

$$\frac{\partial^2}{\partial\mu\partial\sigma^2}(\log f) = \frac{1}{(\sigma^2)^2}(-1)(x-\mu) = -\frac{1}{(\sigma^2)^2}(x-\mu)$$

Put the value of $\frac{\partial^2}{\partial\mu^2}(\log f), \frac{\partial}{\partial(\sigma^2)^2}(\log f)$ and $\frac{\partial^2}{\partial\mu\partial\sigma^2}(\log f)$ in the expression of the Fisher information matrix, we get

$$I(\theta) = -nE\begin{bmatrix} \frac{\partial^2}{\partial^2\mu}\log f & \frac{\partial^2}{\partial\mu\partial\sigma^2}\log f \\ \frac{\partial^2}{\partial\mu\partial\sigma^2}\log f & \frac{\partial^2}{\partial^2(\sigma^2)^2}\log f \end{bmatrix}$$

$$I(\theta) = -n\begin{bmatrix} -E\left[\frac{1}{\sigma^2}\right] & -E\left[\frac{1}{(\sigma^2)^2}(X-\mu)\right] \\ -E\left[\frac{1}{(\sigma^2)^2}(X-\mu)\right] & E\left[\frac{1}{2(\sigma^2)^2}-\frac{1}{(\sigma^2)^3}(X-\mu)^2\right] \end{bmatrix}$$

Diagonal Elements

$$E\left[\frac{\partial^2}{\partial\mu^2}\log f\right] = -E\left[\frac{1}{\sigma^2}\right] = -\frac{1}{\sigma^2}\ (\because E(a)=a)$$

$$E\left[\frac{\partial^2}{\partial(\sigma^2)^2}\log f\right] = E\left[\frac{1}{2\sigma^4}-\frac{1}{\sigma^6}(X-\mu)^2\right] = E\left[\frac{1}{2\sigma^4}\right]-\frac{1}{\sigma^6}E\left[(X-\mu)^2\right]$$

$$= \frac{1}{2\sigma^4}-\frac{1}{\sigma^6}\times\sigma^2\quad\left(\because E\left[(X-\mu)^2\right]=\sigma^2\right)$$

$$= \frac{1}{2\sigma^4} - \frac{1}{\sigma^4} = -\frac{1}{2\sigma^4}$$

Off-diagonal Elements

$$E\left[\frac{\partial^2}{\partial\mu\partial\sigma^2}\log f\right] = E\left[\frac{1}{\sigma^4}(X-\mu)\right] = \frac{1}{\sigma^4}\left[E[X]-E[\mu]\right]$$

$$= \frac{1}{\sigma^4}[\mu-\mu] = 0$$

Put the value of diagonal and off-diagonal elements in the expression of the Fisher information matrix, we get

$$I(\mu,\sigma^2) = -n\begin{bmatrix} -\frac{1}{\sigma^2} & 0 \\ 0 & -\frac{1}{2\sigma^4}\end{bmatrix} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4}\end{bmatrix}$$

We can interpret the above Fisher information matrix as:

The diagonal elements $\left(\frac{n}{\sigma^2}\right)$ and $\left(\frac{n}{2\sigma^4}\right)$ indicate the information about the parameters μ and σ^2, respectively. The off-diagonal elements 0 indicating the parameters μ and σ^2 are uncorrelated in terms of Fisher information.

We now calculate the Fisher information for n = 40 and $\sigma^2 = (0.1)^2 = 0.01$ as

$$I(\mu,\sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4}\end{bmatrix} = \begin{bmatrix} \frac{40}{0.01} & 0 \\ 0 & \frac{40}{2(0.01)^2}\end{bmatrix}$$

$$I(\mu,\sigma^2) = \begin{bmatrix} 4000 & 0 \\ 0 & 200000\end{bmatrix} = \begin{bmatrix} I_{\mu\mu} & I_{\mu\sigma^2} \\ I_{\mu\sigma^2} & I_{\sigma^2\sigma^2}\end{bmatrix}$$

Interpretation:

We can interpret the above Fisher information matrix as follows:

Diagonal Elements (Precision of Estimates):

- **Information about μ:** $I_{\mu\mu}$ = 4000 indicates that the sample provides a high amount of information about the mean diameter μ leading to a very precise estimate.
- **Information about σ²:** $I_{\sigma^2\sigma^2}$ = 200,000 is extremely large, indicating a very precise estimate of σ^2. This implies the sample size is highly effective at detecting inconsistencies in the manufacturing process.

Off-Diagonal Elements (Interaction Between Parameters):

- $I_{\mu\sigma^2}$ = 2600 is large, indicating a strong dependence between μ and σ^2. This means changes in μ significantly affect estimates of σ^2 and vice versa.

