The Gaussian Distribution Meaning, Derivations, Simulations

Giacomo Babudri

January 2024

1 Introduction

The Gaussian Distribution, commonly referred to as the Normal Distribution, is a smooth and symmetrical probability distribution. Its shape resembles a bell curve, and it holds significant importance in the realms of statistics and probability theory, being extensively utilized for various applications.

1.1 Definition

In statistics, a normal distribution or Gaussian distribution is a type of continuous probability distribution for a real-valued random variable. The Gaussian Distribution is defined by its probability density function (PDF), which describes the likelihood of a random variable taking a particular value. For a random variable X following a Gaussian distribution with mean μ and standard deviation σ, the PDF is given by:

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

- x is the variable for which the probability is calculated.
- μ is the mean of the distribution, representing the central value.
- σ is the standard deviation, indicating the spread or dispersion of the distribution.
- π is a mathematical constant approximately equal to 3.14159.
- $\frac{1}{\sigma\sqrt{2\pi}}$: This term is a normalization factor that ensures the area under the curve sums to 1. It accounts for the spread and height of the curve.
- $e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$: This is the exponential term that determines the shape of the bell curve. The argument $\frac{x-\mu}{\sigma}$ standardizes the variable x by subtracting the mean μ and dividing by the standard deviation σ. The square of this standardized value is then multiplied by $-\frac{1}{2}$ and exponentiated.

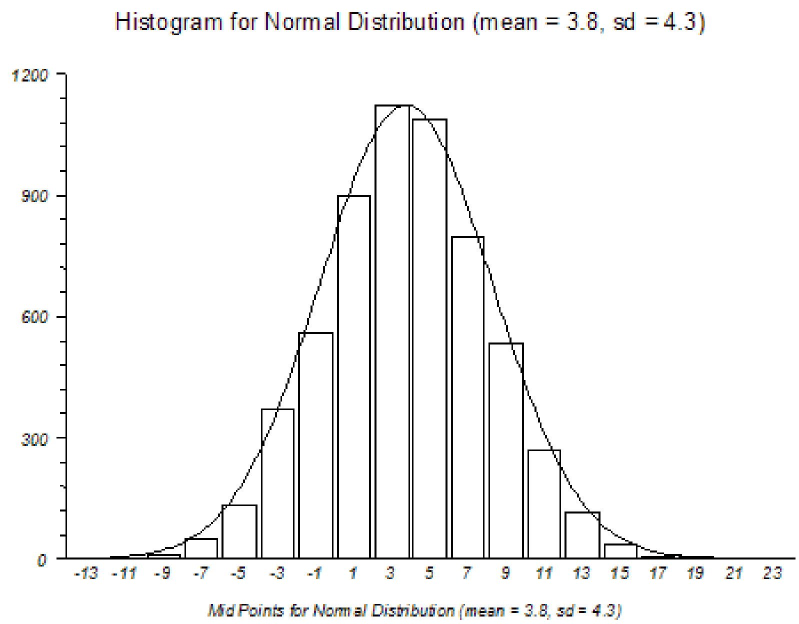

Figure 1: Representation of a gaussian distribution

1.2 Main Features

1. *Symmetry:* The Gaussian Distribution is characterized by perfect symmetry. This means that the probability density function (PDF) is identical on both sides of the mean (μ). The curve reaches its highest point at the mean and extends symmetrically in both directions.

2. *Bell-Shaped Curve:* One of the distinctive features is its bell-shaped curve. The probability density function forms a smooth, symmetric curve resembling a bell. This shape is a result of the likelihood of values decreasing as they move away from the mean.

3. *68-95-99.7 Rule:* This empirical rule provides insights into the distribution's spread. Approximately 68% of the data falls within one standard deviation (σ) of the mean, 95% within two standard deviations, and 99.7% within three standard deviations. This rule underscores the concentration of data around the mean.

4. *Central Limit Theorem (CLT):* A fundamental theorem associated with the Gaussian Distribution, the Central Limit Theorem states that the sum or average of a large number of independent, identically distributed random variables will be approximately normally distributed. This holds true regardless of the original distribution of the variables. The CLT is foundational in statistical analysis and hypothesis testing.

2 Simulation

2.1 Code

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

# Set the mean and standard deviation
mean = 0
std_dev = 1

# Generate random samples from a Gaussian distribution
num_samples = 1000
samples = np.random.normal(mean, std_dev, num_samples)

# Set a greater number of bins in the histogram
num_bins = 10

# Plot the normalized histogram
plt.hist(samples, bins=num_bins, density=True, alpha=0.7, color='blue', label='Histogram')

# Plot the probability density function (PDF)
xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax, 1000)
pdf = norm.pdf(x, mean, std_dev)
plt.plot(x, pdf, color='red', linewidth=2, label='PDF')

# Add vertical lines for one, two, and three standard deviations
for i in range(-3, 4):
    plt.axvline(mean + i * std_dev, linestyle='--', color='green', alpha=0.5)

# Fill the area between the PDF curve and the x-axis
plt.fill_between(x, pdf, alpha=0.1, color='red')

# Show legend and add labels
plt.legend()
plt.title('Gaussian Distribution Simulation')
plt.xlabel('Value')
plt.ylabel('Probability Density')

# Show the plot
plt.show()
```

2.2 Output

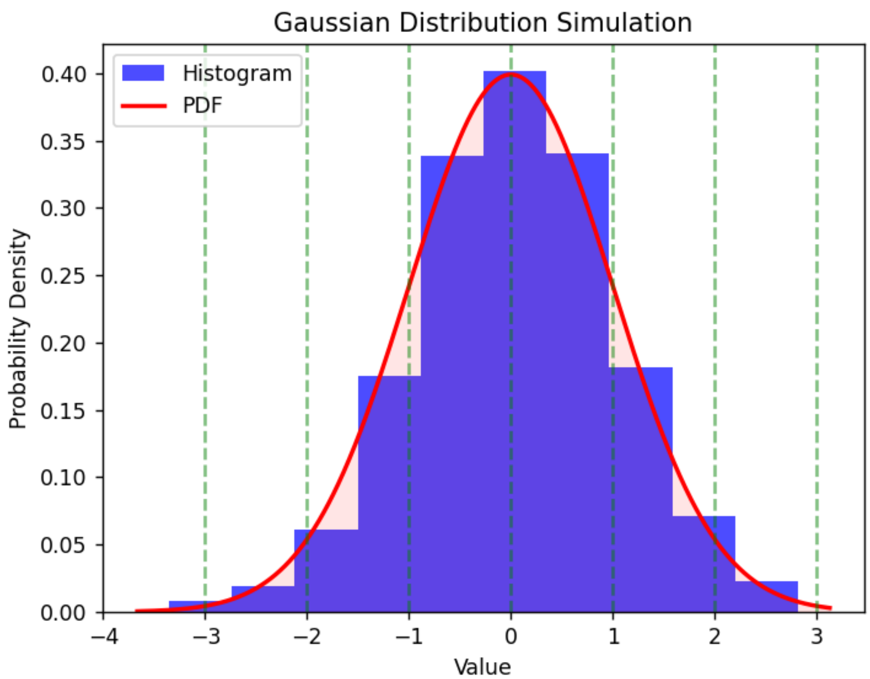

Figure 2: Simulation's output