10.06.2024

Kubilay Kağan Çapak
Institute of Applied Mathematics
Middle East Technical University
Dumlupınar Boulevard, No:1
Çankaya/ ANKARA, Turkey, 06800
kubilay.capak@metu.edu.tr

Project Name: The Normal Distribution And Its Applications

THE REPORT OF THE PROJECT

Introduction

The origin of the normal distribution was introduced by Abraham de Moivre in 1733. He studied the probability of coin flips. Then, he derived a bell shaped distribution, which can be called as a normal curve, whose has many applications of measurements such as the work of observing the movements of planets and stars. Although he worked out the areas under the normal curve and the formula for it, it can be popularized by other mathematicians that are Pierre Laplace, Carl Gauss, Galileo, Pearson, Fisher and Quételet.

The normal distribution is one of the most significant and helpful probability distributions in statistics, natural sciences and the other fields because of its theoretical properties and useful applications. Also, it is more popular and better measure than other probability distributions such as Uniform and Poisson. Then, it is continuous probability distribution which is characterized by its bell-shaped curve. There are several properties of it.

1. The normal distribution is symmetric about the mean μ, which is the peak of the distribution. In fact, the peak of the curve corresponds to the mean of the distribution.

2. In the normal distribution, the three measures of the central tendency, which are mean, median, and mode, are equal and located at the center of the distribution.

3. The spread or the width of the normal distribution is determined by the standard deviation σ. The variance is the square of the standard deviation σ^2. The larger standard deviation σ we have, the wider and flatter curve we can obtain.

4. The tails of the normal distribution approach to the horizontal axis, but they never touch it. This implies that the probability of extreme values is never zero even though it becomes increasingly small. That is, the tails of the normal distribution extend infinitely in both directions without touching the horizontal axis, which shows that every value has a non-zero probability although it becomes infinitesimally small as it moves away from the mean.

5. According to the Empirical Rule (68-95-99.7 Rule) in the normal distribution, approximately 68% of the data falls within one standard deviation from the mean ($\mu \pm \sigma$), about 95% of the data within two standard deviations ($\mu \pm 2\sigma$), and about 99.7% of the data within three standard deviations ($\mu \pm 3\sigma$).

Also, the probability density function of the normal distribution depends on the mean of the distribution μ, the standard deviation σ, the variance σ^2, the constant pi π = 3.14159… and the base of the natural algorithm e = 2.71828… because of the fact that the formula of it is given by:

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

The pth percentile of the data set of the measurements is the value for which at most p% of them are less than that value, and at most (100-p)% of them are above it when we order the values of the set from smallest to largest in the normal distribution.

The normal distribution provides a close approximation to the Binomial when the number of the trials n is large, and the probability of success on the given trial p or 1-p is small. In this case, we can use the normal distribution to approximate the probabilities, which is related to the binomial distribution. Then, the mean and the standard deviation of X are μ = np and σ^2 = np(1-p), respectively. Also, 5 ≤ np and 5 ≤ n(1-p) in this criteria.

The normal distribution has a skewness of zero, where it measures the degree of symmetry of a distribution. Furthermore, it has a kurtosis equal to 3, where it measures the combined weight of the tails of the distribution that is relative to the center of the curve. Also, it has the same kurtosis as mesokurtic distributions.

In addition, the normal distribution is widely used due to the central limit theorem because it plays a crucial and important role on the theorem. It states that the sum of many independent and identically distributed random variables will approximate a normal distribution regardless of the original distribution. In fact, it states that the distribution of the sum of a large number of independent and identically distributed variables will tend to be normal regardless of the shape of the original distribution. The theorem underlies many statistical procedures and justifies the use of the normal distribution in practice. In addition, the theorem plays an important role in sampling distribution, which is beneficial and advantageous for us to estimate and find the proportion of the population. This enables us to give information and make inferences about the population more easily.