Practical Application:

The company can use these Fisher information values to design its sampling strategy. For example:

- If higher precision is needed for μ, e.g., maintaining a tighter tolerance on the mean diameter, they could increase the sample size n.
- Monitoring the variance σ^2 helps ensure consistency in the production process, which directly affects product quality and customer satisfaction.

Now, you can assess your understanding by answering the following Self Assessment Question.

SAQ 2

A manufacturing company monitors the failure times (in hours) of a specific machine component. The failure time follows a gamma distribution whose pdf is given as follows:

$$f(x,\alpha,\beta)=\frac{\beta^{\alpha}}{\lceil\alpha}e^{-\beta x}x^{\alpha-1};\quad x>0,\quad \alpha,\beta>0$$

where,

α (shape parameter) represents the number of failure-causing events.

β (rate parameter) represents the failure rate.

A sample of n = 40 components is tested, and the total failure time recorded is 2600 hours. If α = 2, then estimate the Fisher Information Matrix for (α, β) and interpret the results.

When we estimate an unknown parameter, then we require that the estimate should be accurate and reliable, which means the estimator should be uniformly minimum variance unbiased estimator. One way to find a uniform minimum variance unbiased estimator for a parameter is to use the Cramér-Rao inequality. Let us study it in the next section.

(1893-1985)

Harald Cramér was a Swedish mathematician. He made significant contributions to probability theory, statistics, and the theory of stochastic processes. He developed many fundamental results in statistics, including the concept of Fisher information.

2.4 CRAMÉR-RAO INEQUALITY

The **Cramér-Rao inequality** is a fundamental result in the field of estimation theory and statistics. It provides a lower bound of the variance of an unbiased estimator of a parameter. The Cramér-Rao inequality states that for any unbiased estimator of a parameter, say, θ, the variance cannot be smaller than the inverse of the Fisher Information. This inequality is particularly useful because it establishes a benchmark to evaluate the efficiency of estimators.

The Cramér-Rao inequality is named after two mathematicians: **Harald Cramér** and **Calyampudi Radhakrishna Rao (C.R. Rao)**. They independently gave this inequality. In 1946, Cramér introduced the inequality in the context of unbiased estimators and the Fisher information, whereas Rao, in 1945, connected the lower bound of variance to the concept of Fisher information and explored its implications for multivariate parameter estimation. Let us discuss the Cramér-Rao inequality.

Statement of the Cramér-Rao Inequality

Suppose $X_1, X_2, ..., X_n$ is a random sample of size n taken from a population with probability density $f(x;\theta)$, where θ is a parameter. If T is an unbiased estimator of the parameter θ, then the variance of the estimator T satisfies

$$\text{Var}(T) \geq \frac{1}{nE\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]} = \frac{1}{I(\theta)} \quad \text{for all values of } \theta$$

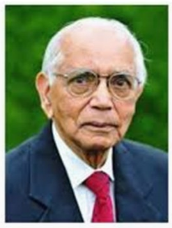
Equivalently

$$\text{Var}(T) \geq \frac{1}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]} \quad \text{for all values of } \theta$$

The inequality is held under some regularity conditions.

The denominator of the inequality is called the **Fisher information,** which measures the amount of information that a random sample carries about an unknown parameter θ. The above inequality is called **Cramér-Rao inequality**, and the right-hand side is called the **Cramér-Rao lower bound** for the variance of the unbiased estimator of the parameter θ.

If an unbiased estimator T whose variance coincides with the Carmer-Rao lower bound of variance, then this estimator will be a uniform minimum variance unbiased estimator (UMVUE) for the parameter θ. The converse of this is not true. It means that not all uniform minimum variance unbiased estimators of a parameter satisfy the Cramér-Rao lower bound. The Cramér-Rao lower bound is based on some conditions. Let us study them.



(1920–2023)

An Indian statistician, Rao is renowned for his pioneering work in statistical theory and applied statistics. His contributions include the development of the Rao-Blackwell theorem, Rao's score test, and the Cramér-Rao inequality.

Regularity Conditions of Cramér-Rao Inequality

In order to assure the validity of the lower bound and the well-defined Fisher information, the Cramér-Rao inequality depends on some specific assumptions which ensure the mathematical validity of the derivations and permit the exchange of differentiation and integration procedures. These assumptions are called **regularity conditions** of the Cramér-Rao inequality. The regularity conditions are as follows:

1. The probability density (or mass) function $f(x;\theta)$ must be **smooth** with respect to the parameter θ so that the first and second partial derivatives of $f(x;\theta)$ with respect to θ must exist and be well-defined.

