Th3: The Gaussian Distribution Meaning, Derivations, Simulations

The Gaussian Distribution

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 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 .

The simplest case of a normal distribution is known as the standard normal distribution or unit normal distribution. This is a special case when $\mu = 0$ and $\sigma = 1$, and it is described by this probability density function (or density):

$$\varphi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

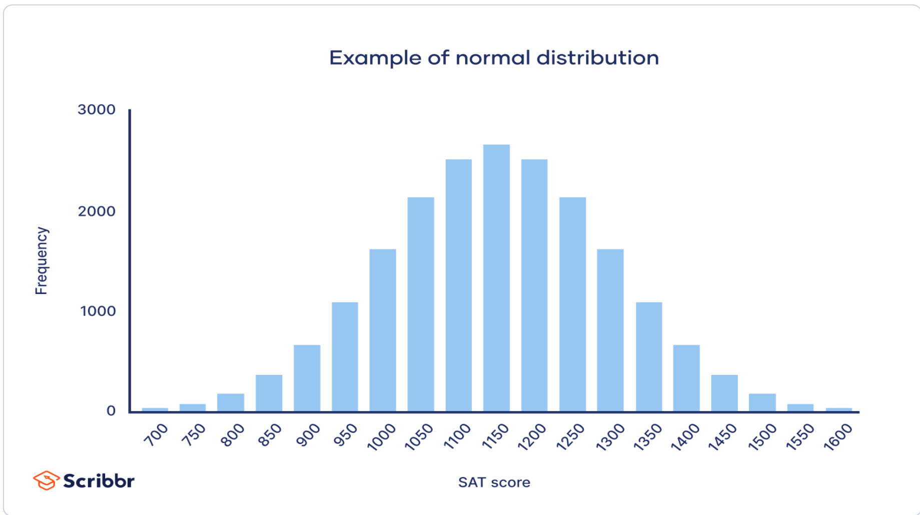
The variable z has a mean of 0 and a variance and standard deviation of 1.

The density $\varphi(z)$ has its peak $1/\sqrt{2\pi}$ at $z = 0$ and inflection points at $z = +1$ and $z = -1$.

A random variable with a Gaussian distribution is said to be normally distributed, and is called a normal deviate.

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.

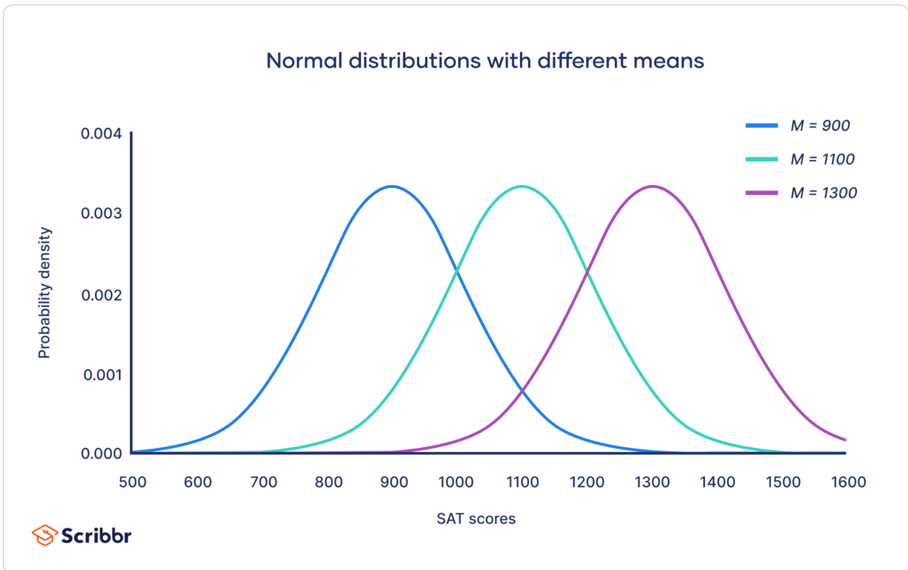
In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the center.



