# 4

# The Fisher Information and the Cramér-Rao Inequality

Suppose that $Y_1, \dots, Y_n$ is a random sample from a population having a common density function $f(y|\theta)$ depending on a parameter $\theta$, the estimation of which is desired. We have already seen a number of ways of measuring the goodness of an estimator $\hat{\theta}$ of $\theta$. In general, we prefer unbiased estimators with as small a variance as possible. In fact, we would ideally like to find the *minimum variance unbiased estimator* (or MVUE) of $\theta$.

The Cramér-Rao inequality gives a lower bound on the variance of any unbiased estimator. In fact, this is how we can determine if an estimator $\hat{\theta}$ is truly the MVUE of $\theta$. If $\text{Var}(\hat{\theta})$ attains the Cramér-Rao lower bound, then it must be the MVUE.

In order to state the Cramér-Rao inequality, we must first introduce the Fisher information.

**Definition 4.1.** If the random variable $Y$ has density function $f(y|\theta)$, then the *Fisher information* is defined as

$$I(\theta) := -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(Y|\theta)\right).$$

**Remark.** It is important to pay attention to the symbols: we are starting with the density function $f(y|\theta)$, and then taking its log, the second derivative with respect to $\theta$, evaluating this expression *at the random variable* $Y$, and finally taking the resulting expectation.

**Remark.** Although we have not stated it explicitly in the definition of the Fisher information, some assumptions on the smoothness (continuity and differentiability) of the density $f(y|\theta)$ as a function of $\theta$ are required. We will not focus too much on such technicalities in Stat 252. It is perhaps worth noting that the definition of the Fisher information involves the second derivative with respect to $\theta$ of $\log f(y|\theta)$ which implicitly assumes that $\log f(y|\theta)$ is twice-differentiable!

**Example 4.2.** Suppose that $Y \sim \text{Exp}(\theta)$ for some parameter $\theta > 0$. Calculate $I(\theta)$.

**Solution.** The density of $Y$ is

$$f(y|\theta) = \frac{1}{\theta} e^{-y/\theta}, \quad y > 0.$$

Therefore,

$$\log f(y|\theta) = -\log \theta - \frac{y}{\theta}$$

so that

$$\frac{\partial}{\partial \theta} \log f(y|\theta) = -\frac{1}{\theta} + \frac{y}{\theta^2} \quad \text{and} \quad \frac{\partial^2}{\partial \theta^2} \log f(y|\theta) = \frac{1}{\theta^2} - \frac{2y}{\theta^3}.$$

We now find the Fisher information is

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log f(Y|\theta)\right) = \mathbb{E}\left(\frac{2Y}{\theta^3} - \frac{1}{\theta^2}\right) = \frac{2\mathbb{E}(Y)}{\theta^3} - \frac{1}{\theta^2} = \frac{2\theta}{\theta^3} - \frac{1}{\theta^2} = \frac{1}{\theta^2}.$$

**Example 4.3.** Suppose that $Y \sim \mathcal{N}(\theta, \sigma^2)$ where $\theta \in \mathbb{R}$ is a parameter and $\sigma^2$ is known. Calculate $I(\theta)$.

**Solution.** The density of $Y$ is

$$f(y|\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\theta)^2}{2\sigma^2}\right\}, \quad -\infty < y < \infty.$$

Note that the density only depends on $\theta$ since we are assuming that $\sigma^2$ is known. Therefore,

$$\log f(y|\theta) = -\log(\sigma\sqrt{2\pi}) - \frac{(y-\theta)^2}{2\sigma^2}$$

so that

$$\frac{\partial}{\partial \theta} \log f(y|\theta) = \frac{y-\theta}{\sigma^2} \quad \text{and} \quad \frac{\partial^2}{\partial \theta^2} \log f(y|\theta) = -\frac{1}{\sigma^2}.$$

We now find the Fisher information is

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log f(Y|\theta)\right) = -\mathbb{E}\left(-\frac{1}{\sigma^2}\right) = \frac{1}{\sigma^2}.$$

We can now state the Cramér-Rao inequality.

