Probability Mass Function (PMF)

The **PMF** of a discrete random variable X is:

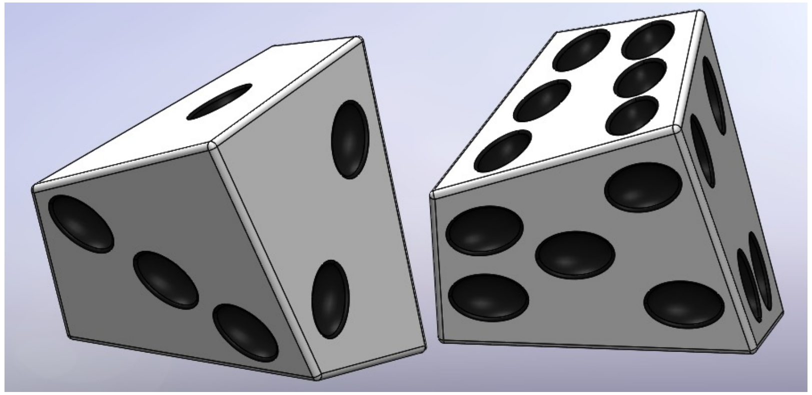
$$P(X = x)$$

It must satisfy:

$$P(X = x) \geq 0$$

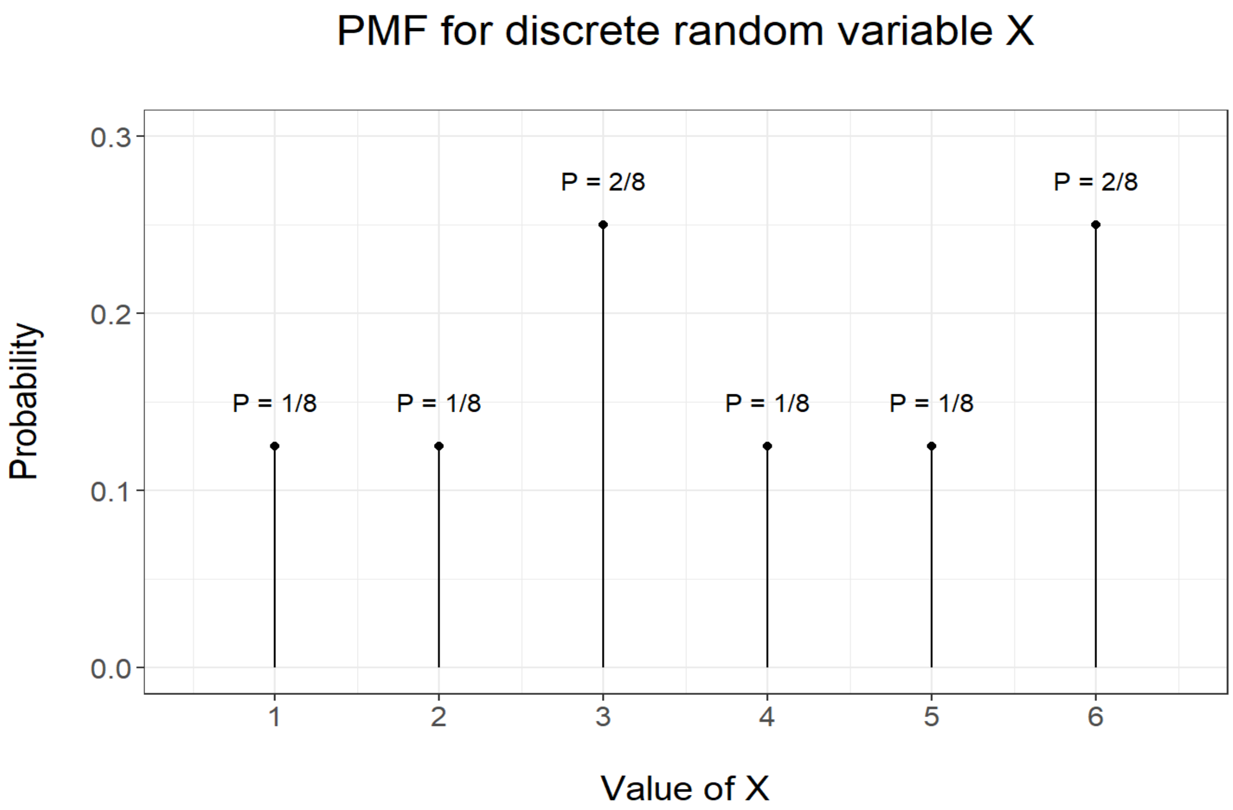
$$\sum_x P\,(X = x) = 1$$

PMF



x	1	2	3	4	5	6
$p(x)$	1/8	1/8	2/8	1/8	1/8	2/8

PMF


PMF for discrete random variable X
Probability
Value of X
0.3
0.2
0.1
0.0
1
2
3
4
5
6
P = 1/8
P = 1/8
P = 2/8
P = 1/8
P = 1/8
P = 2/8

Cumulative Distribution Function (CDF)

The CDF answers the question:

“What is the probability that a random variable is less than or equal to a given value?”

Random variables recap

Random variable = a numerical outcome of a random experiment.

- **Discrete random variable**
- **Continuous random variable**

The definition of the CDF applies to **both discrete and continuous** random variables.

Definition of the CDF

Let X be a random variable.

The CDF of X, denoted by $F(x)$, is defined as:

$$F(x) = P(X \leq x)$$

This means:

For each real number x, the CDF gives the probability that the random variable X is **less than or equal to** x.

