Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/38/20240595

The Gaussian distribution: Derivation method and its applications in selected examples

Shuhao Chen

Xiaoshan No.2 Senior High School, Xiaoshan, Zhejiang, 311251, China

100914@yzpc.edu.cn

Abstract. Gaussian distribution is the most important and common form of distribution in statistical theory, and its density function has more complicated forms. When Gauss dealt with the problem of measurement error, he derived the form of normal distribution from another angle, and proposed the least squares theory based on the normal distribution of error. This helps solve the problem of the probability density distribution of the error, enabling people to make a better statistical measure of the effect of the size of the error. In short, Gaussian distribution is not only a powerful mathematical tool, but also a connection between theory and practical applications. It is also a bridge between theory and practical application. In practice, the Gaussian distribution has a wide range of applications, in the natural sciences, engineering and social sciences and other fields, the Gaussian distribution is used to describe the continuity of random variables, such as measurement error, temperature change, population intelligence level. In this paper, the author briefly deduces the Gaussian distribution derivation process, from the overall and the measurement error of the simple random samples to cut in, to admit that $\bar{x}$ is already the estimates that should be taken. Hopefully, it will be helpful to scholars who are interested in this area.

Keywords: Gaussian distribution, random walk, probability density distribution.

1. Introduction

The Gaussian distribution, also known as the normal distribution, is the most important form of distribution in mathematical and statistical theory. It was proposed and studied by the German mathematician Carl Friedrich Gauss. The Gaussian distribution can be traced back to the study of the binomial distribution. In his study of error theory, Gauss found that many errors conformed to a linear distribution like a bell-shaped distribution, with a high middle and two low sides. With the rapid development of statistics and probability theory, the Gaussian distribution has become a very classic and influential probability distribution model. Its main characteristics of symmetry, concentration, variability and asymptotic can lead to a more realistic probability distribution. In the euro before the German 10 mark banknote not only printed on the Gaussian head, but also printed Gaussian distribution (μ, σ^2) density curve, can be seen Gaussian influence is very big. Gaussian distribution is very friendly to beginners, and it has the largest entropy when the mean and variance are given. According to the central limit theorem, many independent random variables approximately obey the Gaussian distribution. It is the basis of parametric testing and confidence intervals, which is more important for statistical inference.

DOI: 10.54254/2753-8818/38/20240595

The Gaussian distribution, at its most basic, simplifies complex problems. Because of the ease of calculating and applying the normal distribution, it is very useful for constructing predictive models and interpreting data. It is useful to understand the underlying characteristics of a variable, which is useful for accurate estimation. The normal distribution also facilitates the handling of multivariate data. The symmetry, concentration, variability, and asymptotic of the multivariate Gaussian distribution in the multivariate context make it possible to efficiently handle and analyses multidimensional data.

Gaussian distribution has a notable role in several fields. One can use Gaussian mixture models to better select good optimal control strategies for the current microgrid operation optimization, and to propose a real-time scheduling model for the operation of microgrids that considers both the economy and security of the microgrid [1]. Speaking of the bounding box is converted to Gaussian probability distribution, and then based on the comprehensive consideration of the three elements of the bounding box regression, a new loss function can be proposed: non-Gaussian squared distance [2]. On the framework of the sparse Gaussian process, a representative detection data selection strategy is introduced. Based on its strategy, the algorithm complexity is reduced. With the help of the Maanshan Yangtze River Bridge example, it becomes possible to apply Gaussian process to long-term structural real-time health monitoring and so on.

The Gaussian distribution can be extended into many branches, and the scope of application is very broad. In this paper, the most basic derivation of Gaussian distribution and two real-life applications of Gaussian distribution are obtained from the limited information available. At present, the very authoritative academic literature on the internet library of proof material in this regard is also very little. This aroused people's interest, and it also need more people to improve as well as to explore. The core significance of this issue is far more than the understanding and mastery of the theory itself. The more is in the Gaussian core statistical ideas of deep understanding, learning and experience of his extraordinary research perspective and methodology. It will appropriately depart from the traditional way of thinking, perhaps this is a higher theoretical and academic value.