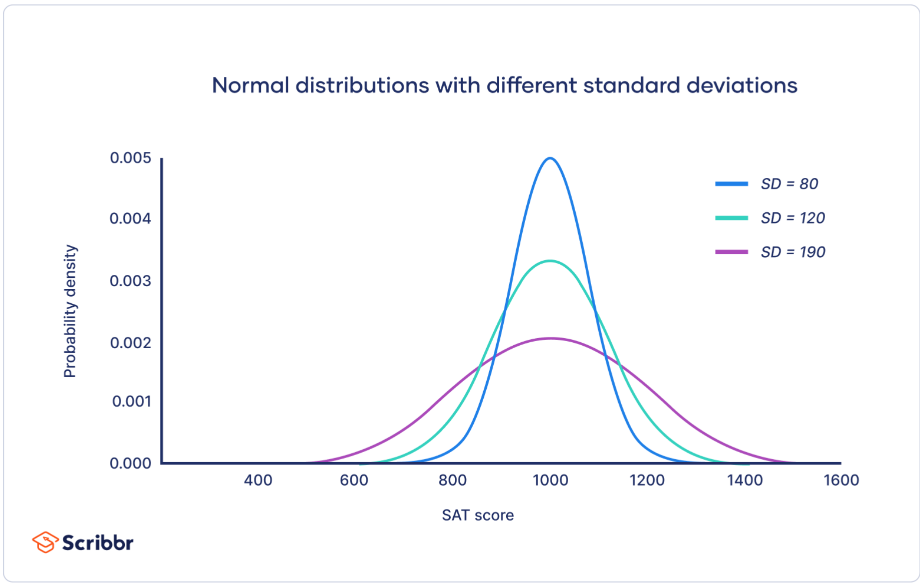
Normal distributions have key characteristics that are easy to spot in graphs:

- The mean, median and mode are exactly the same.
- The distribution is symmetric about the mean, half the values fall below the mean and half above the mean.
- The distribution can be described by two values: the mean and the standard deviation.

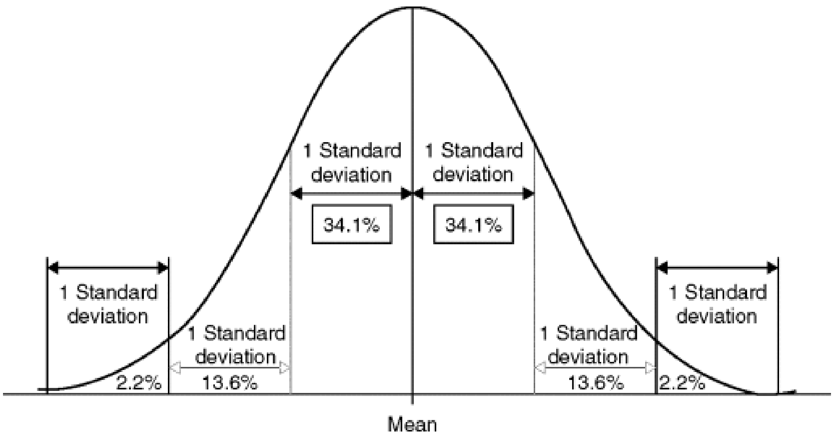
The mean determines where the peak of the curve is centered. Increasing the mean moves the curve right, while decreasing it moves the curve left.



The standard deviation is the scale parameter, so if we increase it the bell will be higher and if we decrease it the bell will be lower.



In a normal curve, approximately 68% of the values fall within one standard deviation of the mean, 95% fall within two standard deviations, and 99.7% fall within three standard deviations.



The derivation of the normal distribution can be found on this page.

The Gaussian distribution occurs in many physical phenomena such as the probability density function of a ground state in a quantum harmonic oscillator. Any particle undergoing diffusion (such as in a mixed liquid) may have its location modeled accurately as a Gaussian distribution as a function of time. Even sepal width of irises have been found to follow a Gaussian distribution.

A Gaussian process is any process in time or space that creates Gaussian distributions within its domain (time , space, etc). They may be used to find non-linear regressions (one problem in machine learning) as well as to reduce dimensionality by identifying which dimensions of a dataset have larger variance and thus may contain more useful information. The Gaussian distribution can be used for example in the following cases:

