Normal Distribution (a.k.a. Gaussian Distribution)

Before we dive into the details, let's understand two key concepts related to normal distribution, mean, and standard deviation.

$$\mu = \frac{1+2+4+7+9}{5}$$

1. What is Mean?

The **mean** (or average) of a dataset is the sum of all values divided by the number of values.

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \dots x_n}{n}$$

Formula:

mean = (x1 + x2 + x3 + ... + xn) / n

It represents the central value of the data.

---

2. What is Standard Deviation?

The **standard deviation** measures how spread out the numbers are from the mean.

Steps to calculate standard deviation:

$$\underbrace{1 \quad 2 \quad 9 \quad 7 \quad 8}$$

1. Find the mean
2. Subtract the mean from each value and square the result
3. Take the average of these squared differences (this is the variance)
4. Take the square root of the variance

$$\underbrace{1 \quad 10^7 \quad -10^8 \quad 10^{120}}$$

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{n}}$$

Formula:

standard deviation (sigma) = sqrt((1/n) * sum((xi - mean)^2))

$\sigma^2 =$ variance.

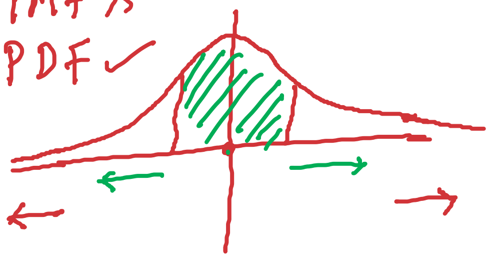
A smaller standard deviation means the data points are close to the mean. A larger standard deviation means the data is more spread out.

What is Normal Distribution?

> A **normal distribution** is a continuous probability distribution that is **bell-shaped and symmetric** around the mean.

PMF ✗

PDF ✓



It describes variables where:

- Most values cluster around the **average (mean)**.
- Extreme values (very high or low) are **rare**.

Think of:

- Heights of people
- Test scores
- Measurement errors

All tend to follow a **normal curve**.

Key Properties:

- **Mean (μ)**: Center of the distribution
- **Standard Deviation (σ)**: Spread of the distribution

Shape of the Curve:

70 10

70 + 20 = 90

70 - 20 = 50

$\mu \pm 1\sigma$

70 + 1×10 = 80

70 - 10 = 60

- Bell-shaped
- Symmetrical
- Peaks at the mean
- About 68% of the data lies within ±1σ, 95% within ±2σ, 99.7% within ±3σ → This is the famous **68–95–99.7 Rule**

70 + 30 100

70 - 30 40

Formula (Plain Text):

f(x) = (1 / (σ * sqrt(2π))) * e^(-(x - μ)² / (2σ²))

$$PDF = \frac{1}{\sigma\sqrt{2\pi}} \times e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Where:

- `μ` = mean
- `σ` = standard deviation
- `e` = Euler's number (≈ 2.718)
- `π` = pi (≈ 3.14159)

This gives the **probability density** for a given value `x` .

> You **don't need to memorize** this formula — but understanding its shape and behavior is essential.

In Python:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Generate values
x = np.linspace(-4, 4, 1000)
mean = 0
std_dev = 1

# Get the probability density
y = norm.pdf(x, loc=mean, scale=std_dev)

# Plot
plt.plot(x, y)
plt.title("Standard Normal Distribution (μ=0, σ=1)")
plt.xlabel("x")
plt.ylabel("Probability Density")
```

```
plt.grid(True)
plt.show()
```

Summary

Property	Value
Type	Continuous
Shape	Bell curve
Key parameters	Mean (μ), Std. Dev (σ)
Formula	$f(x) = (1 / (\sigma\sqrt{2\pi})) * e^{(-(x - \mu)^2 / 2\sigma^2)}$
Python	`scipy.stats.norm.pdf(x, μ, σ)`