2. The range of the probability density (or mass) function $f(x;\theta)$ must be independent of the parameter θ. In mathematical terms, the region of integration should be constant with respect to θ. Thus, the $U(0, \theta)$ distribution, for example, is left out.

3. The **score function**, $\frac{\partial}{\partial\theta}\log f(x;\theta)$, must be integrable with respect to the distribution $f(x;\theta)$. This ensures the expectations involving the score function are well-defined.

4. For continuous random variables, the differentiation of the expected value with respect to parameter θ must be valid under the integral sign, that is,

$$\frac{\partial}{\partial\theta}E\left[g(X)\right]=E\left[\frac{\partial}{\partial\theta}g(X)\right]$$

This condition allows the derivation of the Fisher information and ensures the Cramér-Rao inequality holds.

5. The estimator T must be unbiased for the parameter θ, that is, $E\left[T\right]=\theta$ for all θ.

6. The Fisher information must be finite, that is,

$$0<nE\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]<\infty.$$

These requirements are met in the majority of real-world issues; nevertheless, if any of them are broken, the Cramér-Rao inequality may not hold.

Let us discuss how to obtain the lower bound of variance of an unbiased estimator using Cramér-Rao inequality with the help of some examples.

Example 3: Suppose a multi-brand electronic store records the arrival times of customers, and the time between consecutive arrivals follows an exponential distribution whose probability density function is given as follows:

$$f(x,\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}};\quad x>0,\ \ \theta>0$$

where θ represents the average arrival rate (customers per hour).

To estimate the average arrival rate, the market analyst of the store randomly noted the arrival times of n customers, then

(i) What is the Cramér-Rao lower bound for the variance of the unbiased estimator of the parameter θ?

(ii) Show that the sample mean is the UMVUE of the mean waiting time between customers.

(iii) If the analyst observed 10 customer arrivals, and the true arrival rate is θ = 5 customers per hour, then find the Cramér-Rao lower bond and interpret the results.

Solution: The probability density function of the given exponential distribution with parameter θ is given as follows:

$$f(x,\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}};\quad x>0,\ \ \theta>0$$

Let T be an unbiased estimator of θ. Cramér-Rao lower bound for the variance of the estimator T is given by

$$\mathrm{Var}(T)\geq\frac{1}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]}$$

To find the Cramér-Rao lower bound for the variance, we first compute the Fisher information as discussed in Section 2.2. Therefore, we take natural logarithmic on both sides of the above expression, then we get

$$\log f(x;\theta) = -\log(\theta) - \frac{x}{\theta}$$

We partially differentiate the above expression with respect to the parameter θ, we get

$$\frac{\partial}{\partial \theta}\log f(x;\theta) = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

We now again partially differentiate the above expression with respect to the parameter θ to get the second derivative as follows:

$$\frac{\partial^2}{\partial \theta^2}\log f(x;\theta) = \frac{1}{\theta^2} - \frac{2x}{\theta^3} \quad \left[\because \frac{\partial}{\partial x}\left(\frac{1}{x^n}\right) = -\frac{n}{x^{n+1}}\right]$$

Put the value of $\frac{\partial^2}{\partial \theta^2}\log f(x;\theta)$ in the expression of the Fisher information, we get

$$I(\theta) = -nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right] = -nE\left[\frac{1}{\theta^2} - \frac{2X}{\theta^3}\right]$$

$$= -nE\left[\frac{1}{\theta^2}\right] + \frac{2n}{\theta^3}E[X] = -\frac{n}{\theta^2} + \frac{2n}{\theta^3}\times\theta\left(\because E[X] = \theta\right)$$

$$= -\frac{n}{\theta^2} + \frac{2n}{\theta^3}\times\theta = \frac{n}{\theta^2}$$

Hence, by putting the Fisher information in the Cramér-Rao inequality, we get the lower bound for variance as

$$\text{Var}(T) \geq \frac{1}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]}$$

$$\text{Var}(T) \geq \frac{\theta^2}{n}$$

To show that the sample mean is the UMVUE of the mean time between consecutive customers using Cramér-Rao inequality, first, we have to show

(i) It is an unbiased estimator of the parameter θ, that is,

$$E(T) = \theta$$

(ii) Find the variance of the estimator. If the variance of the estimator $\bar{X}$ is equal to the lower bound of the variance, then the estimator is the UMVUE.

