Fisher Information Matrix

• Fundamental role of data analysis is to **extract information from data**
• Parameter estimation for models is central to process of extracting information
• The Fisher **information matrix** plays a central role in parameter estimation for measuring information

---

Information matrix summarizes the amount of information in the data relative to the parameters being estimated
Problem Setting

- Consider the classical statistical problem of estimating parameter vector $\theta$ from $n$ data vectors $z_1, z_2, \ldots, z_n$
- Suppose have a probability density and/or mass function associated with the data
- The parameters $\theta$ appear in the probability function and affect the nature of the distribution
  - Example: $z_i \sim N(\text{mean}(\theta), \text{covariance}(\theta))$ for all $i$
- Let $\ell(\theta|z_1, z_2, \ldots, z_n)$ represent the likelihood function, i.e., the p.d.f./p.m.f. viewed as a function of $\theta$ conditioned on the data
Selected Applications

• Information matrix is measure of performance for several applications. Four uses are:

1. **Confidence regions for parameter estimation**
   - Uses asymptotic normality and/or Cramér-Rao inequality

2. **Prediction bounds for mathematical models**

3. **Basis for “D-optimal” criterion for experimental design**
   - Information matrix serves as measure of how well $\theta$ can be estimated for a given set of inputs

4. **Basis for “noninformative prior” in Bayesian analysis**
   - Sometimes used for “objective” Bayesian inference
Information Matrix—Definition

- Recall likelihood function $\ell(\theta|z_1, z_2, \ldots, z_n)$
- Information matrix defined as

$$F_n(\theta) = E\left( \frac{\partial \log \ell}{\partial \theta} \frac{\partial \log \ell}{\partial \theta^T} \right)$$

where expectation is w.r.t. $z_1, z_2, \ldots, z_n$

- Equivalent form based on Hessian matrix:

$$F_n(\theta) = -E\left( \frac{\partial^2 \log \ell}{\partial \theta \partial \theta^T} \right)$$

- $F_n(\theta)$ is positive semidefinite of dimension $p \times p$ ($p=\text{dim}(\theta)$)
Information Matrix—Two Key Properties

• Connection of $F_n(\theta)$ and uncertainty in estimate $\hat{\theta}_n$ is rigorously specified via two famous results ($\theta^* = \text{true value of } \theta$):

1. Asymptotic normality:

   $$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{\text{dist}} N(0, \bar{F}^{-1})$$

   where $\bar{F} \equiv \lim_{n \to \infty} F_n(\theta^*)/n$

2. Cramér-Rao inequality:

   $$\text{cov}(\hat{\theta}_n) \geq F_n(\theta^*)^{-1} \text{ for all } n$$

Above two results indicate: greater variability of $\hat{\theta}_n$ ⇒ “smaller” $F_n(\theta)$ (and vice versa)