Example 1 (coin tosses) - Discrete Random Variable

A fair coin is tossed **twice**. Let X= number of heads obtained.

Step 1: List possible values of X

$$X \in \{0, 1, 2\}$$

Step 2: Probability mass function (PMF)

x	0	1	2
P(X=x)	1/4	1/2	1/4

Step 3: Construct the CDF

$$F(x) = P(X \leq x)$$

$F(x) = 0, for\ x < 0$
$F(2) = P(X \leq 2) = 1$
$F(1) = P(X \leq 1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$
$F(0) = P(X \leq 0) = 1/4$

Example 2 (rolling a die) - Discrete Random Variable

Let X be the number shown when a fair die is rolled.

Step 1: Definition of the CDF

$$F(x) = P(X \leq x)$$

Step 2: Compute specific values

$F(6) = 1$

$F(4) = P(X \leq 4) = \frac{4}{6} = \frac{2}{3}$

$F(2) = P(X \leq 2) = P(1) + P(2) = 2/6 = 1/3$

Interpretation

$F(4) = \frac{2}{3}$ means the probability that the outcome is **4 or less** is **66.7%**.

CDF for discrete random variables

Definition

If X is a discrete random variable with PMF $p(x)$, then:

$$F(x) = \sum_{t \leq x} p\,(t)$$

That is, the CDF is the **sum of probabilities** of all values less than or equal to x.