The Normal Distribution In Practice

1. The normal distribution with the mean of 0 and the standard deviation of 1 is called the standard normal distribution. The values from the normal distribution can be transformed to the standard normal distribution by using the z-score formula z = (x – μ)/ σ, where μ is the mean of the distribution, σ is the standard deviation and x is the data point. The transformation converts any normal distribution to the standard normal distribution, which enables us to compare between different normal distributions by standardizing the values and analyze them more easily.

2. The probabilities for the given range of values can be computed by using standard normal distribution tables that provide us with thecumulative distribution functions.

3. Normal distribution assumptions are necessary for doing some hypothesis tests, which helps us to determine confidence intervals, and making predictions about sample data.

Some Applications of The Normal Distribution

1. Many statistical methods assume that data follows the normal distribution. For instance, t-tests and z-tests rely on the assumption.

2. In finance, asset returns, portfolio risks, and other financial variables are assumed to follow the normal distribution that helps in risk assessment and portfolio management.

3. In manufacturing, the normal distribution helps in monitoring process variations, ensuring quality control and process management by analyzing variations in product dimensions and other attributes.

4. Some natural phenomena such as the heights of people, measurement errors in physical experiments, and test scores tend to follow the normal distribution.

The normal distribution serves some fundamental purposes in statistics and various applied fields. Below is a detailed examination of some aims of the normal distribution:

1. The normal distribution plays an important role in modeling natural phenomena. Many biological, physical and social phenomena show the distribution of values that approximate a normal distribution. Its aim is to provide a mathematical model that accurately and properly describes the distribution of real-world measurements. For example, human characteristics such as heights of individuals, blood pressure readings, IQ scores generally follow the normal distribution. Also, it provides us with mathematical framework to represent the naturally occurring variables.

2. The normal distribution plays a significant role in simplifying statistical analysis. It simplifies many aspects of statistical measures. Its aim is to provide a basis for easier calculation and interpretation of statistical measures. In addition, it enables us to facilitate the computation of probabilities, confidence intervals, and other statistical measures by the symmetric bell shape of it. Then, it aims to enable the application of statistical tests that assume normality and simplify understanding the distribution of data and aids in making statistical inferences about the population from the sample thanks to standard statistical tests like t-tests and Empirical Rule, respectively.

3. The normal distribution is important in the central limit theorem. The aim of it is to supply a foundation for inferential statistics. Furthermore, it is assistant for us to use sample data to estimate population parameters, establish confidence intervals, and do hypothesis tests even though the distribution is not normal. That is, it helps us to justify the use of normal approximation of sample data in the analysis and generalize the results of sample to population.

4. The normal distribution is significant in standardization and comparisons. The normal distribution enables us to standardize different datasets, which facilitates comparisons. In fact, the aim of it is to enable standardization of datasets from different populations for comparative analysis. It can be done by transforming data to the standard normal distribution, whose mean is 0 and standard deviation is 1, by using z-scores. Then, it aims to facilitate the calculation of probabilities and critical values by using the standard normal distribution tables.

5. The normal distribution is effective in decision making under uncertainty. In fact, the aim is to help us to make decisions in the presence of uncertainty. It is in fields such as engineering. Then, it aims to predict the future outcomes that are based on historical data which facilitates better planning in some fields like healthcare. Furthermore, it allows us to assess risk and optimize portfolio in finance. Then, it is helpful to develop quality and process control by understanding variations and deviations from the norm.

6. The normal distribution enable us to facilitate data science. The aim is to optimize data analysis processes. Then, it is assistant and considerable to improve and advance data preprocessing steps such as transformation, which ensure that data satisfies the assumptions of some analytical models.

7. The normal distribution plays an important role in analyzing sports and some demografic trends. In sports, it is very effective for us to evaluate and compare the athletes' performances by using the performance metrics of athletes like running times. In demografic trends, it is very assistant for us to understand situations by using the characteristics of populations like the distribution of income.