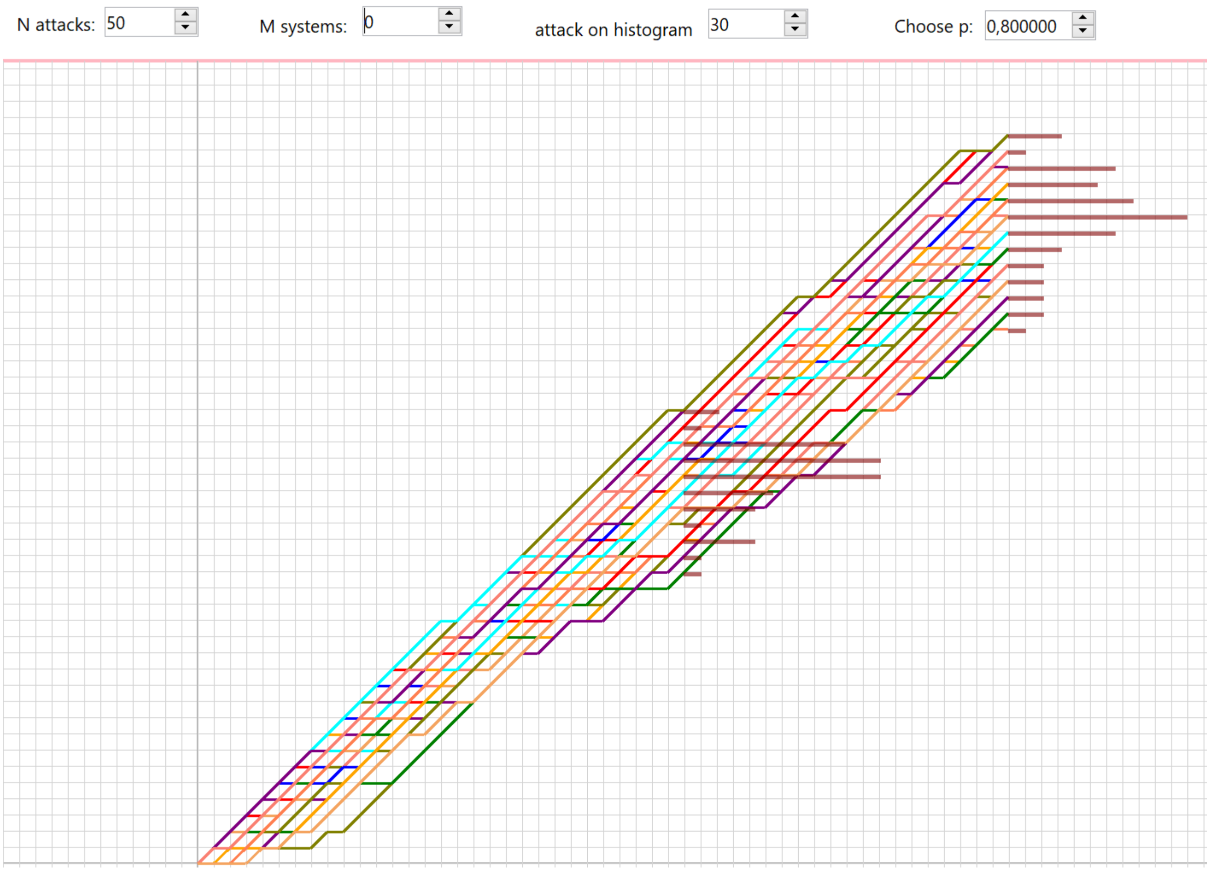
- Designing Standardized Tests: Standardized tests are designed so that test-taker scores fall within a Gaussian distribution.
- Statistical Tests: Many statistical tests can be derived from a Gaussian distribution.
- Quantum Mechanics: A Gaussian distribution can be used to describe the ground state of a quantum harmonic oscillator.

The Gaussian distribution is very useful because:

- It is ever-present as a dataset with finite variance turns into Gaussian as long as the dataset with free feature-probabilities is permitted to raise in size.
- It is the most significant probability distribution in statistics as it turns many natural phenomena such as age, height, test-scores, IQ scores, and sum of the rolls of two cubes and so on.
- Reviews and conclusions resulting from such analysis are intuitive. That is also easy to explain to audiences with basic knowledge of statistics.

Using the simulator on this page is possible to make a simulation of a Gaussian distribution.
To see how Gaussian distribution changes according to its parameters there is a simulator on this page.
Using the previous two simulators is possible to calculate the data and then visualize the Gaussian distribution of the simulation.

One simulation that shows the Gaussian distribution is the one made in homework 3, where there are m systems and n attacks, each attack has a probability p of success. So if we choose a high number for N and M, we can see that the final distribution has the bell shaped form, so it is like a normal distribution.



Bibliography
-https://en.wikipedia.org/wiki/Normal_distribution
-https://www.scribbr.com/statistics/normal-distribution/
-https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/gaussian-distribution
-https://www.probabilitycourse.com/chapter4/4_2_3_normal.php
-https://medium.com/@curiousincosmos/normal-distribution-probability-density-function-derivation-872c4f9d514d
-https://deepai.org/machine-learning-glossary-and-terms/gaussian-distribution
-https://plainenglish.io/blog/gaussian-distribution-in-machine-learning-fa6b2a53ab53
-https://academo.org/demos/gaussian-distribution/

-https://www.mathsisfun.com/data/normal-distribution-simulator.html