

Overview of Probability and Statistics: Concepts, Methods, and Applications

Yinchu Liu

*New Channel, Qingdao, China
prc372516@163.com*

Abstract: Probability theory studies the likelihood of an event to occur. This review paper provides a summary of the concepts of probability and statistics. It discusses the evolution of probability theory from its origins to the transitions it has undergone to help scholars better understand and apply it. The paper further discusses theories by some scholars and their significance and relevance. It emphasizes the importance of statistics and probability as methods that can be applied in scientific research, public policy, finance, and daily decision-making to make predictions from data. Finally, the paper makes recommendations on the need for developing better statistical methods and algorithms in the wake of big data. It also predicts that machine learning and artificial intelligence innovations will benefit data analysis and influence how data is understood and processed.

Keywords: Probability theory, statistics, events, data

1. Introduction

In mathematics, probability theory quantifies the likelihood of various events. This branch of study provides a numerical description of the probability related to the occurrence of particular events, where probabilities range from 0 and 1. The nearer the probability to 1, the more the chances presented for an event to possibly occur. For instance, when a coin is flipped, the likelihood that it will be either a head or a tail is 0.5, indicating an equal likelihood for each outcome. Probability is fundamental in everyday reasoning and critical in insurance, science, and engineering, where various parties frequently make decisions facing uncertainty.

The human fascination with probability dates back thousands of years, with early evidence found in games involving dice and cards, which were prevalent in many ancient cultures. These games were among the first to introduce the general populace to the concept of chance and probability, albeit without a rigorous mathematical framework [1]. The formal mathematical treatment of probability did not emerge until much later, despite these early interests. It was not until the 16th and 17th centuries that significant strides were made towards a formal mathematical theory of probability, driven by the needs of gamblers and mathematicians like Gerolamo Cardano, Pierre de Fermat, and Blaise Pascal, among others. These pioneers set good starting points for probability theory, initially proposing games of chance, leading to a broader application in scientific inquiry.

Probability theory transitioning from heuristic interpretations of something happening by chance to becoming a rigorous mathematical framework was pivotal in its evolution [2]. Mathematicians such as Jacob Bernoulli, who coined the law of large numbers, greatly pioneered the transformation. Later, during the 20th century, Andrey Kolmogorov established the axiomatic foundations of

probability theory. Kolmogorov's formalism concept, based on set theory, defined the entire field of probability, allowing more complex phenomena to be modeled and analyzed systematically. This axiomatic approach clarified many earlier theories and paved the way for the integration of probability with other mathematical disciplines, which highly enhances the precision and scope of statistical analysis.

In recent years, probability theory has been indispensable in theoretical and applied contexts. It underpins the statistical methods used in hypothesis testing, predictive modeling, and decision theory. As data is the new oil in the world today, probability is now applied in the field of data science, where they are used to handle large volumes of complex data. Moreover, the study of probability has led to richer understandings in fields as diverse as quantum mechanics, genetics, economics, and even philosophy, underscoring its fundamental role in advancing human knowledge and technology.

In conclusion, probability theory is a key to unlocking insights in a world replete with uncertainties. The transition from ancient games to modern algorithms reflects its profound impact and utility. The transition also illustrates its central role in shaping how scientists think and how scholars make decisions in different disciplines. This review paper will explore the basic concepts of probability and statistics and how they are applied.

The paper outlines the basic concepts of probability theory, starting with its core definitions, terms, and axioms. We then introduce significant probability theorems such as Bayes' Theorem and the Law of Total Probability. These theorems provide a clear transition into discussions on continuous and discrete distributions. As the narrative of mathematics unfolds, the intricate relationship between the disciplines of statistics and probability theory becomes evident. Subsequently, the focus shifts to statistics, where we introduce essential concepts such as populations and samples. This sets the stage for a thorough exploration of fundamental statistical principles, including the Law of Large Numbers, the Central Limit Theorem, and the construction and application of confidence intervals. Following this detailed structure, the paper will discuss the deep interconnections and applications of probability and statistics in analyzing and interpreting data across various domains.