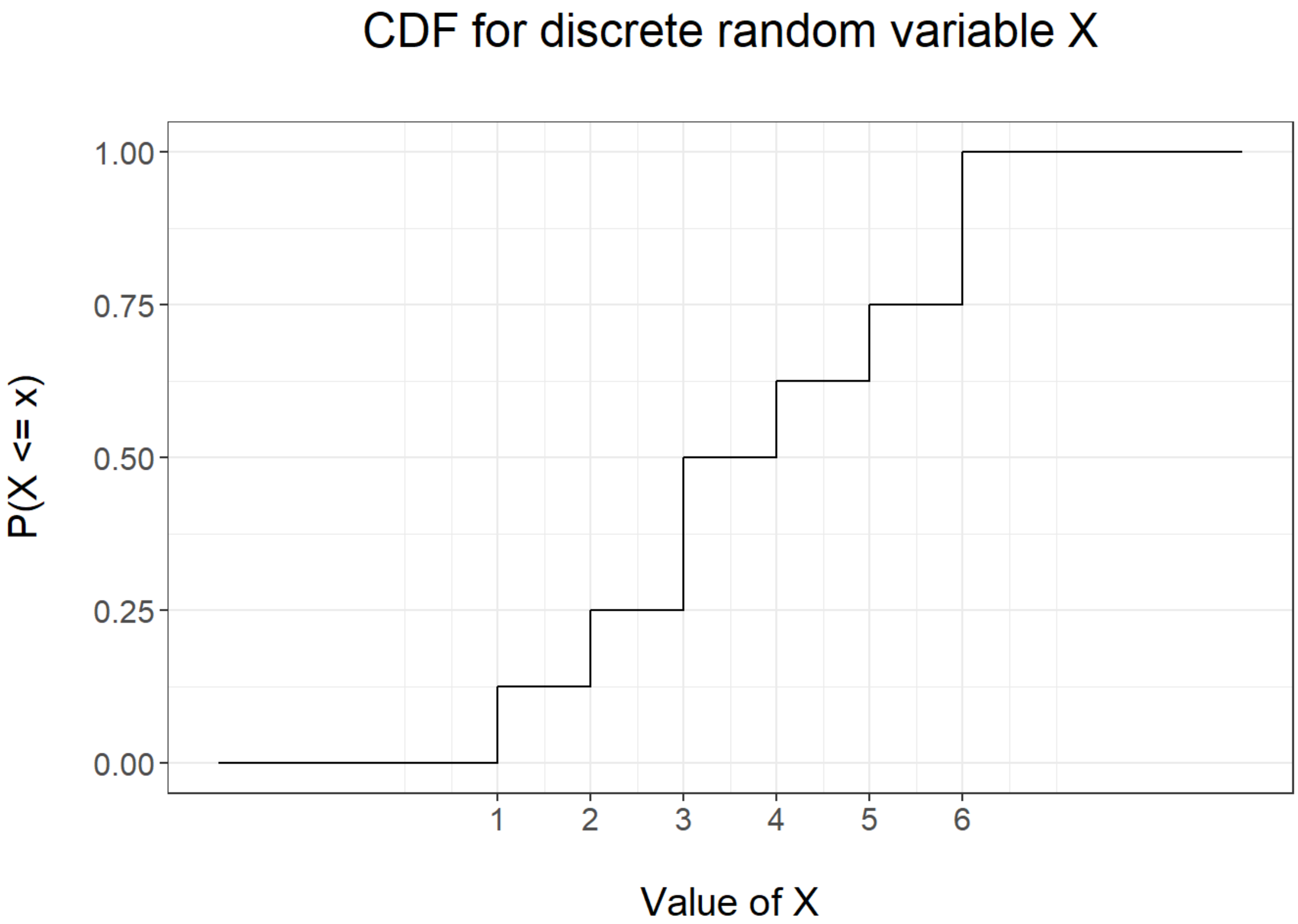
Example (tossing a fair die)

Let X be the outcome when a fair die is rolled.