Thus, we consider,

$$E(\bar{X}) = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] \quad \text{[By the definition of the sample mean]}$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right] \quad \left[\because E(aX + bY) = aE(X) + bE(Y)\right]$$

Since $X_1, X_2, \ldots, X_n$ are randomly drawn from the same population (exponential distribution) whose mean is θ and variance θ^2, therefore,

$$E(X_1) = E(X_2) = ... = E(X_n) = E(X) = \theta$$

Thus,

$$E(\bar{X}) = \frac{1}{n}\left(\underbrace{\theta + \theta + ... + \theta}_{n-\text{times}}\right) = \frac{1}{n}(n\theta) = \theta$$

Hence, the sample mean $(\bar{X})$ is an unbiased estimator of the population parameter θ.

We now find the variance of the estimator $(\bar{X})$ as

If X and Y are two independent random variables, then

Var(aX±bY) = a^2 Var(X) + b^2 Var(Y).

$$Var(\bar{X}) = Var\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right]$$

$$= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + ... + Var(X_n)\right]$$

$$= \frac{1}{n^2}\left(\underbrace{\theta^2 + \theta^2 + ... + \theta^2}_{n\text{-times}}\right)$$

$$Var(\bar{X}) = \frac{1}{n^2}(n\theta^2) = \frac{\theta^2}{n}$$

Since the variance of the estimator sample mean ($\bar{X}$) is equal to the lower bound of the variance, therefore, the sample mean ($\bar{X}$) is the UMVUE of the parameter θ.

Here, n = 10 and θ = 5 then we can calculate the Cramér-Rao lower bound as

$$Var(T) \geq \frac{\theta^2}{n} = \frac{25}{10} = 2.5$$

The Cramér-Rao lower bound tells us that the lowest possible variance for an unbiased estimator of θ is 2.5, and our estimator sample mean attains this bound, therefore, it is the best possible estimator (efficient). If another estimator had a variance greater than 2.5, it would be suboptimal, and if an estimator had a variance lower than 2.5, it must be biased.

Example 4: If the number of patients arriving at an emergency room of a hospital per day follows Poisson distribution λ. To estimate the number of patients, a medical researcher of the hospital counted the number of patients who arrived at the emergency room on randomly selected days. If they are denoted by $X_1, X_2, \ldots, X_n$, then

(i) What is the Cramér-Rao lower bound for the variance of the unbiased estimator of the parameter λ?

(ii) Show that the sample mean $(\bar{X})$ is the UMVUE of the parameter λ.

Solution: The probability mass function of the Poisson distribution with parameter λ is given as follows:

$$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}; \quad x = 0, 1, 2, ... \;\; \& \lambda > 0$$

Let T be an unbiased estimator of λ. Cramér-Rao lower bound for the variance of the estimator T is given by

$$\text{Var}(T) \geq \frac{1}{-nE\left[\frac{\partial^2}{\partial \lambda^2}\log f(x;\lambda)\right]}$$

To find the Cramér-Rao lower bound for the variance, we first compute the Fisher information as discussed in Section 2.2. Therefore, we take natural logarithmic on both sides of the probability mass function, then we get

$$\log p(x;\lambda) = -\lambda + x\log(\lambda) - \log(x!)$$

We partially differentiate the above expression with respect to the parameter λ, we get

$$\frac{\partial}{\partial \lambda}\log p(x;\lambda) = -1 + \frac{x}{\lambda}$$

We now again partially differentiate the above expression with respect to the parameter θ to get the second derivative as follows:

$$\frac{\partial^2}{\partial \lambda^2}\log f(x;\lambda) = -\frac{x}{\lambda^2} \quad \left[\because \frac{\partial}{\partial x}\left(\frac{1}{x^n}\right) = -\frac{n}{x^{n+1}}\right]$$

Put the value of $\frac{\partial^2}{\partial \theta^2}\log f(x;\lambda)$ in the expression of the Fisher information, we get

$$I(\theta) = -nE\left[\frac{\partial^2}{\partial \lambda^2}\log f(x;\lambda)\right] = -nE\left[-\frac{X}{\lambda^2}\right]$$

$$= \frac{n}{\lambda^2}E[X] = \frac{n}{\lambda^2} \times \lambda = \frac{n}{\lambda}$$

Hence, by putting the Fisher information in the Cramér-Rao inequality, we get the lower bound for variance as

$$\text{Var}(T) \geq \frac{1}{-nE\left[\frac{\partial^2}{\partial \lambda^2}\log f(x;\lambda)\right]}$$

$$\text{Var}(T) \geq \frac{\lambda}{n}$$

To show that the sample mean is the UMVUE of the average number of the patients using Cramér-Rao inequality, first, we have to show

(iii) It is an unbiased estimator of the parameter λ, that is,

$$E[T] = \lambda$$

(iv) Find the variance of the estimator. If the variance of the estimator $\overline{X}$ is equal to the lower bound of the variance then the estimator is the UMVUE.