2. Probability theory

The modern framework of probability theory began to solidify in the 16th century, which marks a significant shift from its rudimentary use in gambling to a rigorous mathematical discipline. This transformation was fueled by the need to solve more complex problems arising from real-world scenarios beyond games of chance. As the theory developed, it incorporated both combinatorial methods, which were suitable for discrete events, and, later, analytical techniques to handle continuous variables [3]. This expansion was crucial for addressing a broader range of problems in scientific research and practical applications.

The foundational concepts of probability originate from the classical definition formulated by pioneers like Pierre-Simon Laplace and Jacob Bernoulli. In the classical interpretation, it is argued that probability is the proportion of the wanted outcomes in a condition where outcomes might not be biased. This approach, while straightforward, relies on the assumption of equal likelihood and does not require extensive empirical data. In contrast, the statistical definition of probability, which emerged to address limitations in the classical approach, discusses how frequently an event occurs over numerous trials [4]. This definition assumes that when there are more experiments then the experimental probability converges to the theoretical probability. Although there are many interpretations of probability theory, one can construct the foundation of probability theory through a strict and brief set of definitions and axioms.

The evolution of probability theory required a more formal structure to address complex analytical challenges. This need led to the development of a robust set of axioms by Andrey Kolmogorov in the 20th century, which provided a solid foundation for both theoretical and applied probability. These

axioms formalized probability in terms of set theory, allowing for a systematic approach to both discrete and continuous phenomena. The shift to a more axiomatic system in probability theory not only streamlined earlier methods but also enhanced the capacity to model and predict outcomes in increasingly diverse fields. This foundational shift has enabled probability theory to address not just games of chance but also to predict phenomena in areas ranging from genetics to economics, which reflects its critical role in modern science and decision-making.

Before detailed discussions of probability theory in Section 2, an overview is provided to outline the forthcoming content. Section 2.1 introduces the fundamental terminologies and axioms of probability theory, laying the necessary foundational groundwork for understanding the more complex concepts that follow. Section 2.2 progresses to explore Bayes' Theorem together with the Law of Total Probability, pivotal components for linking basic principles with practical application. Lastly, Section 2.3 discusses both discrete and continuous distributions, elaborating on their distinct characteristics and significance in various probabilistic models.

2.1. Terminologies and axioms in probability theory

In probability theory, a clear understanding of several fundamental terminologies is essential before discussing the subject further. We first define an experiment or trial as any procedure or action with a possible infinite repetition chance and is restricted to limited outcomes that might be possible—collectively, the space of the sample. This sample space uses the symbol Ω . It includes all possible experimental results— words, letters, numbers, or symbols, and can have a limited number of possibilities, a countably infinite number of possibilities, or an uncountably infinite number of possibilities. Each outcome within this set is referred to as a sample point, or an element.

To illustrate, consider tossing a coin: it is an experiment because it can be performed repeatedly and yields a specific set of potential results, heads or tails. Here, the sample space is $\Omega = (\text{heads}, \text{tails})$, forming the basis for defining probabilities and examining random events. Events—the sample space's subsets—if can range from simple to complex. A simple event involves a single outcome, whereas a complex event encompasses several outcomes. For example, when a dice is rolled, the subset $E_1 = (1,3,5)$ represents the event of that it can be an odd number, a complex event, just as $E_2 = (2,4,6)$ shows the possibility that it can be an even number. Further, events can be either non-mutually exclusive, which may overlap, or mutually exclusive, which cannot occur concurrently. Additionally, if events A and B satisfy $P(A \cap B) = P(A) \cdot P(B)$, they get labelled as independent events.