x	1	2	3	4	5	6
P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6

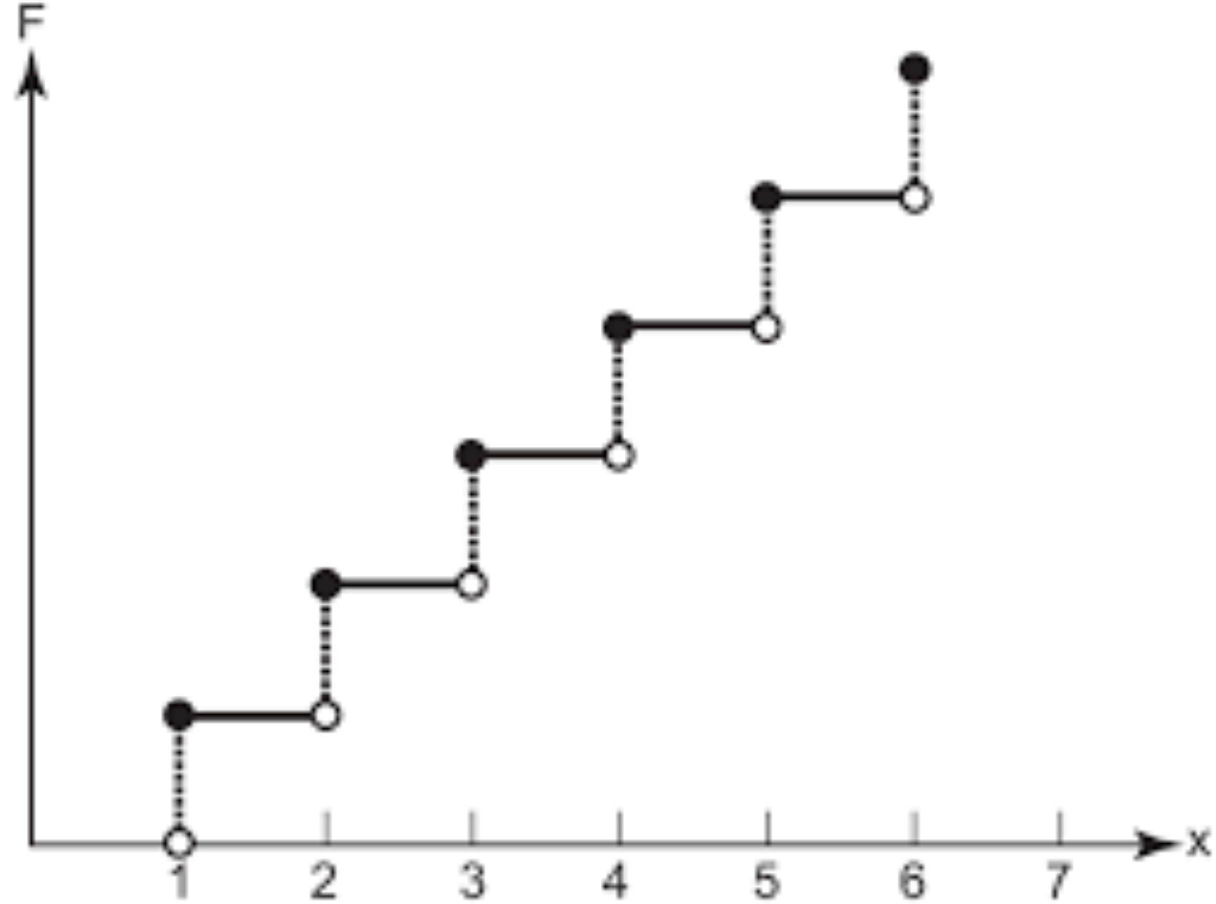
Now compute the CDF:

- $F(1) = P(X \leq 1) = 1/6$
- $F(2) = P(X \leq 2) = 2/6$
- $F(3) = 3/6 = 1/2$
- $F(6) = 1$



CDF for discrete random variable X
1.00
0.75
0.50
0.25
0.00
P(X <= x)
1
2
3
4
5
6
Value of X

Graphical Interpretation of the cumulative distribution function for a fair six-sided die.