Thus, we consider,

$$E(\overline{X}) = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] \qquad \text{[By the definition of the sample mean]}$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] \quad \left[\because E(aX + bY) = aE(X) + bE(Y)\right]$$

Since $X_1, X_2, \ldots, X_n$ are randomly drawn from the same population (Poisson distribution) whose mean and variance are the same λ, therefore,

$$E(X_1) = E(X_2) = ... = E(X_n) = E(X) = \lambda$$

Thus,

$$E(T) = \frac{1}{n}\left(\underbrace{\lambda + \lambda + ... + \lambda}_{n-\text{times}}\right) = \frac{1}{n}(n\lambda) = \lambda$$

Hence, the sample mean $(\bar{X})$ is an unbiased estimator of the population parameter λ.

We now find the variance of the estimator $(\bar{X})$ as

> If X and Y are two independent random variables, then
>
> $Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y)$.

$$Var(\bar{X}) = Var\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right]$$

$$= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + ... + Var(X_n)\right]$$

$$= \frac{1}{n^2}\left(\underbrace{\lambda + \lambda + ... + \lambda}_{n\text{-times}}\right) = \frac{1}{n^2}(n\lambda)$$

$$Var(\bar{X}) = \frac{\lambda}{n}$$

Since the variance of the estimator $\bar{X}$ is equal to the lower bound of the variance, therefore, the sample mean is the UMVUE of the parameter λ.

I think you understand how to apply the Cramér-Rao inequality to find the UMVUE estimator of a parameter. You may have the curiosity to apply this inequality yourself. For that, you can try the following Self Assessment Question.

SAQ 3

Suppose a sensor measures the temperature in a room, and the readings follow a normal distribution with unknown mean temperature μ and known variance σ^2. A statistical analyst is interested in obtaining an efficient estimator of the mean temperature. The analyst randomly noted the reading of the room temperature n times, say $X_1, X_2, \ldots X_n$. Show that the sample mean is the UMVUE for the mean temperature of the room using the Cramér-Rao inequality.

After understanding the Cramér-Rao lower bound for variance of an unbiased estimator, let us study some limitations of the Cramér-Rao inequality in the next section.

2.5 LIMITATION OF CRAMÉR-RAO INEQUALITY

In the previous section, you studied the Cramér-Rao inequality and how you can find the best estimator with its help. The Cramér-Rao inequality has some limitations, some of them are listed as follows:

The main limitations of the Cramér-Rao lower bound approach are given as follows:

1. The Cramér-Rao inequality applies only to unbiased estimators. In many situations, estimators are slightly biased but asymptotically unbiased as the sample size grows.
2. The assumptions of Cramér-Rao inequality are not satisfied by every density function $f(x; \theta)$ such as uniform.
3. All the unbiased estimators do not attain the Cramér-Rao lower bound.

Hence, in any one of these situations, one does not know whether an estimator is a uniform minimum variance unbiased estimator or not.

There is an alternative way of looking for UMVU estimates when the approach by means of the Cramér–Rao inequality fails to produce such an estimate. This approach hinges heavily on the concept of sufficiency and completeness. We will discuss these alternatives in the next unit.

You can try the following Self Assessment Question before moving to the next section.

SAQ 4

Write the main limitations of the Cramér-Rao inequality.

We now end this unit by giving a summary of what we have covered in it.

2.6 SUMMARY

In this unit, we have covered the following points:

- Fisher information quantifies the amount of information a random variable carries about an unknown parameter of a probability distribution.
- We can calculate the Fisher information that a sample carries as follows:

$$I(\theta) = -nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]$$

- The Fisher information matrix is a generalization of the Fisher information for problems involving multiple parameters.
- The Cramér-Rao inequality provides a lower bound of the variance of an unbiased estimator of a parameter.
- According to the Cramér-Rao inequality, if T is an unbiased estimator of the parameter θ, then

$$\mathrm{Var}(T) \geq \frac{1}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]} \quad \text{for all values of } \theta$$

2.7 TERMINAL QUESTIONS

A meteorologist wants to estimate the probability of rain on a given day in a specific month so that they advise people to carry umbrellas or schedule outdoor events based on the estimated probability of rain. If the probability of rain on a given day is modelled by the Beta(θ,1) whose probability density function is given by

$$f(x;\theta) = \theta x^{\theta-1}; \quad x > 0, \quad \theta > 0$$

where θ reflects the number of observed rainy days.

(i) What is the Cramér-Rao lower bound for the variance of the unbiased estimator of the parameter θ?

