The Gaussian Distribution

Wendy Rosettini

November 2023

1 Introduction

In statistics, a normal distribution or Gaussian distribution is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is

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

The parameter μ is the mean or expectation of the distribution (and also its median and mode), while the parameter σ is its standard deviation. The variance of the distribution is σ^2. A random variable with a Gaussian distribution is said to be normally distributed, and is called a normal deviate.

Normal distributions are important in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known. Their importance is partly due to the central limit theorem. It states that, under some conditions, the average of many samples (observations) of a random variable with finite mean and variance is itself a random variable—whose distribution converges to a normal distribution as the number of samples increases. Therefore, physical quantities that are expected to be the sum of many independent processes, such as measurement errors, often have distributions that are nearly normal.

Moreover, Gaussian distributions have some unique properties that are valuable in analytic studies. For instance, any linear combination of a fixed collection of normal deviates is a normal deviate. Many results and methods, such as propagation of uncertainty and least squares parameter fitting, can be derived analytically in explicit form when the relevant variables are normally distributed.

A normal distribution is sometimes informally called a bell curve. However, many other distributions are bell-shaped (such as the Cauchy, Student's t, and logistic distributions). For other names, see Naming.

The univariate probability distribution is generalized for vectors in the multivariate normal distribution and for matrices in the matrix normal distribution.

Source: [1]

[1] https://en.wikipedia.org/wiki/Normal$_{d}istribution$

Derivation of the Normal Distribution

Probability Density Function (PDF) of the Normal Distribution:

The probability density function $f(x)$ for a normal distribution with mean μ and standard deviation σ is given by:

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

Derivation Steps:

1. **Start with the Standard Normal Distribution:** The standard normal distribution has a mean (μ) of 0 and a standard deviation (σ) of 1. Its probability density function is:

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

2. **Transformation to the General Normal Distribution:** To derive the general normal distribution with mean μ and standard deviation σ, we perform a linear transformation on the standard normal variable Z:

$$X = \mu + \sigma Z$$

where X is the variable of interest following a normal distribution with mean μ and standard deviation σ.

3. **Change of Variables:** Use the change of variables formula for probability density functions:

$$f_X(x) = f_Z\left(\frac{x-\mu}{\sigma}\right)\left|\frac{dx}{dz}\right|$$

Substitute $X = \mu + \sigma Z$ into this formula.

4. **Plug in the Standard Normal PDF:** Substitute the standard normal PDF $\phi(z)$ into the formula, using the transformed variable $\frac{x-\mu}{\sigma}$.

5. **Simplify:** Simplify the expression, combining terms and simplifying the exponent.

6. **Final Result:** After simplification, you should arrive at the general form of the normal distribution PDF:

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

This derivation involves basic techniques in probability theory, calculus, and algebra. It's worth noting that the normal distribution has many important properties, and its ubiquity in statistics is due to the central limit theorem, which states that the distribution of the sum (or average) of a large number of independent, identically distributed random variables approaches a normal distribution.

Simulation of Gaussian Distribution Using Box-Muller Algorithm

1. Generation of Gaussian Data

The `generateGaussian` function generates a standard Gaussian sample using the Box-Muller algorithm.

2. Creation of an Array of Data

An array `data` is created containing 1000 Gaussian samples generated by the `generateGaussian` function.

3. Calculation of the Histogram

- The `data` array is divided into 20 bins.
- For each bin, the number of samples falling within that bin is calculated.

4. Drawing the Histogram

- Using the HTML canvas, a histogram is drawn representing the frequency of samples in each bin.
- The bins are displayed on the x-axis with labels showing the central value of each bin.
- The y-axis represents the frequency of data in the bins.
- The color of the bins is graduated to create a more visually appealing effect.

No libraries used for this code try my code on the website: [2]

[2] http://wendy.altervista.org/gauss.html

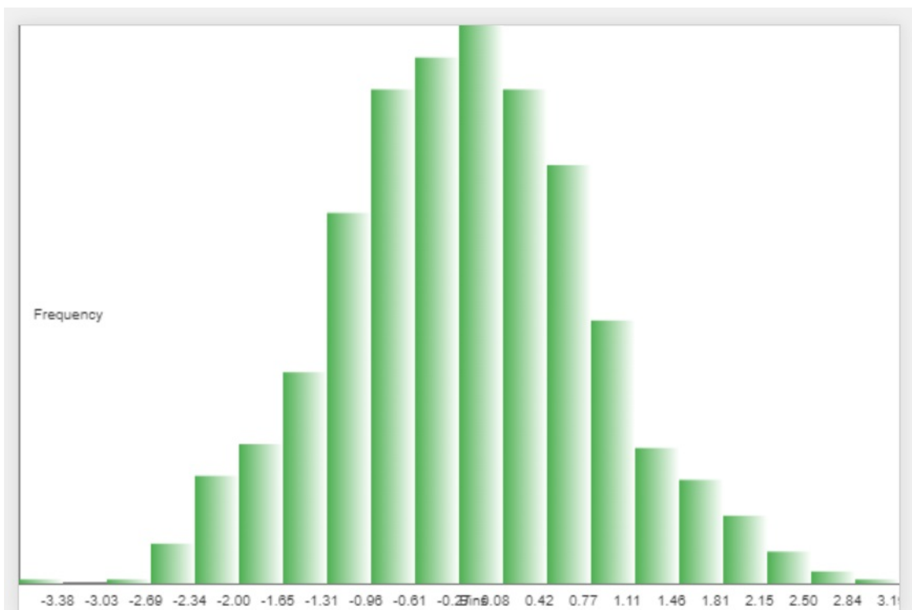

Figure 1: The Gaussian distribution