Fisher Information

Statistics · Module 2 — Statistical Models

Node 2.5

Position in Knowledge Graph

System:
Statistics

Structural Domain:
Statistical Models and Likelihood

Node:
2.5

Formal Definition

Let $\ell(\theta)$ be the log-likelihood for a single observation.

The **Fisher Information** is defined as:

$$\mathcal{I}(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right].$$

Equivalently, under regularity conditions:

$$\mathcal{I}(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right].$$

For n i.i.d. observations:

$$\mathcal{I}_n(\theta) = n\mathcal{I}(\theta).$$

Intuition

Fisher Information measures how sharply peaked the likelihood is.

- Flat likelihood → low information
- Sharp likelihood → high information

It quantifies how much the data tells us about the parameter.

It is curvature of the log-likelihood in expectation.

Structural Role

Fisher Information underlies:

- Cramér–Rao lower bound
- Asymptotic normality of MLE
- Efficiency
- Information geometry
- Hypothesis testing

It introduces metric structure on parameter space.

Mathematical Development

Score Variance Form

Since:

$$S(\theta) = \frac{\partial}{\partial \theta} \log f_\theta(X),$$

and

$$\mathbb{E}[S(\theta)] = 0,$$

we have:

$$\mathcal{I}(\theta) = \mathrm{Var}(S(\theta)).$$

Thus, information equals variance of the score.

Example: Bernoulli(p)

$$\log f_p(x) = x \log p + (1-x)\log(1-p).$$

Score:

$$S(p) = \frac{x}{p} - \frac{1-x}{1-p}.$$

Then:

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

Information is highest near boundaries.

Cross-Links

- **Linear Algebra:** Information matrix is positive semidefinite.
- **AI:** Natural gradient uses Fisher Information metric.
- **Economics:** Efficiency bounds in estimation.
- **Quantum:** Quantum Fisher Information generalizes classical case.

Structural Compression

Invariant:
Information ≥ 0.

Mechanism:
Curvature of expected log-likelihood.

Cross-System Echo:
In optimization, curvature determines step size and stability.

Exercises

1. Prove equivalence of the two definitions of Fisher Information.
2. Compute Fisher Information for Normal mean with known variance.
3. Explain why information scales linearly with sample size.
4. Show that score has zero expectation under regularity conditions.

Statistics · Module 2 — Statistical Models · Node 2.5
Concepts: fisher-information, likelihood

hbar.university — own.your.cognition