(ii) If the meteorologist collected data for 20 days and the true number of observed rainy days is θ = 16 then find the Cramér-Rao lower bond and interpret the results.

2.8 SOLUTIONS / ANSWERS

Self Assessment Questions (SAQs)

1. To find the Fisher information, we follow the same procedure as discussed in Example 1. That is, first of all, we write the probability mass function of the distribution.

The probability mass function of the Bernoulli distribution with parameter p is given as follows:

$$P[X = x] = p(x) = p^x (1-p)^{1-x}; \quad x = 0,1$$

We now take natural logarithmic on both sides of the above expression, then we get

$$\log f(x;p) = \log f = x\log(p) + (1-x)\log(1-p)$$

We partially differentiate the above expression with respect to the parameter p, we get

$$\frac{\partial}{\partial p}\log f(x;p) = x\left(\frac{1}{p}\right) + (1-x)\left(\frac{1}{1-p}\right)(-1) = x\left(\frac{1}{p}\right) - (1-x)\left(\frac{1}{1-p}\right)$$

We now again partially differentiate the above expression with respect to the parameter p to get the second derivative as follows:

$$\frac{\partial^2}{\partial p^2}\log f(x;p) = -\frac{x}{p^2} - \frac{(1-x)}{(1-p)^2} \quad \left[\because \frac{\partial}{\partial x}\left(\frac{1}{x}\right) = -\frac{1}{x^2}\right]$$

Put the value of $\frac{\partial^2}{\partial \theta^2}\log f(x;\theta)$ in the expression of the Fisher information, we get

$$I(p) = -nE\left[\frac{\partial^2}{\partial p^2}\log f(x;p)\right] = -nE\left[-\frac{x}{p^2} - \frac{(1-x)}{(1-p)^2}\right]$$

$$= n\left[\frac{1}{p^2}E[X] + \frac{1}{(1-p)^2}E[1-X]\right]$$

$$= n\left[\frac{1}{p^2}p + \frac{1}{(1-p)^2}(1-p)\right]$$

$$= n\left[\frac{1}{p} + \frac{1}{(1-p)}\right] = n\left(\frac{1-p+p}{p(1-p)}\right)$$

$$I(p) = \frac{n}{p(1-p)}$$

Hence, the required Fisher information is $I(p) = \frac{n}{p(1-p)}$.

We now estimate the probability p that a learner has their own laptop on the basis of sample information. Since p is the mean of Bernoulli distribution (population mean) so we estimate it by the sample mean as follows:

1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1+1+1+0+1+1+1+1+1+0+1+1+1+1+1+1}{16} = 0.875$$

We now calculate the Fisher information for n = 12 and $\hat{p} = 0.75$ as

$$I(\theta) = \frac{n}{p(1-p)} = \frac{16}{0.875(1-0.875)} = \frac{16}{0.109} = 146.789$$

We can calculate the variance of the estimator as

$$Var(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

For $I(\theta)$ = 146.789

$$Var(\hat{\theta}) \geq \frac{1}{146.789} = 0.0068$$

A very small variance indicates that the sample provides a high degree of information about the parameter p and we get a reliable and consistent estimate of p.

2. To solve this question, first, we find the Fisher information matrix, and then we estimate it using the given information.

We know that the probability density function of the gamma distribution with shape parameter α and rate parameter β is given as follows:

$$f(x, \alpha, \beta) = \frac{\beta^{\alpha}}{\overline{|\alpha}} e^{-\beta x} x^{\alpha-1}; \quad 0 < x < \infty, \alpha, \beta > 0$$

We take natural logarithmic on both sides of the above expression, then we get

$$\log f(x, \alpha, \beta) = \log f = \alpha \log(\beta) - \log\left(\overline{|\alpha}\right) - \beta x + (\alpha - 1)\log(x)$$

There are two parameters α and β, therefore, to find the Fisher information matrix, first, we partially differentiate the above expression with respect to the parameters α and then β, we get

$$\frac{\partial}{\partial \alpha}(\log f) = \log(\beta) - \psi(\alpha) + \log(x) \quad \left[\because \frac{\partial}{\partial \alpha}(\log \alpha) = \psi\left(\overline{|\alpha}\right)\right]$$

where $\psi(\alpha) = \frac{\overline{|\alpha}'}{\overline{|\alpha}}$ is the digamma function.