Probability refers to the proportion of events in the space of the sample. It reflects how possible it is for an event to happen. Probability is expressed in the range of 0 and 1. 0 indicates that the event won't happen, while 1 indicates that the event is guaranteed to happen [5]. Mathematically, probability is represented using a function known as the probability measure. The probability measure, P , shows the likelihood of each possible outcome in the space of the sample. The probability that an event A , will occur is represented by the function $P(A)$. For example if $\Omega = (a, b, c)$ and all outcomes have an even chance, then $P(a) = \frac{1}{3}, P(b) = \frac{1}{3}, P(c) = \frac{1}{3}$.

An uncertain event, associating each possible outcome with a real number can be represented as a random variable. Formally, a random variable X is a function $X: \Omega \rightarrow R$ represents the sample space of a random experiment. Via X , every result in Ω is equivalent to a numerical number. For instance, in a coin toss experiment, if $\Omega = (\text{Heads}, \text{Tails})$, X could be described as follows: $X(\text{Heads}) = 1$ and $X(\text{Tails}) = 0$

Mathematician Andrey Kolmogorov significantly advanced the formalization of probability theory into a rigorous mathematical framework in the early 20th century. In 1933, Kolmogorov published his seminal work, "Foundations of the Theory of Probability." In the book, the author introduced a

set of axioms that laid the groundwork for modern probability theory. This axiomatic approach provided a clear and consistent foundation for probability, addressing many existing logical inconsistencies in earlier formulations and allowing probability theory to be applied more broadly in various fields.

There are three fundamental principles in Kolmogorov's axioms. The first concept of the three axioms is Non-negativity. This axiom states that for any event A , the probability of A is non-negative: $P(A) \geq 0$. This is a fundamental requirement because probability shows how likely an event may occur, and a negative probability is impractical. This axiom ensures that probabilities align with our intuitive understanding of likelihood and avoids the illogical concept of negative probabilities.

Normalization is the second axiom. The axiom describes the probability of the entire sample space as $P(\Omega) = 1$. The sample space Ω indicates all possible outcomes, and its total probability must be 1. It represents the certainty that there will be an outcome. This axiom ensures that the total probability is properly normalized, preventing probabilities from exceeding or falling short of this logical boundary.

The third axiom is Additivity (or Countable Additivity). It states that for any countable collection of mutually exclusive events E_1, E_2, E_3 the probability of their union is equal to the sum of their probabilities: $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$. This principle ensures that probabilities are consistent in the case of a countable number of mutually exclusive events. It is crucial to ensure that the addition of probabilities is coherent and aligns with our intuitive understanding of how probabilities combine. Together, these three axioms form the foundation of modern probability theory, allowing for a rigorous and consistent approach to analyzing uncertainty. Kolmogorov's axiomatic system has enabled probability theory to be applied effectively across various disciplines, including statistics, finance, and physics. According to the above definitions and axioms, we can further draw some important preliminary conclusions, which are also the premises of subsequent theories.

Conditional probability and marginal probability are two important terms. These two measurements are very general and appear frequently in probability theory. The possibility of an event happening, given the occurrence of another event, is measured by conditional probability. Given event B , the conditional probability of event A is defined as follows: $P(A | B) = P(A \cap B) / P(B)$ for events A and B with $P(B) > 0$. This concept is fundamental to the development of Bayesian probability and statistical inference, and it is essential for comprehending dependent events.

Marginal probability is how likely an event may occur without considering the impact of other factors or events. They come from a joint probability distribution, which expresses how likely several things may happen at once. Assume that X and Y are two discrete random variables. The probability that X and Y will occur jointly is given by the joint probability distribution $P(X = x, Y = y)$. The marginal probability of X is computed by adding up all of the potential values of Y : $P(X = x) = \sum_y P(X = x, Y = y)$. Similarly, one can add up the probabilities of all potential values of X to find the marginal probability of Y :

$$P(Y = y) = \sum_x P(X = x, Y = y).$$

2.2. Law of total probability and Bayes' Theorem

In probability theory and statistics, the Law of Total Probability is a fundamental concept in probability and statistics. Scholars use it as a framework when working out the chance that an event will occur looking at various independent and comprehensive scenarios. The Law simplifies complex probabilistic problems, facilitating both theoretical analysis and practical applications. Consider

$(B_1, B_2, B_3, \dots, B_n)$ as part of the space of the sample S , where each B_i is an independent event with others, then the chance that event A will occur is represented like:

$$P(A) = \sum_{i=1}^n P(A | B_i) \cdot P(B_i)$$

Events $(B_1, B_2, B_3, \dots, B_n)$ are comprehensive and independent, indicating a division of the sample space. Given B , the probability of A under some conditions is shown as $(A | B_i)$. $P(B_i)$ is the probability of B_i .

