The Gaussian Distribution: meaning, derivations and simlations

Pietro Colaguori 1936709

November 2023

1 Meaning

Also known as the normal distribution or bell curve, the Gaussian distribution, a fundamental probability distribution, crucially impacts various fields: statistics, physics, finance; and many other natural and social sciences. Its meaning and characteristics are the following:

1. **Symmetry and Bell Shape**: The Gaussian distribution is symmetric around its mean, i.e. it forms a bell-shaped curve. The apex of this curve aligns with the mean. This symmetry suggests that observing a value on either side of the mean carries an equal probability: in other words, it is equiprobable to observe identical values on both sides.

2. **Mean, Median and Mode**: The mean (average), median, and mode (most frequent value) in a Gaussian distribution all equate: they are equal to one another; moreover, they situate at the distribution's center.

3. **Parameters**: Two parameters characterize the Gaussian distribution: the mean (μ) and the standard deviation (σ). The center of the distribution is determined by its mean, whereas its spread or dispersion is controlled by its standard deviation. A wider bell curve results from a larger standard deviation.

4. **The 68-95-99.7 Rule**: The rule: the concentration of data around the mean in a Gaussian distribution is underscored by this: approximately 68% falls within one standard deviation of the mean, 95% within two standard deviations and 99.7% within three.

5. **Central Limit Theorem (CLT)**: The Gaussian distribution is closely connected to the Central Limit Theorem. According to the CLT, the distribution of the sum (or average) of a large number of independent and identically distributed random variables will be approximately normally distributed, regardless of the shape of the original distribution.

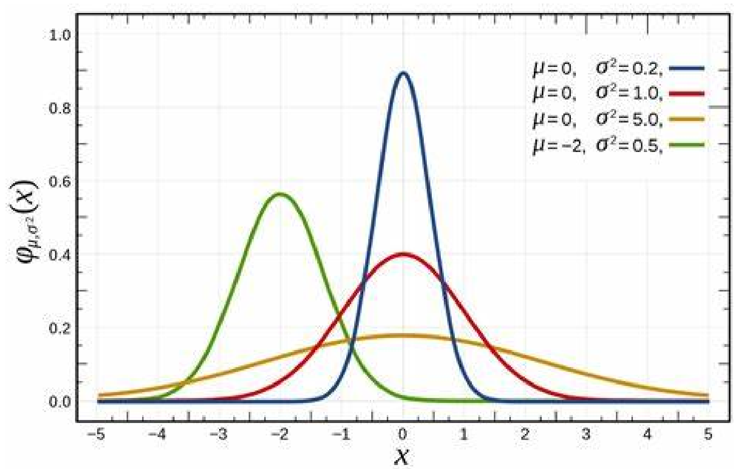

Figure 1: Gaussian Distribution with various values of μ and σ.

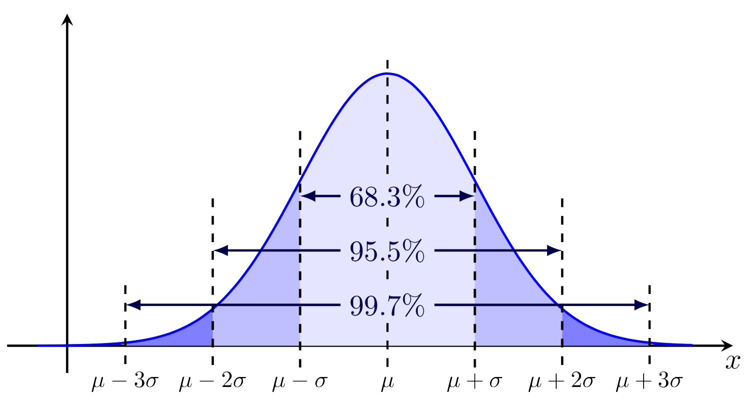

Figure 2: Graphical representation of the 68-95-99.7 rule.

6. **Common in Nature**: Many natural phenomena and measurement errors exhibit behavior that approximates a Gaussian distribution. This makes the Gaussian distribution a useful model for a wide range of applications.

