Th3: The Gaussian Distribution

Federico Detomaso

1 Meaning

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

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

The parameter μ s 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 independent normal deviates is a normal deviate. Many results and methods, such as propagation of uncertainty and least squares[5] parameter fitting, can be derived analytically in explicit form when the relevant variables are normally distributed.

2 Definitions

2.1 Standard Normal Distribution

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):

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

The variable z has a mean 0 and a variance and standard deviation of 1. The density $\varphi(z)$ has its peak $\frac{1}{\sqrt{2\pi}}$ at $z = 0$ and inflection points at $z = +1$ and $z = -1$.

Although the density above is most commonly known as the standard normal, a few authors have used that term to describe other versions of the normal distribution. Carl Friedrich Gauss, for example, once defined the standard normal as

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

which has a variance of 1/2, and Stephen Stigler once defined the standard normal as

$\varphi(z) = e^{-\pi z^2}$,

which has a simple functional form and a variance of $\sigma = \frac{1}{(2\pi)}$.

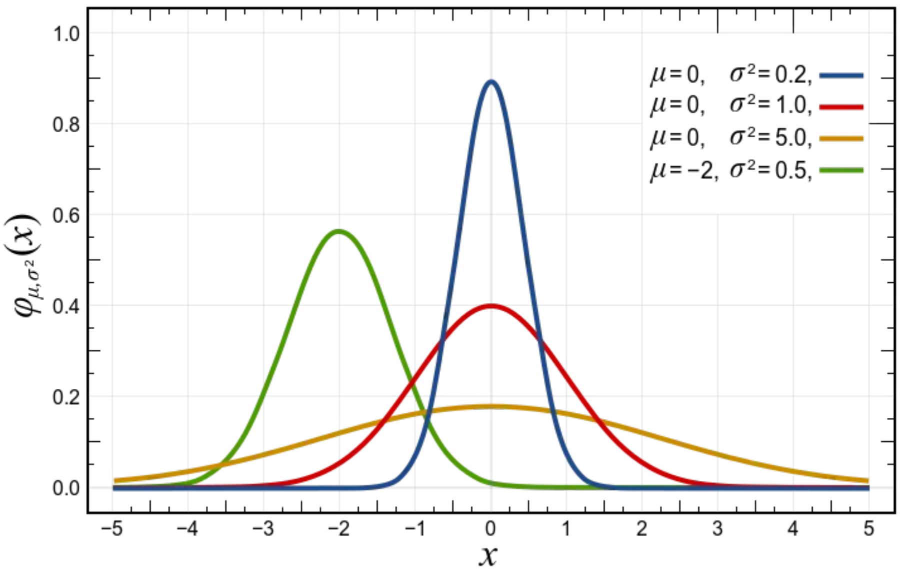
Every normal distribution is a version of the standard normal distribution, whose domain has been stretched by a factor σ (the standard deviation) and then translated by μ (the mean value).

2.2 Notation

The probability density of the standard Gaussian distribution (standard normal distribution, with zero mean and unit variance) is often denoted with the Greek letter ϕ. The alternative form of the Greek letter phi, φ, is also used quite often.

The normal distribution is often referred to as $N(\mu, \sigma^2)$. Thus when a random variable X is normally distributed with mean μ and standard deviation σ, one may write

$X \sim N(\mu, \sigma^2)$.



3 Derivations

The probability density function (PDF) of the standard normal distribution is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

The cumulative distribution function (CDF) of the standard normal distribution is given by:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt$$

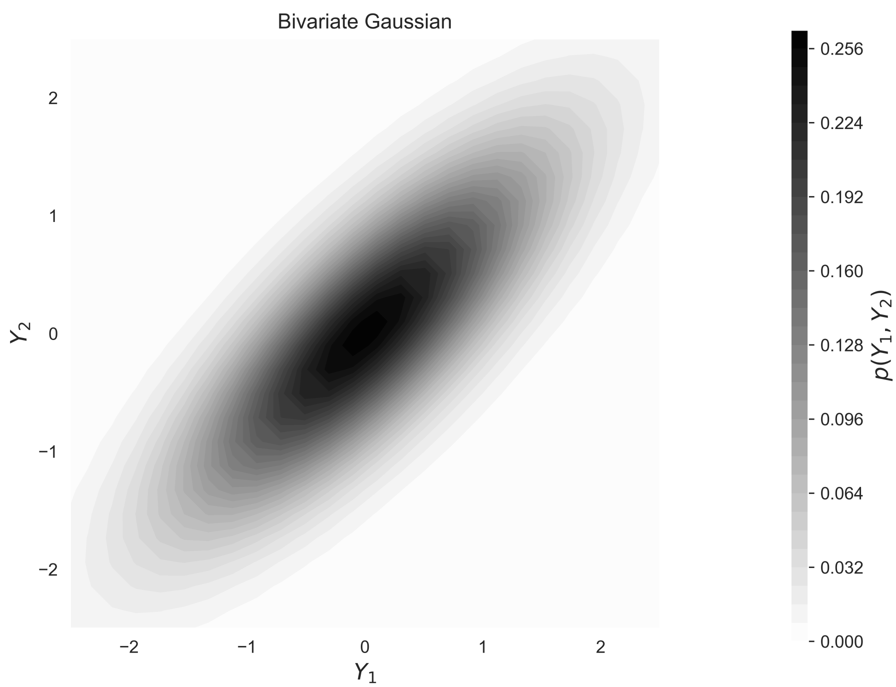
To find the probability density function (PDF), differentiate the CDF with respect to x:

$$f(x) = \frac{d}{dx} \Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Therefore, the PDF of the standard normal distribution is the derivative of its CDF.

4 Simulations

The generating mechanism of the Gaussian distribution is the central limit theorem. The central limit theorem states that the distribution of the sum of many independent samples of a random variable (RV) with finite mean and variance tends to a Gaussian distribution as the number of samples increases. The characteristic bell shape of the univariate Gaussian distribution and the characteristic ellipsoidal contours of the bivariate distribution are well known, see the figure.



Example of bivariate distribution of correlated variables which correlation $\rho = 0.8$.

References

[1] https://en.wikipedia.org/wiki/Normal_distribution

[2] https://towardsdatascience.com/the-gaussian-distribution-explained-19d8417fb047#:~:text=A%20Gaussian%20distribution%20is%20one,returns%20and%20height%20in%20populations.

[3] https://geostatisticslessons.com/lessons/multigaussian