2. Derivation of Gaussian distribution

To begin with, the following three Lemmas are useful to derivative the Gaussian distribution.

Lemma 1: If the function $g(x)$ is an even function with a second order derivative, then $g'(x)$ is an odd function and $g''(x)$ is again an even function.

Lemma 2: If the function $g(x)$ satisfies the following conditions: a) $g(0) = 0$, b) $g(x)$ is derivable and the derivative function is continuous, c) $g'(x)$is even function, d) For any natural number m and real number x satisfying $g'(mx) = g'(x)$, then for any real number x, the function $g(x)$ must have the form $g(x) = cx$ (where c is a constant).

Lemma 3: The integral value of the function e^{-x^2} over the whole domain of real numbers is π, namely, $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$.

The original derivation of Gaussian density function is motivated by an awareness of the law of error. Density function always makes the arithmetic mean of the sample becomes the best of the overall truth estimates. Let the overall truth value be θ, there is a simple random on measurement error $x_1, x_2, \cdots x_n$. Consider the estimator that maximizes the following likelihood function

$$L(\hat{\theta}) = \max_{\theta} L(\theta) = \max_{\theta} f(x_1 - \theta) f(x_2 - \theta) \cdots f(x_n - \theta). \quad (1)$$

Unlike the idea of maximum likelihood estimation given the overall density function $f(x)$, what value of θ maximises the above equation. This value will be an estimate of the parameter (average of) θ being sought.

The density function of the error to be determined is written in terms of $f(x)$. According to the condition that in the absence of systematic errors, the measured value is always the one which is further away from the true value. The smaller the probability of its occurrence, and it should be symmetrically distributed to the left and right of the true value. Therefore, the density function of the error $f(x)$ should

DOI: 10.54254/2753-8818/38/20240595

be a continuous even function. In addition, it can be allowed to have further excellent mathematical properties, such as derivative functions with second-order continuity.

Mathematically, the issue of $\max_{\theta} f\ (x_1-\theta)f(x_2-\theta)\ldots f(x_n-\theta)$equates to following problem $\max_{\theta} \ln[\ f\ (x_1-\theta)f\ (x_2-\theta)\ldots f(x_n-\theta)]$. Therefore, For the equation $\max \sum_{I=1}^{n} \ln\big(f(x_i-\theta)\big)$ to take a large value, one must have

$$\sum_{I=1}^{n} \ln'\theta\big(f(x_i-\theta)\big)=\sum_{i=1}^{n}\frac{f'\theta(x_i-\theta)}{f(x_i-\theta)}\ . \tag{2}$$

Let quote function $g\ (x)=\frac{f'(x)}{f(x)}$, then $\sum_{i=1}^{n} g(x_i-\ \theta)=0$. It is easy to get $g\ (x)$ form, by using the even function property of $f(x)$. Applying the conclusion of Lemma 1, one can show that $g\ (x)$ is an odd function. So, there are $g\ (x)=-g\ (-x)$, $g\ (0)$ = 0. Taking natural numbers m, and $n=m+1\ , so\colon x_1=x_2=\cdots=x_m=-x, x_{m+1}=mx$. This time $\hat{\theta}=\bar{x}=0$ due to $\sum_{i=1}^{n} g\big(x_i=\hat{\theta}\big)=\sum_{i=1}^{n} g(x_i)=0$ and using the previous equation yields $g(mx)=mg(x)$. The above equation holds for all natural numbers m and real numbers x. From this, assuming that $g(x)$ is derivable and derivative function is continuous, both sides of the equation are derived with respect to X respectively, one must

$$\frac{dg(mx)}{g(mx)}*\frac{d(mx)}{dx}=m\frac{dg(x)}{dx}. \tag{3}$$

Then $g'(mx)=\ g'(x)$ because of $g'(x)=\frac{f''(x)f'(x)-[f'(x)]^2}{[f'(x)]^2}$. Since $f(x)$ is even function, based on lemma 1, $f'(x)$ is odd function and $f''(x)$ is even function, hence $g'(x)$ is even function. Since $f(x)$ has a second order continuous derivative function, so $g'(x)$is continues. Then, $g(x)$ satisfies all the conditions of Lemma 2. Applying its conclusion yields $\frac{f'(x)}{f(x)}=cx$ and $g(x)=cx$. Because of $\frac{f'(x)}{f(x)}=\frac{d[ln(f(x))]}{dx}$, thus $ln\big(f(x)\big)=\int cxdx\ =\frac{1}{2}cx^2+c'$.so

