The Gaussian Integral for the Normal Distribution in Machine Leaning

Chinnaraji Annamalai
School of Management, Indian Institute of Technology, Kharagpur, India
Email: anna@iitkgp.ac.in
https://orcid.org/0000-0002-0992-2584

Abstract: The Gaussian distribution is a normal probability distribution that is a key concept in statistics and the backbone of machine learning. In normal distribution, data are symmetrically distributed with no skew. A data scientist needs to know about the normal distribution for data analysis. In this article, we can learn and understand how to use the Gaussian integral in the probability density function for a normal distribution and applications of machine learning.

MSC Classification codes: 60E05, 68Q87

Keywords: computation, gaussian distribution, dataset

1. Introduction

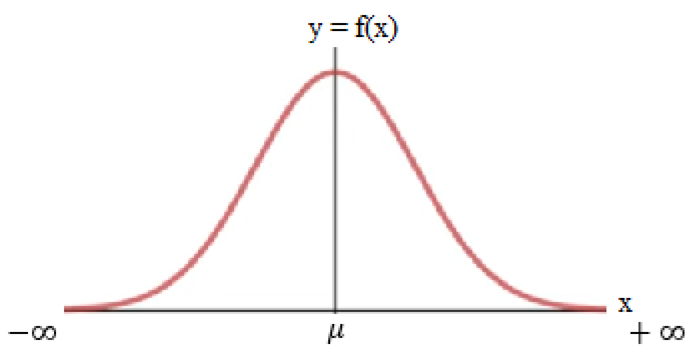
In a normal distribution, a bell-shaped curve is a symmetrical graph that expresses a collection of values about the dataset(distribution). Actually, the normal distribution is a continuous probability distribution wherein values or scores lie in a symmetrical form mostly situated around the mean. The area under the bell curve is equal to 1. In this article, the author introduces a technique in connection with the probability density function for the normal distribution using the Gaussian integral [1-10].

2. Gaussian Integral in Normal Distribution

The Gaussian integral plays an important role in the normal probability distribution with Z-score. Z-score or Z-value in the probability density function for a normal distribution is statistically defined by

$$Z = \frac{(\text{observed value} - \text{mean})}{\text{standard deviation}} \Rightarrow Z = \frac{x - \mu}{\sigma},$$

where x is test value, σ is mean, and σ is standard deviation.
In a normal distribution, the bell-shaped graph is shown below:



The probability density function for a normal distribution is defined using Z-score by

https://doi.org/10.33774/coe-2024-p4j6k **ORCID:** https://orcid.org/0000-0002-0992-2584 Content not peer-reviewed by Cambridge University Press. **License:** CC BY 4.0

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \quad \Rightarrow f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-u)^2}{\sigma^2}} \text{ OR } \quad y = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-u)^2}{\sigma^2}}.$$

Let prove the probability density function of the normal distribution as follows:

Definition of the normal distribution: Data are said to be normally distributed if the rate at which the frequencies fall off is proportional to the distance of the score from the mean and to the frequencies themselves. This statement can be modeled as a differential equation.

$$\frac{dy}{dx} = -k(x-\mu)y, \text{where } k \text{ } is \text{ a possitive constant.}$$

By separating the variables in the above differential equation, we get

$$\frac{dy}{y} = -k(x-\mu)dx$$

Integrating the differential equation on both sides:

$$\int \frac{dy}{y} = \int -k(x-\mu)dx \ \Rightarrow \ln y = \frac{-k(x-u)^2}{2} + c$$

Applying exponentials on both sides:

$$e^{\ln y} = e^{\frac{-k(x-u)^2}{2}+c} \Rightarrow y = e^{\frac{-k(x-u)^2}{2}} e^c$$

Let $e^c = C$, where e^c is a constant.

$$y = Ce^{-\frac{1}{2}k(x-u)^2}. \qquad (1)$$

Since a normal distribution is symmetric about the mean, the area of the bell curve is 1.

Therefore, $\displaystyle\int_{-\infty}^{\infty} Ce^{-\frac{k}{2}(x-u)^2} dx = 1.$

Let $u^2 = \frac{k}{2}(x-u)^2 \ \Rightarrow u = \sqrt{\frac{k}{2}}(x-u); du = \sqrt{\frac{k}{2}}dx \Rightarrow \sqrt{\frac{2}{k}}du = dx$

$$C\sqrt{\frac{2}{k}}\int_{-\infty}^{\infty} e^{-u^2} du = 1$$

Squaring on both sides:

$$\left[C\sqrt{\frac{2}{k}}\int_{-\infty}^{\infty} e^{-u^2} du\right]^2 = 1^2 \ \Rightarrow \frac{2C^2}{k}\left[\int_{-\infty}^{\infty} e^{-u^2} du\right]^2 = 1. \qquad (2)$$

Using the Fubini's theorem, we can solve the above equation

$$\int_{-\infty}^{\infty} f(x,y)dxdy \quad = \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy$$

The Gaussian integral denotes that $\displaystyle\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$

$Proof.$ Let $\displaystyle \mathrm{I} = \int_{-\infty}^{\infty} e^{-x^2} dx = 2\int_{0}^{\infty} e^{-x^2} dx = 2\int_{0}^{\infty} e^{-y^2} dy$

https://doi.org/10.33774/coe-2024-p4j6k **ORCID:** https://orcid.org/0000-0002-0992-2584

$$\mathrm{I}^2 = 2\int_0^\infty e^{-x^2}dx \times 2\int_0^\infty e^{-y^2}dy = 4\int_0^\infty\int_0^\infty e^{-(x^2+y^2)}dxdy$$