Similarly,

$$\frac{\partial}{\partial\beta}(\log f) = \frac{\alpha}{\beta} - x$$

We now again partially differentiate the above expression with respect to the parameters α and β to get the second derivative as follows:

$$\frac{\partial^2}{\partial\alpha^2}(\log f) = 0 - \psi'(\alpha) + 0 = -\psi'\left(\overline{|\alpha}\right)$$

$$\frac{\partial^2}{\partial\beta^2}(\log f) = -\frac{\alpha}{\beta^2}$$

$$\frac{\partial^2}{\partial\alpha\partial\beta}(\log f) = \frac{1}{\beta}$$

Put the value of $\frac{\partial^2}{\partial\alpha^2}(\log f)$, $\frac{\partial^2}{\partial\beta^2}(\log f)$ and $\frac{\partial^2}{\partial\alpha\partial\beta}(\log f)$ in the expression of the Fisher information matrix, we get

$$I(\alpha,\beta) = -nE\begin{bmatrix} \frac{\partial^2}{\partial\alpha^2}(\log f) & \frac{\partial^2}{\partial\alpha\partial\beta}(\log f) \\ \frac{\partial^2}{\partial\alpha\partial\beta}(\log f) & \frac{\partial^2}{\partial\alpha\partial\beta}(\log f) \end{bmatrix}$$

$$I(\alpha,\beta) = -nE\begin{bmatrix} -\psi'(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & -\frac{\alpha}{\beta^2} \end{bmatrix} = \begin{bmatrix} n\psi'(\alpha) & -\frac{n}{\beta} \\ -\frac{n}{\beta} & \frac{n\alpha}{\beta^2} \end{bmatrix} \quad \left(\because E\left[\psi'(\alpha)\right] = \psi'(\alpha), \; E\left[\frac{1}{\beta}\right] = \frac{1}{\beta}, E\left[\frac{\alpha}{\beta^2}\right] = \frac{\alpha}{\beta^2}\right)$$

We now estimate the Fisher information matrix. It is given that n = 40, α = 2, total failure time = 2600 hours.

Therefore, the average failure time per component is 2600/40 = 65 hours.

The average failure time per component of gamma distribution = $\frac{1}{\beta} = 65$

We can calculate the tri-gamma function $\psi'(\alpha)$ as

$$\psi'(\alpha) = \frac{1}{\alpha^2} + \frac{1}{(\alpha+1)^2} + \frac{1}{(\alpha+1)^3} + \ldots = \frac{1}{2^2} + \frac{1}{(2+1)^2} + \frac{1}{(2+1)^3} + \ldots$$

$$\psi'(2) \approx 0.25 + 0.111 + 0.065 = 0.426$$

Now, we compute the Fisher Information elements as follows:

$$I(\alpha,\beta) = \begin{bmatrix} n\psi'(\alpha) & -\frac{n}{\beta} \\ -\frac{n}{\beta} & \frac{n\alpha}{\beta^2} \end{bmatrix} = \begin{bmatrix} 40 \times 0.426 & -40 \times 65 \\ -40 \times 65 & 40 \times 2 \times 65^2 \end{bmatrix}$$

$$I(\alpha,\beta)=\begin{bmatrix}17.04 & -2600\\ -2600 & 338000\end{bmatrix}=\begin{bmatrix}I_{\alpha\alpha} & I_{\alpha\beta}\\ I_{\alpha\beta} & I_{\beta\beta}\end{bmatrix}$$

We can interpret the above Fisher information matrix as follows:

Diagonal Elements (Precision of Estimates):

- $I_{\alpha\alpha}$ =17.04 suggests moderate precision in estimating α.
- $I_{\beta\beta}$ = 338000 is extremely large, indicating a very precise estimate of β. This is expected because failure rate estimates are highly sensitive to large amounts of data.

Off-Diagonal Elements (Interaction Between Parameters):

- $I_{\alpha\beta}$ = 2600 is large, indicating a strong dependence between α and β. This means changes in α significantly affect estimates of β and vice versa.

3. To show that the sample mean is the UMVUE using the Cramér-Rao inequality,

(i) first, we have to find the lower bound of the variance using the Cramér-Rao inequality.

(ii) Check whether the sample mean is an unbiased estimator of the parameter μ, that is,

$$E(\bar{X})=\mu$$

(iii) Find the variance of the estimator. If the variance of the estimator $\bar{X}$ is equal to the lower bound of the variance, then the estimator is the UMVUE.

Therefore, we first find the Cramér-Rao lower bound.