$$f(x)=e^{\frac{1}{2}cx^2+c'}=Me^{\frac{1}{2}cx^2}\ \big(M\triangleq e^{c'}\big). \tag{4}$$

The above $f(x)$ is obviously always greater than 0. In order to find a density function, it must also have an integral value of 1 over the entire real number field. thus there is $\int_{-\infty}^{\infty} Me^{\frac{1}{2}cx^2}dx=1$ obviously, and c must be a constant less than zero. Writing $c=-\frac{1}{\sigma^2}(\sigma>0)$, the above equation becomes $\int_{-\infty}^{\infty} Me^{-\frac{1}{2\sigma^2}x^2}dx=1$. Applying the conclusions of lemma 3, $\int_{-\infty}^{\infty} e^{-x^2}dx=\sqrt{\pi}$. Let $y=\frac{1}{\sqrt{2}\sigma}$, then $dx=\sqrt{2\sigma}$. Hence, it yields $1=\int_{-\infty}^{\infty} Me^{-\frac{1}{2\sigma^2}x^2}dx=M\int_{-\infty}^{\infty} e^{-y^2}\big(\sqrt{2}\sigma\ dy\big)=M\big(\sqrt{2}\sigma\big)\int_{-\infty}^{\infty} e^{-y^2}dy=M\big(\sqrt{2}\sigma\big)\sqrt{\pi}$. So, there is $M=\frac{1}{\sigma\sqrt{2\pi}}$. Thus, the Gaussian distribution density function has the form

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{1}{2\sigma^2}x^2}. \tag{5}$$

3. Applications

3.1. Application 1: Representative monitoring data selection strategy

When using civil engineering structures, some harsh environmental changes as well as some uncontrollable factors can cause structural damage or even ruptures. Structural health monitoring technology can be realized through sensing devices and related algorithms to assess the condition of

DOI: 10.54254/2753-8818/38/20240595

bridges and monitor damage [3]. Environmental influences such as wind, temperature, traffic, humidity, and solar radiation can also cause corresponding changes in the bridge structure. They directly affect the performance of vibration-based damage diagnostics at various levels, such as early damage monitoring damage monitoring, damage localization and damage quantification [4]. The effects here need to be effectively considered in the structural condition assessment, otherwise it may cause bias or even misjudgment of the structural assessment results [5]. Based on the monitoring data of Donghai Bridge for 6 years, Zhou and Sun investigated the effects of traffic loads, wind loads, and structural temperature on the modal frequency changes in different time scales, and gave explanations at the mechanical level [6]. Based on the monitoring data of Tsing Ma Bridge during the typhoon period, Wang et al. used a variationally anisotropic Gaussian process to fit the main beam strain and anisotropic noise changes in typhoon conditions better and heteroskedastic noise variations under typhoon conditions [7].

Among these methods, Gaussian process is an effective method for constructing this type of model and is widely used in structural condition assessment [8,9]. Due to the properties of the probabilistic framework, it can consider to take account uncertainty in the data and give predictions based on probability distributions. Based on the Gaussian process to representation monitoring data selection strategy, it can reduce the amount of training data when using the Gaussian process algorithm to fit the effects of multiple environmental factors in bridge structures, improve the efficiency of the valuation, and this means significantly reduces the complexity of the algorithm.

The representative monitoring data selection strategy mentioned in this paper is mainly based on the Gaussian process development. The next step will be introduced the main theory of Gaussian process. The structural monitoring data y can be expressed as $y = f(x) + \varepsilon$ where x is the environmental monitoring variable, $f(x)$ is the structural response induced by environmental changes, and ε is the model error. The mapping relationship between the environmental variables x and $f(x)$ is fitted by a Gaussian process model [10].