Conclusion

The normal distribution is the fundamental concept in science, with wide and prevalent applications in some fields. The properties such as the symmetry and the empirical rule make it a powerful tool for data analysis, inference, and decision-making. Understanding and comprehending the normal distribution is necessary and beneficial for interpreting data and implementing statistical methods accurately and properly. The aim of the normal distribution is to provide us with strong and versatile framework for understanding, modeling and analyzing data. The properties and theoretical foundations facilitate statistical procedures and decision-making in some fields. By approximating the natural phenomena and enabling standardization, the normal distribution remains the cornerstone of statistical theory and practice.

References

-Jaynes, Edwin J. (1996, December 5). Probability Theory: The Logic of Science. Retrieved from http://www-biba.inrialpes.fr/Jaynes/cc07s.pdf

-Rosenfeld, B. (n.d.). History of Statistics 2. Origin of the Normal Curve – Abraham DeMoivre (1667-1754). Retrieved from https://higherlogicdownload.s3.amazonaws.com/AMSTAT/1484431b-3202-461e-b7e6-ebce10ca8bcd/UploadedImages/Classroom_Activities/HS_2__Origin_of_the_Normal_Curve.pdf

-History of Normal Curve and Normal Distribution. (n.d.). Retrieved from https://www.scribd.com/document/171098296/History-of-Normal-Curve-and-Normal-Distribution

-Chen, J. (2024, March 13). Normal Distribution: What It Is, Uses, and Formula. Retrieved from https://www.investopedia.com/terms/n/normaldistribution.asp#:~:text=The%20Bottom%20Li

ne-
,Normal%20distribution%2C%20also%20known%20as%20the%20Gaussian%20distribution%2C%20is%20a,defined%20by%20the%20standard%20deviation

-Turney, S. (2024, January 29). What Is Kurtosis? | Definition, Examples & Formula. Retrieved from https://www.scribbr.com/statistics/kurtosis/#:~:text=Kurtosis%20is%20a%20measure%20of,(medium%20tails)%20are%20mesokurtic

-Z score table. (n.d.). Retrieved from https://www.z-table.com/

-Hayes, A. (2024, May 29). Empirical Rule: Definition, Formula, Example, How It's Used. Retrieved from https://www.investopedia.com/terms/e/empirical-rule.asp

-Frost, J. (n.d.). Z-Score Table. Retrieved from https://statisticsbyjim.com/hypothesis-testing/z-table/

-Measures of Position. (n.d.). Retrieved from https://online.stat.psu.edu/stat800/book/export/html/741#:~:text=The%20pth%20percentile%20of%20the,%25%20at%20or%20above%20it

-Bobbitt, Z. (2021, January 29). Normal Approximation to Binomial: Definition & Example. Retrieved from https://www.statology.org/normal-approximation/

-Ganti, A. (2024, April 22). Central Limit Theorem (CLT): Definition and Key Characteristics. Retrieved from https://www.investopedia.com/terms/c/central_limit_theorem.asp

-Normal(Gaussian) Distribution. (n.d.). Retrieved from https://www.probabilitycourse.com/chapter4/4_2_3_normal.php#:~:text=The%20CDF%20of%20the%20standard%20normal%20distribution%20is%20denoted%20by,is%20widely%20used%20in%20probability

-Hypothesis Testing. (n.d.). Retrieved from https://online.stat.psu.edu/stat200/book/export/html/441

-The Central Limit Theorem for Proportions. (n.d.). Retrieved from https://openstax.org/books/introductory-business-statistics-2e/pages/7-3-the-central-limit-theorem-for-proportions

-Skewness and Kurtosis. (n.d.). Retrieved from https://financetrain.com/skewness-and-kurtosis

-Normal Distribution and Standard Normal Distribution. (n.d.). Retrieved from https://www.almabetter.com/bytes/tutorials/applied-statistics/normal-distribution-and-standard-normal-distribution

-Shaikh, M. (2023, March 3). The Normal Distribution and Its Applications in Quality Control and Process Improvement. Retrieved from https://www.linkedin.com/pulse/normal-distribution-its-applications-quality-control-process-shaikh