7. **Parametric and Continous**: The Gaussian distribution is parametric, meaning its shape is fully determined by its parameters (mean and standard deviation). It is a continuous distribution, defined for all real numbers

8. **Maximum Entropy**: In information theory, the Gaussian distribution is noteworthy for being the probability distribution with maximum entropy among all distributions with a given mean and variance. This means it contains the least amount of information.

9. **Statistical Inference**: The Gaussian distribution is often used in statistical inference, hypothesis testing, and confidence interval construction due to its mathematical tractability and its appearance in the context of the CLT.

In summary, the Gaussian distribution is a fundamental and versatile probability distribution that describes the distribution of many natural and human-made phenomena. Its properties make it a central concept in statistics and probability theory, with broad applications in various scientific and engineering disciplines.

2 Derivation

The Probability Density Function (PDF) of the Gaussian Distribution is the following:

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

Deriving the Gaussian Distribution involves several mathematical steps and concepts.

1. **Normalization Constant**: The Gaussian distribution must be normalized so that the area under the curve integrates to 1. This property is ensured by the multiplicative constant:

   $$\frac{1}{\sqrt{s\pi}\sigma}$$

2. **Exponential Term**: the exponential term

   $$-\frac{(x-\mu)^2}{2\sigma^2}$$

   is crucial in shaping the bell curve. It arises from the probability density function of a standard normal random variable, which has a bell-shaped distribution centered at 0 with a standard deviation of 1. This exponential element can also be written in another form, which sometimes is more convenient:

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

3 Simulation

Using Python, I can generate and display a Gaussian Distribution of a desired μ and σ, which are taken as user input. Below I will show both the code and the result obtained.

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

def gaussian_simulation(mean, std_dev, sample_size):
    # Generate random samples from a normal distribution
    samples = np.random.normal(mean, std_dev, sample_size)

    # Plot histogram
    plt.hist(samples, bins=30, density=True, alpha=0.6, color='g')

    # Plot the theoretical probability density function
    xmin, xmax = plt.xlim()
    x = np.linspace(xmin, xmax, 100)
    p = norm.pdf(x, mean, std_dev)
    plt.plot(x, p, 'k', linewidth=2)

    title = f'Simulation of Gaussian Distribution\n(mean={mean}, std_dev={std_dev})'
    plt.title(title)
    plt.xlabel('Value')
    plt.ylabel('Probability Density')
    plt.grid(True)

    # Display the plot using Streamlit
    st.pyplot()

# Sidebar for user input
```

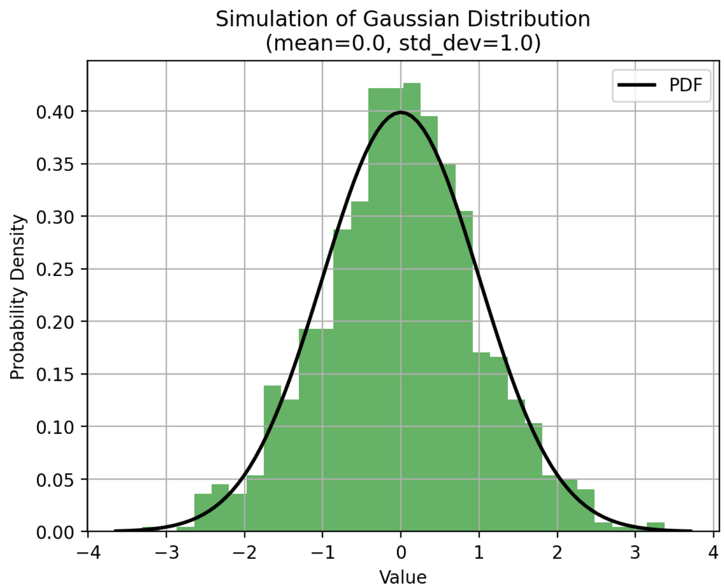

Figure 3: Result of the Python simulation, with $\mu = 0$ and $\sigma = 1$.

```python
mean = st.sidebar.number_input('Mean:', value=0.0, step=0.1)
std_dev = st.sidebar.number_input('Standard Deviation:', value=1.0, step=0.1)
sample_size = st.sidebar.number_input('Sample Size:', value=1000, step=100)

# Display simulation
if st.button('Run Simulation'):
    gaussian_simulation(mean, std_dev, sample_size)
```