Let $y^2 = (xt)^2 \Rightarrow y = xt \Rightarrow t = \frac{y}{x}; dy = xdt. y = 0 \Rightarrow t = 0, y = \infty \Rightarrow t = \infty.$

$$4\int_{t=0}^{t=\infty}\int_{x=0}^{x=\infty} e^{-(x^2+x^2t^2)}\,dx(xdt) = \int_{t=0}^{t=\infty}\int_{x=0}^{x=\infty} xe^{-x^2(1+t^2)}\,dxdt$$

Let $u = x^2(1+t^2); \frac{1}{2(1+t^2)}du = xdx; x = 0 \Rightarrow u = 0; x = \infty \Rightarrow u = \infty.$

$$4\int_{t=0}^{t=\infty}\frac{1}{2(1+t^2)}\int_{u=0}^{u=\infty} e^{-u}\,dudt = 4\int_{t=0}^{t=\infty}\frac{1}{2(1+t^2)}\left[\frac{e^{-u}}{-1}\right]_0^\infty dt = 4\int_{t=0}^{t=\infty}\frac{1}{2(1+t^2)}dt$$

Here, $\left[\frac{e^{-u}}{-1}\right]_0^\infty = -(e^{-\infty} - e^0) = -(0-1) = 1.$

$$4\int_{t=0}^{t=\infty}\frac{1}{2(1+t^2)}dt = 2\int_{t=0}^{t=\infty}\frac{1}{(1+t^2)}dt = 2[\tan^{-1} t]_0^0 = 2(\tan^{-1}\infty - \tan^{-1} 0)$$

Then, $\mathrm{I}^2 = 2(\tan^{-1}\infty - \tan^{-1} 0) = 2\left(\frac{\pi}{2} - 0\right) = \pi \Rightarrow \mathbf{I} = \sqrt{\pi}. \left(\because \mathrm{I} = \int_{-\infty}^{\infty} e^{-x^2}dx\right)$

From the above equation, we obtain the Gaussian integral as follows:

$$\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi} \Rightarrow \int_{-\infty}^{\infty} e^{-u^2}du = \sqrt{\pi}. \quad (3)$$

Substituting the equation (3) in the equation (2), we get

$$\frac{2C^2}{k}\left[\int_{-\infty}^{\infty} e^{-u^2}du\right]^2 = 1 \Rightarrow \frac{2C^2}{k}\left[\sqrt{\pi}\right]^2 = 1 \Rightarrow C^2 = \frac{k}{2\pi} \Rightarrow C = \sqrt{\frac{k}{2\pi}}. \quad (4)$$

Substituting (4) in the equation (1)

$$y = Ce^{-\frac{1}{2}k(x-u)^2} = \sqrt{\frac{k}{2\pi}}e^{-\frac{1}{2}k(x-u)^2}. \quad (5)$$

Substituting $k = \frac{1}{\sigma^2}$ in the equation (5) according to $z = \frac{x-\mu}{\sigma}$, we conclude that

$$y = \sqrt{\frac{1}{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(x-u)^2}{\sigma^2}}.$$

$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, which is the probability function for a normal distribution.

3. The Gaussian Distribution in Machine learning

In data science and machine learning, the Gaussian distribution plays a vital role in various aspects, particularly in probabilistic modeling and statistical inference. Actually, machine learning is a branch of artificial intelligence focused on enabling systems to learn from data, uncover patterns, and autonomously make decisions. In today's era dominated by data, the machine learning is transforming industries ranging from healthcare to finance, offering robust tools for predictive analytics, automation, and informed decision-making.

,

4. Conclusion

In this article, author has well used the Gaussian integral for the normal distribution in machine learning. The normal distribution, also known as the Gaussian distribution, is a key concept in statistics and the backbone of machine learning. In normal distribution, data are symmetrically distributed with no skew. A data scientist needs to know about the normal distribution for data analysis.

References

[1] Annamalai, C. (2023) Factorial Theorem: An Alternative to Gamma Function. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4376857.

[2] Annamalai, C. (2022) Factorials, Integers, and Factorial Theorems for Computing and Cryptography. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4189241.

[3] Annamalai, C. (2023) Factorial Theorem for Computing the Factorial of Positive Real Number. *COE, Cambridge University Press.* https://dx.doi.org/10.33774/coe-2023-7p472.

[4] Annamalai, C. (2022) Combinatorial Theorems in Factorials with Multinomial Computation. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4185700.

[5] Annamalai, C. (2023) Review on the Gamma Function and Error Correction. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.4397585.

[6] Annamalai, C. (2023) Gamma Function derived from Factorial Function based-Pi Function. *COE, Cambridge University Press*. https://dx.doi.org/10.33774/coe-2023-wv4rc.

[7] Annamalai, C. (2023) Gamma Function derived from Natural Logarithm and Pi Function. *COE, Cambridge University Press*. https://dx.doi.org/10.33774/coe-2023-wv4rc-v3.

[8] Annamalai, C. (2023) An Alternative Method for the Gamma Function derived from Natural Logarithm and Pi Function. *COE, Cambridge University Press*. https://doi.org/10.33774/coe-2023-wv4rc-v4.

[9] Annamalai, C. (2023) Computation of the Gamma Function of Half using the Gaussian Integral. *Zenodo.* https://doi.org/10.5281/zenodo.13849817.

[10] Annamalai, C. (2004) The Gaussian Integral. *The ResearchGate.* https://doi.org/10.13140/RG.2.2.10969.33122.