Conditional Probability and Conditional Expectation

Statistics · Module 1 — Probability Foundations

Node 1.7

Position in Knowledge Graph

System:
Statistics

Structural Domain:
Probability Foundations

Node:
1.7

Formal Definition

Conditional Probability

For events $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Conditional Expectation (Discrete Case)

If Y is a discrete random variable,

$$\mathbb{E}[X \mid Y = y] = \sum_x x \, \mathbb{P}(X = x \mid Y = y).$$

General Definition (Measure-Theoretic Form)

The **conditional expectation** of X given a sigma-algebra $\mathcal{G} \subseteq \mathcal{F}$ is a random variable $\mathbb{E}[X \mid \mathcal{G}]$ such that:

1. $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable.
2. For all $G \in \mathcal{G}$,

$$\int_G \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} = \int_G X \, d\mathbb{P}.$$

Intuition

Conditional probability updates belief given information.

Conditional expectation updates averages given information.

If expectation is balance, conditional expectation is balance **after learning something**.

It represents the best prediction of X given available information.

Structural Role

Conditional expectation is the backbone of:

- Bayesian inference
- Regression
- Martingale theory
- Filtering
- Sequential decision-making

It formalizes how information changes predictions.

Mathematical Development

Law of Total Probability

If $\{B_i\}$ partitions Ω,

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Law of Total Expectation

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X \mid Y]\big].$$

This is sometimes called the **tower property**.

Best Prediction Property

Conditional expectation minimizes mean squared error:

$$\mathbb{E}[X \mid Y] = \arg\min_g \mathbb{E}\big[(X - g(Y))^2\big].$$

This connects probability directly to regression.

Worked Example

Let $X \sim \text{Bernoulli}(p)$ and suppose we observe an event B.

Then:

$$\mathbb{P}(X = 1 \mid B) = \frac{\mathbb{P}(B \mid X = 1)\mathbb{P}(X = 1)}{\mathbb{P}(B)}.$$

This is Bayes' rule.

Cross-Links

- **Linear Algebra:** Conditional expectation is an orthogonal projection in L^2.
- **AI:** Bayesian updating and predictive modeling.
- **Economics:** Expectations conditional on information sets.
- **Quantum:** State update after measurement.

Structural Compression

Invariant:
Conditioning preserves total expectation (tower property).

Mechanism:
Projection onto information sigma-algebra.

Cross-System Echo:
In ML, regression projects outcomes onto feature space.

Exercises

1. Prove the law of total probability.
2. Prove the tower property.
3. Show that conditional expectation minimizes mean squared error.
4. Derive Bayes' rule from the definition of conditional probability.

Statistics · Module 1 — Probability Foundations · Node 1.7
Concepts: conditional-probability, expectation-value

hbar.university — own.your.cognition