The Law of Total Probability essentially simplifies calculating $P(A)$ by factoring in all possibilities (partitions B_i) that event A will occur. The breakdown leverages the idea that any event A can be seen within one of these partitions. For example, using the Law of Total Probability, calculating the chance that an employee can be promoted—within a company with different departments—is possible (e.g., Sales, Marketing, Engineering). Each department (partition) has its probability of promotion. By summing these probabilities weighted by the departmental distribution, you obtain the overall probability of promotion.

Bayes' Theorem is named from the 18th-century statistician Thomas Bayes and represents a cornerstone of Bayesian statistics. It offers a way to adjust a hypothesis's probability when new data or evidence becomes available. This theorem is frequently used in many scientific and practical domains and is essential for decision-making processes where uncertainty is present.

From the definition of conditional probability: for events A and B , the following formula is applied:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

Rearranging this equation gives:

$$P(A \cap B) = P(A | B) \cdot P(B).$$

Similarly, the conditional probability $P(B | A)$ uses the following formula:

$$P(B | A) = \frac{P(A \cap B)}{P(A)}.$$

Substituting $P(A \cap B)$ from the earlier equation into this expression yields:

$$P(B | A) = \frac{P(A | B) \cdot P(B)}{P(A)},$$

known as the Bayes' Theorem [6].

In probability theory and statistical inference, Bayes' Theorem is a fundamental tool. It can be used to systematically update probabilities based on new evidence, enabling predictions that are more accurate and decisions. Anyone involved in data analysis, decision-making, or probabilistic modeling should understand and be able to apply Bayes' Theorem is crucial for.

This theory has classical applications in information filtering. For example, Bayes' Theorem is applicable in filtering spam email based on the email's content. Let:

- S be the possibility that an email is spam.
- W be the possibility of the occurrence of a word in the email.

We are interested in $P(S | W)$, the likelihood than an email can be labelled spam considering that it contains a specific word. Using Bayes' Theorem:

$$P(S | W) = P(W | S) \cdot P(S)P(W)$$

$$P(W) = P(W | S) \cdot P(S) + P(W | \bar{S}) \cdot P(\bar{S})$$

where $P(W | S)$ is the likelihood that a word will appear in non-spam emails and $P(S)$ is the likelihood that an email is not spam.

Bayes' Theorem is also related to the Law of Total Probability. $P(A)$ is computed as follows

$$P(A) = \sum_{i=1}^n P(A | B_i) \cdot P(B_i)$$

Bayes' Theorem utilizes this result in the denominator to adjust probabilities based on the latest available information. This interplay highlights the theorem's role in probabilistic reasoning and statistical inference.

2.3. Discrete and continuous distribution

Expectation and Variance are fundamental concepts that describe the characteristics of random variables. They are key to explaining the central tendency and dispersion of random variables, which helps us understand and predict the pattern of random events.

Expectation, often denoted as $E[X]$, shows the mean of random variables. It reflects the mean of the random variable in the long-term over many trials. In practical applications, the expectation helps in making decisions and predictions. For instance, in gambling, the expectation can help assess the fairness of the game. The expectation is computed by looking at the random variable's probability distribution, which involves considering the weighted average of possible values. The expectation in discrete random variables is the all possible values' weighted average, while it is computed through the integration of the probability density function in continuous random variables.