CDF for continuous random variables

Definition

If X is a continuous random variable with probability density function (PDF) $f(x)$, then:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

Here:

The CDF is the **area under the PDF curve** from $-\infty$ to x

Applications of the CDF

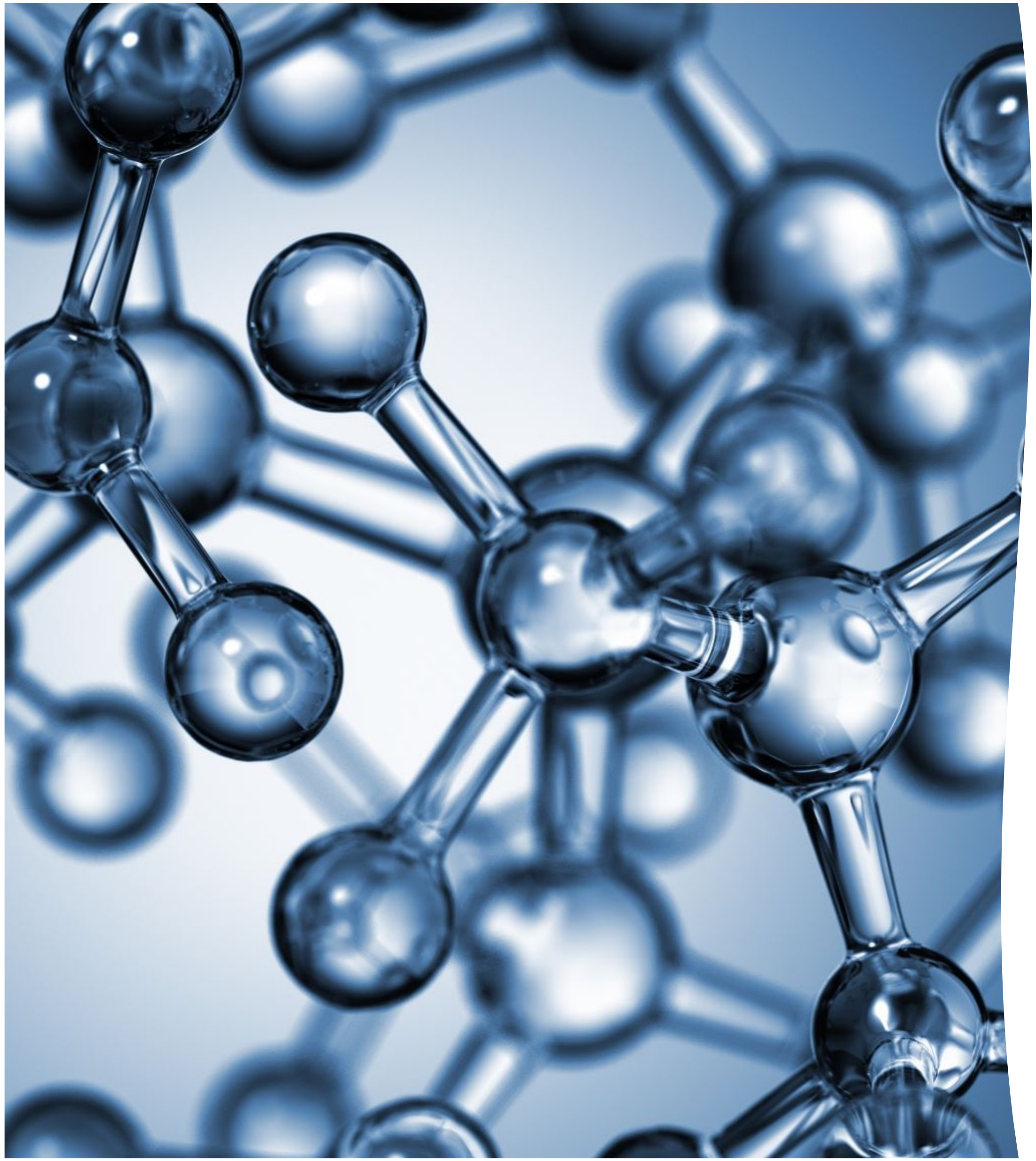
The CDF is widely used in:

- Probability calculations
- Reliability and survival analysis
- Queueing theory
- Economics and finance
- Hypothesis testing
- Simulation and random number generation

Common mistakes to avoid

Confusing PMF/PDF with CDF

Forgetting that the CDF is **non-decreasing**



Summary

- The CDF gives cumulative probabilities
- It applies to both discrete and continuous variables
- It completely characterizes a probability distribution
- It is essential for computing probabilities and understanding distributions

Measures of central tendency and dispersion

Expectation (Mean), Median, Mode, Variance, and Standard Deviation

1. **Expectation (Mean)**

The **expectation**, also called the **mean** or **expected value**, represents the *average outcome* of a random variable.