Gaussian process is a stochastic process defined by a collection of random variables, and any number of this elements can form a joint Gaussian distribution. In other words, a Gaussian process is a generalization of the multivariate Gaussian distribution, while retaining many of the mathematical properties of the multivariate Gaussian distribution. A Gaussian process can be represented by the mean function $m(x)$ and the covariance function $k(x, x')$. Here, $m(x) = E[f(x)]$, $k(x, x') = \left[\left(f(x) - m(x)\right)\left(f(x') - m(x')\right)\right]$. At this point, the Gaussian process can be expressed as the form of $f(x) \sim gp\left(m(x), k(x, x')\right)$. Here, the author introduces the hyperparameters Ψ that adjust $m(x)$ and $k(x, x')$. When the hyperparameters are determined, the probability density function for a given position x can be given by the following equation $p(x|\Psi) = \mathcal{N}(m, k)$ [11].

In the training phase of this method, the structure is considered healthy and there are no structural changes. In this case, the modelling error is mainly caused by sensor noise, which can be considered as an independent and identically distributed Gaussian white noise. Its probability density function can be given by the following equation $p(\varepsilon|\sigma^2) = \mathcal{N}(0, \sigma^2 I)$, where σ^2 is the variance of the noise. The hyperparameters of the Gaussian model can be obtained by maximizing the marginal likelihood function, which is given by the following equation $p(y|x) = \mathcal{N}(m(x), K(x, x) + \sigma^2 I)$, which is

$$p(y|x) = (2\pi)^{-\frac{N}{2}}\left|K(x,x) + \sigma^2 I\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[y - m(x)]^T\left[K(x,x) + \sigma^2 I\right]^{-1}[y - m(x)]\right\}. \quad (6)$$

Given the training data X and the corresponding noisy observations y and the input X_*, the posterior distribution of the Gaussian process prediction f_* is given by the following equation $f_*|X, y, X_* \sim N(\mu_{f_*}, C_{f_*})$, in which $\mu_{f_*} = K(X_*, X)[K(X, X) + \sigma^2 I]^{-1} y$ and $C_{f_*} = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma^2 I]^{-1} K(X, X_*)$ [12].

The resulting f_* can be regarded as the predicted value of the structural response due to external excitation, and its difference ε from the observed value y, is model error, can be used as an indicator of the health of the hand model. As mentioned earlier, in healthy condition, the model error consists only of sensor noise. When the bridge encounters sudden conditions, cumulative damage or sensor failures,

DOI: 10.54254/2753-8818/38/20240595

the model error shows some new characteristics. On other words, it exhibits a non-Gaussian distribution nature, which corresponds to the model as the monitored values exceed the confidence interval given by the model predictions. This type of phenomenon can be used to determine the true condition of the bridge. The methodology is applicable to generic environmental factor-structural response correlation modelling.

3.2. Application 2: Mixed Gaussian noise sparse representation model

Sparse Representation technology uses the guidance of a priori information in the model to make the signal energy highly aggregated, and then it can detect the weak characteristic information generated by the early failure of machinery. Recently, many scholars have applied the sparse theory to the fault diagnosis of mechanical equipment [13]. He et al. based on the kinetic response mechanism of the smooth and shock modulation, and established a modulation dictionary, and realized the fault diagnosis of gearboxes [14]. Diwu and coworkers proposed a collaborative two-periodic group sparse diagnostic model based on the a priori knowledge that the bearing rolling body passes through a localized damage region to provoke the resonance frequency of the system [15]. The classical sparse representation theory is built on the assumption that the disturbance noise satisfies the Gaussian distribution law of independent homogeneous distribution. However, in the context of modern engineering applications, the complex interference components make the noise distribution of vibration signals present a mixed distribution form of multiple components, and thus, it is difficult for the classical sparse representation model to obtain useful information directly from the above mixed distribution of interference noise.