Variance, denoted as $Var(X)$, calculates the degree of how spread or how dispersed a random variable's values in the proximity of its mean. If the variance is higher, it means there is greater variability and uncertainty in the random variable's values. The standard deviation provides a more intuitive sense of the spread of values around the mean. Variance quantifies the spread of the random variable, reflecting the level of risk and uncertainty. Practically, scholars use the variance and standard deviation to evaluate risk and random variables' variability.

Scholars use distributions to understand patterns of random variables. They describe the spread of random variables over the possible values. Distributions can be discrete or continuous. The two types of distribution have unique characteristics and applications, which are crucial for modeling different types of random phenomena.

Discrete distribution describes the likelihood of outcomes for discrete random variables—those that possibly have a finite number of distinct values. These distributions are used when dealing with scenarios where outcomes are clear and separate. For example, the number of times that series of trials succeed or the number of items found in a sample. For instance, the amount times that it will be a head when a coin is flipped 10 times follows a binomial distribution.

A discrete probability distribution matches each possible value of a discrete random variable to its suitable probability. Using the probability mass function (PMF) each value of the discrete random variable can have a probability related to it. If X identifies as a discrete random variable, from the PMF $p(x)$ probability X is exactly equal to x . Formally, $p(x) = P(X = x)$, and x is a specific value in the group of possible outcomes of X . The PMF need to meet two conditions:

1. The likelihood of every possibility: $p(x) \geq 0$, due to the first axiom of probability.
2. The sum of likelihoods over possibilities must equal to 1: $\sum_x p(x) = 1$, according to the second axiom of probability.

The discrete random variable X , has the expectation $E[X]$ defined as the all possible values' weighted average. Mathematically, the formula for a discrete random variable's expectation is:

$$E[X] = \sum_{i=1}^N x_i \cdot P(X = x_i),$$

where x_i stands for all the possibilities random variable X can possess, N is the number of elements in the space of the sample, and $P(X = x_i)$ is the probability of X taking the value x_i .

Variance $Var(X)$ measures how much the values of X deviate from its expectation. For a discrete random variable, variance is calculated by:

$$Var(X) = E[(X - E[X])^2] = \sum_{i=1}^N (x_i - E[X])^2 \cdot P(X = x_i).$$

Alternatively, variance can be simplified using:

$$Var(X) = E[X^2] - (E[X])^2$$

where $E[X^2]$ is the expectation of the square of X :

$$E[X^2] = \sum_{i=1}^N x_i^2 \cdot P(X = x_i).$$

The Poisson distribution models the amount of times an event occurs in a set period or spatial area. Events occur independently and consistently. The Poisson distribution is a classic example of a discrete distribution, as it models the countable outcomes of events, making it an essential tool in probability and statistics. Discrete probability distributions define the probabilities associated with a random variable that corresponds to many individual outcomes [7]. Since the Poisson distribution counts the events that are inherently discrete and countable, it falls under the category of discrete distributions.

The Poisson distribution counts the times that an event occurs in a set interval of time or space, given these events are independent and occur consistently. The distribution shown by parameter λ , representing the average rate or intensity of events.

Let X be a random variable representing the amount of times an event occurs in a set interval. If X is in line with Poisson distribution, its probability mass function (PMF), is represented by:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

λ is the average rate of that an event will occur (the mean number of events). x represents number of events, which must be a non-negative integer $(0, 1, 2, \dots)$. e is the natural logarithm's (2.71828 by estimate).

In the Poisson distribution, both the mean (expected value) and the variance equate to the λ . That is: $E[X] = \lambda$ $Var(X) = \lambda$. In the Poisson distribution, the probability that future events will occur is independent of past events. In other words, past occurrences do not influence the probability of future occurrences. If two independent Poisson processes are observed in the same period, the total number of events follows another Poisson distribution with an average rate equal to the sum of the two individual rates.

Continuous random variable X has the expectation $E[X]$ of all possible values' weighted average given by the probability density function (PDF). Mathematically, the continuous random variable's expectation is shown by:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

where $f_X(x)$ shows X 's probability density function of.