For **discrete random variables**:

$$E(X) = \sum x \ P(X = x)$$

For **continuous random variables**:

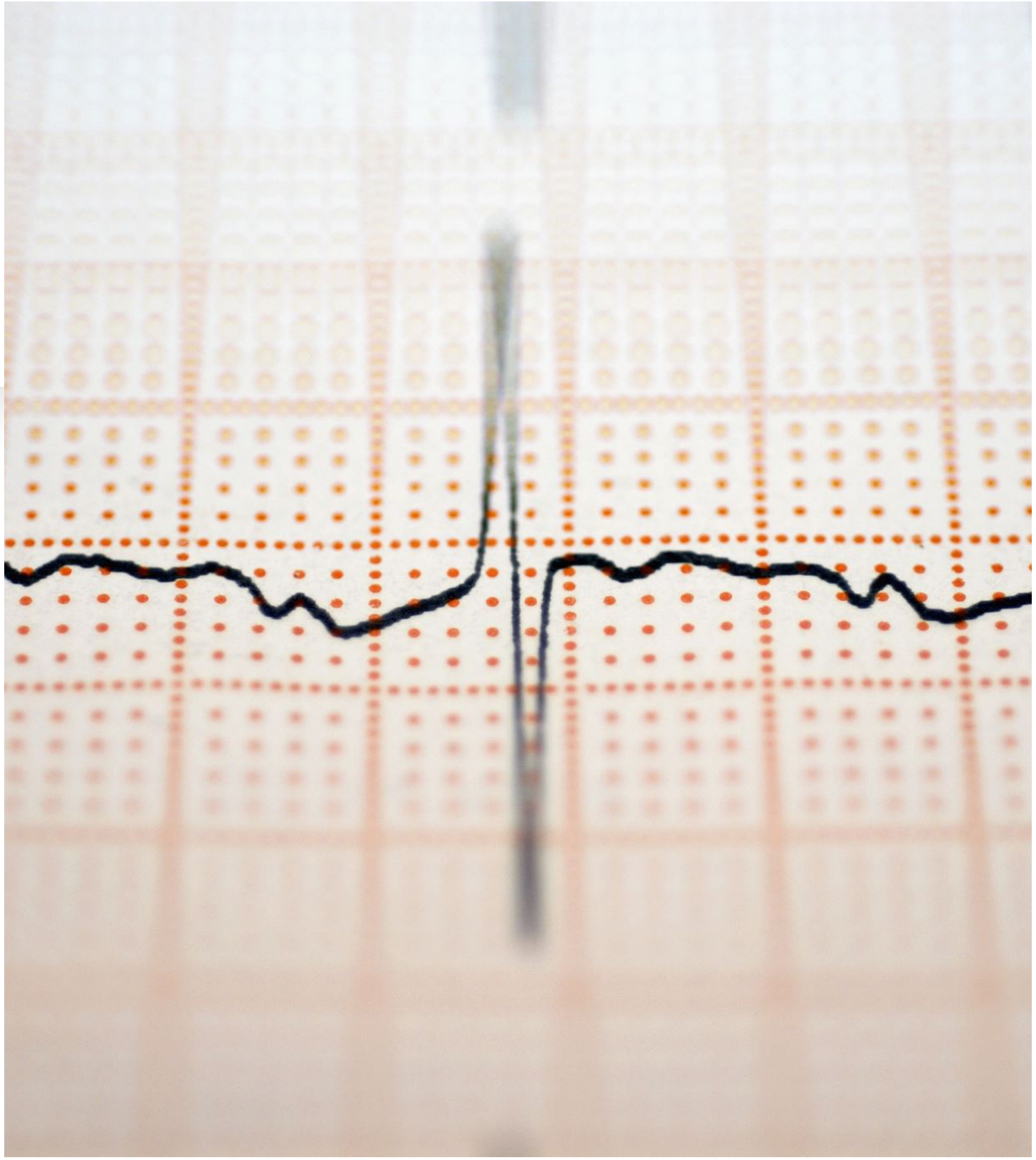
$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Example (discrete)

