Distribution Functions

Statistics · Module 1 — Probability Foundations

Node 1.4

Position in Knowledge Graph

System:
Statistics

Structural Domain:
Probability Foundations

Node:
1.4

Formal Definition

The **cumulative distribution function (CDF)** of a random variable X is the function:

$$F_X(x) = \mathbb{P}(X \leq x).$$

The CDF satisfies:

1. F_X is non-decreasing.
2. $\lim_{x \to -\infty} F_X(x) = 0$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. F_X is right-continuous.

Intuition

The CDF accumulates probability from left to right.

At any real number x, it tells us:

> "How much probability mass lies at or below this value?"

It compresses the full distribution of X into a single monotonic function.

Structural Role

The distribution function:

- Fully characterizes the law of a random variable.
- Enables computation of probabilities:

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

- Bridges discrete and continuous cases into a unified framework.

All density and mass functions are derived from the CDF.

Mathematical Development

Discrete Case

If X takes values x_i with probabilities p_i, then:

$$F_X(x) = \sum_{x_i \le x} p_i.$$

The CDF has jump discontinuities at mass points.

Continuous Case

If X has density $f_X(x)$, then:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

The density is the derivative of the CDF:

$$f_X(x) = \frac{d}{dx} F_X(x).$$

Worked Example

Fair Die

Let $X \in \{1, 2, 3, 4, 5, 6\}$ with equal probability.

Then:

$$F_X(3) = \frac{3}{6} = 0.5.$$

The CDF increases in steps of $\frac{1}{6}$.

Cross-Links

- **Linear Algebra:** CDFs generalize to multivariate distributions via joint probability measures.
- **AI:** Model outputs often represent CDF or probability mass functions.
- **Economics:** Risk analysis depends on distribution tails.
- **Quantum:** Measurement outcome probabilities form cumulative distributions over eigenvalues.

Structural Compression

Invariant:
CDF is monotone and bounded between 0 and 1.

Mechanism:
Accumulation of probability mass via integration or summation.

Cross-System Echo:
In calculus, integration accumulates area; in learning, loss aggregates over data.

Exercises

1. Prove that every CDF is right-continuous.
2. Show that a CDF uniquely determines a probability measure on $\mathbb{R}$.
3. Sketch the CDF of a Bernoulli(p) random variable.

Statistics · Module 1 — Probability Foundations · Node 1.4
Concepts: random-variables, probability-space

hbar.university — own.your.cognition