For the observed signal y, if the useful signal has sparsity in the small wavelet domain and the sparsity coefficients are here the coefficients x in the small wavelet domain, then the noise component $r = y - Dx$ in the observed signal. Here, D is the wavelet inverse transform matrix, and the bd4 wavelet is introduced as the wavelet transform basis function. All its components $r_i(i = 1,2,\cdots\cdots N)$ are independently and identically distributed in a one only mixed Gaussian

$$r_i \sim M(K, \prod, M, \sum), \tag{7}$$

where each component r_i of the noise signal component is independent of each other. It obeys the same distribution form, K denotes the number of Gaussian sub-distributions, $\prod = \{\pi_1, \pi_2 \cdots \pi_k\}$ is the mixing ratio parameter of each Gaussian sub-distribution only. In addition, $\prod = \{\mu_1, \mu_2 \cdots \mu_k\}$ is the mean of each Gaussian sub-distribution. In order to simplify the problem, it can be assumed that the noise component as well as the mean value of each sub-distribution is zero, $\sum = \{\sigma_1^2, \sigma_2^2 \cdots \sigma_k^2\}$ is the Gaussian sub-distribution of each the variance of each Gaussian sub-distribution. Its probability density function is

$$P(r_i) = \sum_{k=1}^{k} \pi_k p_k(r_i|\mu_k, \sigma_k^2). \tag{8}$$

In this formula, $\sum_{k=1}^{k} \pi_k = 1; p_k(r_i|\mu_k, \sigma_k^2)$ is the probability density function of each Gaussian subdistribution, μ_k is the mean, σ_k is the standard deviation. Since maximum a posteriori estimation and sparse representation go in the same direction, the logarithmic form of the likelihood function is first computed from the underlying probability

$$\log P(y|x) = \log\left(\prod_{i=1}^{n} p_k(r_i|\mu_k, \sigma_k^2)\right). \tag{9}$$

Due to the different parameters of each sub-distribution of the mixed Gaussian distribution, it is necessary to determine that each sample belongs to each Gaussian sub-distribution, otherwise the above equation cannot be computed. Therefore, a hidden variable γ needs to be introduced to categorize the belonging of the samples to form a usable full data set $(y_i, \gamma_{i,1}, \gamma_{i,2} \cdots \gamma_{i,k})\ i = 1,2 \cdots n$. In this formula,

DOI: 10.54254/2753-8818/38/20240595

if y_i generated by the t-th Gaussian distribution, then $\gamma_{i,t} = 1, \gamma_{i,k\neq t} = 0$. Thus, the new log-likelihood function is

$$\log P(y|x) = \log\left(\prod_{i=1}^{n} p_k\left(r_i|\mu_k, \sigma_k^2\right)\right) = \sum_{i=1}^{n} \log p_k\left(r_i|\mu_k, \sigma_k^2\right) = \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} p\left(r_i, \gamma_{i,k}|\mu_k, \sigma_k^2\right)\right]. \quad (10)$$

According to Jensen's inequality, the log likelihood of the observed data has the following inequality,

$$\log P(y|x) = \sum_{i=1}^{n} \log\left[\sum_{k=1}^{k} \frac{Q\left(\gamma_{i,k}\right)}{Q\left(\gamma_{i,k}\right)} p\left(r_i, \gamma_{i,k}|\mu_k, \sigma_k^2\right)\right] \geq \sum_{i=1}^{n}\sum_{k=1}^{K}\left[Q\left(\gamma_{i,k}\right)\log\frac{p\left(r_i, \gamma_{i,k}|\mu_k, \sigma_k^2\right)}{Q\left(\gamma_{i,k}\right)}\right], \quad (11)$$

where $Q\left(\gamma_{i,k}\right)$ is the probability distribution of the hidden variable $\gamma_{i,k}$. The right-hand side of the above inequality is denoted as $L(\Omega)$ for simplicity, where $\Omega = \{x, \prod, \sum\}$.

