Likelihood Function

Statistics · Module 2 — Statistical Models

Node 2.3

Position in Knowledge Graph

System:
Statistics

Structural Domain:
Statistical Models and Likelihood

Node:
2.3

Formal Definition

Let:

$$X_1, \ldots, X_n$$

be observed data from a statistical model:

$$\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}.$$

The **likelihood function** is defined as:

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

viewed as a function of θ, with the data fixed.

Under the i.i.d. assumption:

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i).$$

Intuition

Probability asks:

> Given θ, how likely is the data?

Likelihood asks:

> Given the data, how plausible is θ?

The same formula. Different perspective.

Likelihood reverses the direction of reasoning.

Structural Role

Likelihood is the engine of:

- Maximum likelihood estimation (MLE)
- Likelihood ratio tests
- Asymptotic theory
- Information measures
- Bayesian updating (via Bayes' rule)

It is the bridge between model and inference.

Mathematical Development

Key Distinction

Probability:

$$f_\theta(x)$$

is a function of data x.

Likelihood:

$$L(\theta; x)$$

is a function of parameter θ.

The functional role changes.

Example: Bernoulli Model

If $X_i \sim \text{Bernoulli}(p)$,

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}.$$

This simplifies to:

$$L(p) = p^{\sum x_i}(1-p)^{n-\sum x_i}.$$

Likelihood depends only on the number of successes.

Worked Example

Let:

$$\sum x_i = k.$$

Then:

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

The parameter p that maximizes this will be the MLE.

Cross-Links

- **Linear Algebra:** Log-likelihood often leads to quadratic forms.
- **AI:** Empirical risk minimization equals negative log-likelihood minimization.
- **Economics:** Structural estimation uses likelihood maximization.
- **Quantum:** Likelihood used in quantum state tomography.

Structural Compression

Invariant:
Likelihood is the joint density viewed as function of parameter.

Mechanism:
Fix data, vary parameter.

Cross-System Echo:
In ML, model parameters are optimized against fixed dataset.

Exercises

1. Derive likelihood for Poisson model.
2. Show that likelihood depends only on sufficient statistics in Bernoulli case.
3. Explain difference between probability and likelihood.

Statistics · Module 2 — Statistical Models · Node 2.3
Concepts: likelihood

hbar.university — own.your.cognition