Log-Likelihood and the Score Function

Statistics · Module 2 — Statistical Models

Node 2.4

Position in Knowledge Graph

System:
Statistics

Structural Domain:
Statistical Models and Likelihood

Node:
2.4

Formal Definition

Given the likelihood function:

$$L(\theta; x_1, \ldots, x_n),$$

the **log-likelihood** is defined as:

$$\ell(\theta) = \log L(\theta).$$

Under i.i.d. sampling:

$$\ell(\theta) = \sum_{i=1}^{n} \log f_\theta(x_i).$$

Score Function

The **score function** is the gradient of the log-likelihood:

$$S(\theta) = \frac{\partial}{\partial \theta} \ell(\theta).$$

For vector parameters:

$$S(\theta) = \nabla_\theta \ell(\theta).$$

Intuition

Taking logs turns products into sums.

Optimization becomes tractable.

The score measures:

> How sensitive the likelihood is to changes in the parameter.

If the score is zero, we are at a stationary point.

Structural Role

Log-likelihood enables:

- Maximum Likelihood Estimation (MLE)
- Derivation of estimating equations
- Asymptotic normality
- Fisher Information
- Numerical optimization

The score is the foundation of modern inference.

Mathematical Development

Additivity

Because of independence:

$$\ell(\theta) = \sum_{i=1}^{n} \ell_i(\theta).$$

The score becomes:

$$S(\theta) = \sum_{i=1}^{n} S_i(\theta).$$

This additive structure drives asymptotic theory.

Example: Bernoulli Model

If:

$$L(p) = p^k (1-p)^{n-k},$$

then:

$$\ell(p) = k \log p + (n-k) \log(1-p).$$

Score:

$$\frac{d}{dp}\ell(p) = \frac{k}{p} - \frac{n-k}{1-p}.$$

Setting score to zero gives:

$$\hat{p} = \frac{k}{n}.$$

Cross-Links

- **Linear Algebra:** Score is gradient; Hessian gives curvature.
- **AI:** Training neural networks minimizes negative log-likelihood.
- **Economics:** Structural models solved via likelihood maximization.
- **Quantum:** Log-likelihood used in state reconstruction algorithms.

Structural Compression

Invariant:
Log preserves maxima of likelihood.

Mechanism:
Product → sum → gradient optimization.

Cross-System Echo:
In ML, loss gradients guide parameter updates.

Exercises

1. Derive log-likelihood for Poisson model.
2. Compute score for Normal mean with known variance.
3. Show that maximizing likelihood equals maximizing log-likelihood.
4. Explain why logs simplify asymptotic analysis.

Statistics · Module 2 — Statistical Models · Node 2.4
Concepts: likelihood, cost-function

hbar.university — own.your.cognition