According to the above, since the useful signal is sparse in the wavelet domain, the prior distribution is set to the Laplace distribution, while the overall sparse framework under the Bayesian perspective can be written according to $\hat{x}^{MAP} arg\max_x P(x|y) = arg\max_x P(y|x)P(x)$:

$$\hat{x} = \max_x\left\{\log\left(L(\Omega)\right) + \log\left(\prod_{i=1}^{n} P(x_i)\right)\right\}, \quad (12)$$

where $P(x_i) = \frac{1}{2\sigma_x} exp\left(-\frac{|x_i|}{\sigma_x}\right)$ is the probability density function of the Laplace distribution.

4. Conclusion

In this paper, the derivation of the Gaussian density function is carried out, and then two real-life examples are cited. In civil engineering, the key structural and environmental parameters of bridges require long-term on-site monitoring and systematic analysis, and their Gaussian distributions can be used in the strategy of representative inspection data selection for better judgement of the real condition of bridges. In addition, sparse representations of mixed Gaussian noise can be used to better fit experimental observed. The sparse representation of hybrid Gaussian noise can better fit the noise distribution in the experimentally observed signals. Obviously, Gaussian distribution can be used in many areas of life, and the function is very basic and powerful. One goal that this paper achieves is presenting a complete and clear derivation of the Gaussian distribution. To this end, the Gaussian distribution can be used in the field of complete examples and one can write a clear specific for which step. Therefore, the author hopes that this article can give a general understanding and inspiration to those beginners who are new to Gaussian distribution.

References

[1] Li L S, Feng W T, Pan K J, Zheng Y X, Deng B Y, Jing Z Y. Diwu Z K, Cao H R, Wang L. (2024). Study of real-time control strategies for microgrids considering prediction uncertainty. Sichuan Power Technology, 47(1), 22-26.

[2] Li R, Li Y. (2024). Target detection based on nonlinear Gaussian squared distance loss. Journal of applide sciences- Electronics and Information Engineering, 42(1),1-12.

[3] Sun L, Shang Z, Xia Y, et al. (2020). Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. Journal of Structural Engineering, 5, 04020073.

[4] Zhang C, Mousavi A, Masri S F, et al. (2022). Vibration feature extractio using signal processing techniques for structural health monitoring: A review. Mechanical Systems and Signal Processing, 177, 109175.

DOI: 10.54254/2753-8818/38/20240595

[5] Sohn H. (2007). Effects of environmental and operational variability on structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851): 539-560.
[6] Zhou Y, Sun L. (2019). Effects of environmental and operational actions on the modal frequency variations of a sea-crossing bridge: A periodicity perspective. Mechanical Systems and Signal Processing, 131: 505-523.
[7] Wang Q A, Zhang C, Ma Z G, et al. (2022). Modelling and forecasting of SHM strain measurement for a large-scale suspension bridge during typhoon events using variational heteroscedastic Gaussian process. Engineering Structures, 251, 113554.
[8] Rasmussen C E, Williams C K I. (2006). Gaussian processes for machine learning. Cambridge, MA: MIT press.
[9] Wang J. (2024). An intuitive tutorial to Gaussian processes regression. IEEE Computing in Science & Engineering, 25(4), 4-11.
[10] Zhang Y M, Wang Hao, Mao J X. (2022). Sparse Gaussian process regression for predicting the typhoon-induced response of long-span bridges. China Civil Engineering Journal, 55(10), 72-79.
[11] Li Y, Ding Y, Zhao H, et al. (2022). Data-driven structural condition assessment for high-speed railway bridges using multi-band FIR filtering and clustering. Structures, 41, 1546-1558.
[12] Petersen W, Øiseth O, Lourens E. (2022). Wind load estimation and virtual sensing in long-span suspension bridges using physics-informed Gaussian process latent force models. Mechanical Systems and Signal Processing, 170, 108742.
[13] Zhang H, Chen X F, Du Z H, et al. (2016). Nonlocal sparse model with adaptive structural clustering for feature extraction of aero-engine bearings. Journal of Sound & Vibration, 368, 223-248.
[14] He G L, Ding K, Lin H B. (2016). Fault feature extraction of rolling element bearings using sparse representation. Journal of Sound & Vibration, 366(31): 514-527.
[15] Diwu Z K, Cao H R, Wang L, et al. (2021). Collaborative double sparse period-group lasso for bearing fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 70, 3507110.