We know that the probability density function of the normal distribution with mean μ and known variance σ^2 which is given as follows:

$$f(x;\mu)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2};\quad -\infty<x<\infty,\ -\infty<\mu<\infty$$

We take natural logarithmic on both sides of the above expression, then we get

$$\log f(x;\mu)=\log 1-\frac{1}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}(x-\mu)^2$$

We partially differentiate the above expression with respect to the parameters μ as follows:

$$\frac{\partial}{\partial\mu}\log f(x;\mu)=0-0-\frac{1}{2\sigma^2}2(x-\mu)(-1)=\frac{1}{\sigma^2}(x-\mu)$$

We now again partially differentiate the above expression with respect to the parameter μ to get the second derivative as follows:

$$\frac{\partial^2}{\partial\mu^2}\log f(x;\mu)=-\frac{1}{\sigma^2}$$

Put the value of $\frac{\partial^2}{\partial \mu^2}\log f(x;\mu)$ in the expression of the Fisher information, we get

$$I(\theta) = -nE\left[\frac{\partial^2}{\partial \mu^2}\log f(x;\mu)\right]$$

$$= -nE\left[-\frac{1}{\sigma^2}\right] = \frac{n}{\sigma^2}$$

Hence, by putting the Fisher information in the Cramér-Rao inequality, we get the lower bound for variance as

$$Var(\bar{X}) \geq \frac{1}{-nE\left[\frac{\partial^2}{\partial \mu^2}\log f(x;\mu)\right]}$$

$$Var(T) \geq \frac{\sigma^2}{n}$$

We now show that the sample mean $(\bar{X})$ is an unbiased estimator of the population parameter μ. Thus, we consider,

$$E(\bar{X}) = E\left[\frac{X_1 + X_2 + ... + X_n}{n}\right] \quad \text{[By the definition of the sample mean]}$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] \quad \left[\because E(aX + bY) = aE(X) + bE(Y)\right]$$

Since X_1, X_2,…, X_n are randomly drawn from the same population (normal distribution) whose mean is μ and variance 25, therefore,

$$E(X_1) = E(X_2) = ... = E(X_n) = E(X) = \mu$$

Thus,

$$E(\bar{X}) = \frac{1}{n}\Big(\underbrace{\mu + \mu + ... + \mu}_{n-times}\Big)$$

$$= \frac{1}{n}(n\mu) = \mu$$

Hence, the sample mean $(\bar{X})$ is an unbiased estimator of the population parameter μ.

We now find the variance of the estimator $(\bar{X})$ as

> If X and Y are two independent random variables, then
>
> Var(aX±bY) = a^2 Var(X) + b^2 Var(Y).

$$Var(\bar{X}) = Var\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right]$$

$$= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + ... + Var(X_n)\right]$$

$$= \frac{1}{n^2}\Big(\underbrace{\sigma^2 + \sigma^2 + ... + \sigma^2}_{n\text{-times}}\Big)$$

$$= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

Since the variance of the estimator $\bar{X}$ is equal to the lower bound of the variance, therefore, the sample mean $\bar{X}$ is the UMVUE of the parameter μ.

Terminal Questions (TQs)

The probability density function of the given beta distribution with parameter θ is given as follows:

$$f(x;\theta) = \theta x^{\theta-1}; \quad x > 0, \quad \theta > 0$$

Let T be an unbiased estimator of θ. The Cramér-Rao lower bound for the variance of the estimator T is given by

$$\text{Var}(T) \geq \frac{1}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]}$$

To find the Cramér-Rao lower bound for the variance, we first compute the Fisher information. Therefore, we take natural logarithmic on both sides of the above expression, then we get

$$\log f(x;\theta) = \log(\theta) + (\theta - 1)\log(x)$$

We partially differentiate the above expression with respect to the parameter θ, we get

$$\frac{\partial}{\partial\theta}\log f(x;\theta) = \frac{1}{\theta} + \log(x)$$

We now again partially differentiate the above expression with respect to the parameter θ to get the second derivative as follows:

$$\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) = -\frac{1}{\theta^2} \quad \left[\because \frac{\partial}{\partial x}\left(\frac{1}{x^n}\right) = -\frac{n}{x^{n+1}}\right]$$

Put the value of $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)$ in the expression of the Fisher information, we get

$$I(\theta) = -nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right] = -nE\left[-\frac{1}{\theta^2}\right] = \frac{n}{\theta^2}$$

Hence, by putting the Fisher information in the Cramér-Rao inequality, we get the lower bound for variance as

$$\text{Var}(T) \geq \frac{1}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]}$$

$$\text{Var}(T) \geq \frac{\theta^2}{n}$$

Here n = 20 and θ = 16 then we can calculate the Cramér-Rao lower bound as

$$\text{Var}(T) \geq \frac{\theta^2}{n} = \frac{16^2}{20} = 21.33$$

The Cramér-Rao lower bound tells us that the lowest possible variance for an unbiased estimator of θ is 21.33.