**Theorem 4.4** (Cramér-Rao Inequality)**.** *Suppose that $Y_1, \ldots, Y_n$ is a random sample from a population having a common density function $f(y|\theta)$ depending on a parameter $\theta$, and let $\hat{\theta}$ be an unbiased estimator of $\theta$ based on $Y_1, \ldots, Y_n$. If $f(y|\theta)$ is a smooth function of $y$ and $\theta$, then*

$$\operatorname{Var}(\hat{\theta}) \ge \frac{1}{nI(\theta)}.$$

*where the Fisher information $I(\theta)$ is given by*

$$I(\theta) := -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log f(Y|\theta)\right)$$

*for some random variable $Y$ having density $f(y|\theta)$.*

**Remark.** The importance of this result is the following. Suppose that you are able to find some estimator $\hat{\theta}$ whose variance is $\mathrm{Var}(\hat{\theta}) = \frac{1}{nI(\theta)}$. Since you have found the estimator whose variance attends the lower bound of the Cramér-Rao inequality your search for the minimum variance unbiased estimator is over! It is not possible to find another unbiased estimator with smaller variance.

**Example 4.5.** Suppose that $Y_1, \ldots, Y_n$ is a random sample from an $\mathrm{Exp}(\theta)$ population depending on a parameter $\theta > 0$. Prove that $\hat{\theta} := \overline{Y}$ is the minimum variance unbiased estimator of $\theta$.

**Solution.** Since $Y_i \sim \mathrm{Exp}(\theta)$, we know that $\mathbb{E}(Y_i) = \theta$ and $\mathrm{Var}(Y_i) = \theta^2$ for all $i$. Therefore, as in Theorem 2.8, if $\hat{\theta} := \overline{Y}$, then $\mathbb{E}(\hat{\theta}) = \mathbb{E}(Y_1) = \theta$ so that $\hat{\theta}$ is, in fact, an unbiased estimator of $\theta$. Moreover, since $Y_1, \ldots, Y_n$ are i.i.d., we also conclude using Theorem 2.8 that

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{n}\,\mathrm{Var}(Y_1) = \frac{\theta^2}{n}.$$

In Example 4.2 above we determined that $I(\theta) = \theta^{-2}$. Therefore,

$$\frac{1}{nI(\theta)} = \frac{1}{n\theta^{-2}} = \frac{\theta^2}{n} = \mathrm{Var}(\hat{\theta}).$$

Since the lower bound of the Cramér-Rao inequality has been attained, we conclude that $\overline{Y}$ must be the minimum variance unbiased estimator of $\theta$.

**Example 4.6.** Suppose that $Y_1, \ldots, Y_n$ is a random sample from a $\mathcal{N}(\theta, \sigma^2)$ where $\sigma^2$ is known but $\theta \in \mathbb{R}$ is a parameter. Prove that $\hat{\theta} := \overline{Y}$ is the minimum variance unbiased estimator of $\theta$.

**Solution.** Since $Y_i \sim \mathcal{N}(\theta, \sigma^2)$, we know that $\mathbb{E}(Y_i) = \theta$ and $\mathrm{Var}(Y_i) = \sigma^2$ for all $i$. Therefore, as in Theorem 2.8, if $\hat{\theta} := \overline{Y}$, then $\mathbb{E}(\hat{\theta}) = \mathbb{E}(Y_1) = \theta$ so that $\hat{\theta}$ is, in fact, an unbiased estimator of $\theta$. Moreover, since $Y_1, \ldots, Y_n$ are i.i.d., we also conclude using Theorem 2.8 that

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{n}\,\mathrm{Var}(Y_1) = \frac{\sigma^2}{n}.$$

In Example 4.3 we determined that $I(\theta) = \sigma^{-2}$. Therefore,

$$\frac{1}{nI(\theta)} = \frac{1}{n\sigma^{-2}} = \frac{\sigma^2}{n} = \mathrm{Var}(\hat{\theta}).$$

Since the lower bound of the Cramér-Rao inequality has been attained, we conclude that $\overline{Y}$ must be the minimum variance unbiased estimator of $\theta$.