Suppose a fair die is rolled.

$$E(X) = \frac{1+2+3+4+5+6}{6} = 3.5$$

You will never roll a 3.5, but over many rolls, the *average* outcome approaches 3.5.



2. Median

The **median** is the **middle value** of an ordered dataset.

- If the number of observations is **odd** → the middle value
- If **even** → average of the two middle values

The median is **robust to extreme values (outliers)**.

Example

Dataset: $2, \ 4, \ 5, \ 7, \ 100$

- Mean = 23.6
- Median = **5**

The median better represents the *typical* value because it is not distorted by the outlier 100.

3. Mode

The **mode** is the value that occurs **most frequently** in a dataset.

A dataset may have:

- **One mode** (unimodal)
- **Two modes** (bimodal)
- **More than two modes** (multimodal)
- **No mode**

Example

Dataset: $3,\ 4,\ 4,\ 5,\ 6$

Mode = **4**

Comparing mean, median, and mode

- **Symmetric distribution**:
  Mean = Median = Mode
- **Right-skewed distribution**:
  Mean > Median > Mode
- **Left-skewed distribution**:
  Mean < Median < Mode

4. Variance

While the mean tells us the center, it does **not** tell us how spread out the data is.
The **variance** measures the **average squared deviation from the mean**.

For a population:

$$\sigma^2 = \frac{1}{N} \sum (x - \mu)^2$$

For a sample:

$$s^2 = \frac{1}{n-1} \sum (x - \bar{x})^2$$

Example

Consider the following dataset representing the number of hours five students studied for an exam: 2, 4, 6, 8, 10

Step 1: Calculate the Mean

$$\bar{x} = \frac{2 + 4 + 6 + 8 + 10}{5} = \frac{30}{5} = 6$$

Step 2: Find deviations from the mean

Value (x)	$\bar{x}$
2	−4
4	−2
6	0
8	2
10	4

Step 3: Square the deviations

x - $\bar{x}$	(x - $\bar{x}$)^2
−4	16
−2	4
0	0
2	4
4	16

Step 4: Compute the variance

Population variance:

$$\sigma^2 = \frac{16 + 4 + 0 + 4 + 16}{5} = \frac{40}{5} = 8$$

Sample variance:

$$s^2 = \frac{16 + 4 + 0 + 4 + 16}{4} = \frac{40}{4} = 10$$

5. Standard deviation

The **standard deviation** is the **square root of the variance**.

$$\sigma = \sqrt{\sigma^2}, \qquad s = \sqrt{s^2}$$

It measures spread **in the same units as the data**, making it easier to interpret.

Interpretation

- Small standard deviation → data points are close to the mean
- Large standard deviation → data points are widely dispersed

Example

Mean height = 170 cm

Standard deviation = 5 cm

Most individuals have heights between **165 cm and 175 cm**.

Why these measures matter together

Mean → central tendency

Median → robust center

Mode → most common value

Variance & Standard deviation
→ variability and consistency

Applications

- Quality control (defective items)
- Genetics (inheritance patterns)
- Communication systems (bit errors)
- Medical testing (positive/negative outcomes)
- Computer science (algorithms, cryptography)

Summary & Conclusion

- Discrete probability deals with countable outcomes
- Random variables allow us to quantify uncertainty
- PMF, CDF, expectation, and variance summarize distributions
- Discrete distributions like the binomial distribution are widely used in practice