Continuous random variable X has the variance $Var(X)$ is similarly defined as:

$$Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2$$

Variance is also represented by:

$$Var(X) = E[X^2] - (E[X])^2$$

where $E[X^2]$ is the expectation of the square of X :

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) dx$$

The methods for calculating expectation and variance differ between discrete and continuous random variables, but the underlying concepts are consistent. The expectation describes the weighted average of possible values, while variance measures the dispersion of these values. For discrete random variables, expectation and variance are computed using summation, whereas for continuous random variables, integration is employed.

The normal distribution, also called the Gaussian distribution, is significant and popular in continuous probability distributions in statistics. Its bell-shaped curve and symmetry make it prevalent in both natural phenomena and social science data. This article will explore the definition, key features, and practical applications of the normal distribution.

The normal distribution is defined by two parameters: the mean (μ) and the standard deviation (σ). The mean indicates the central location of the distribution, while the standard deviation describes the spread or dispersion of the data. The probability density function (PDF) of a normal distribution is given by the formula:

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

In this formula, $\exp$ denotes the exponential function, $\sqrt{2\pi\sigma^2}$ is the normalization factor that ensures the total area under the curve is 1, $(x - \mu)^2$ represents the squared distance from the mean. The graph of the PDF of the normal distribution is a smooth, bell-shaped curve that is symmetric about the mean μ , making it highly useful in various statistical analyses.

The normal distribution is a fundamental concept. It is nobly recognizable by its bell-shaped curve, which is formed by the probability density function. This curve is smooth and peaks at the μ , signifying that most data points are crowded close to the mean. The normal distribution is also notable for its symmetry: the curve is identical on either side of the mean, reflecting that data is evenly distributed around this central value. This symmetry implies a balanced distribution and follows a consistent pattern on the two sides of the mean.

The empirical rule, or the three-sigma rule, is a normal distribution's important characteristic. It indicates that close to 68.3% of the data points are found in a mean's standard deviation, close to 95.4% are found in the proximity of two standard deviations, and nearly 99.7% are in the proximity of three standard deviations. From the rule, understand a clear understanding of how data is spread and how values are distributed in comparison to the mean.

The mathematical properties of the normal distribution further define its shape and behavior. The center of the distribution is determined by μ , on the other hand the width is dictated by σ . A wider curve is from a larger standard deviation that has data more spread out in relation to the mean, whereas a narrower curve is from a smaller standard deviation that has data more closely clustered around the mean. These features collectively describe the distribution's form and are essential for analyzing and interpreting statistical data.

The normal distribution has broad applications across various fields. In biology, many natural traits such as human height and weight usually take after a normal distribution. In a case where 170 cm is the mean height of a population, and the standard deviation is 10 cm, the height of most individuals is close to the mean.

3. Statistics

In mathematics, statistics focuses on collecting, analyzing, interpreting, presenting, and organizing data [8]. It provides tools and methodologies for understanding and making inferences about complex data sets that are used to make decisions in economics, business, science, and social sciences. Statistics involves two main areas: descriptive statistics, which summarizes and describes the data set's features through measures such as mean, median, and standard deviation, and inferential statistics, which uses samples in probability theory to predict or generalize a population [9]. This discipline is fundamental for designing experiments, conducting surveys, and making data-driven decisions, highlighting data trends, patterns, and relationships. Individuals or organizations use statistical methods to make informed decisions, test hypotheses, and systematically solve practical problems.

3.1. Terminologies

In statistical research, the concepts of population, sample, and the sampling process are fundamental [10]. The population is defined as study objects in the form of a whole set of individuals or observations that are study objects. The population can fall in the category of finite or infinite. The population could consist of all city residents or all employees in a company. The subset of individuals or observations taken from the population is called a sample. The sample should represent the population to make accurate inferences about its characteristics. Samples are chosen by either stratified, simple random, or systematic methods. The sampling method can be applied based on their strengths and weaknesses. The following steps need to be followed when sampling: defining the population, having the collection or list of the population), choosing the method to be used for sampling, and determining how big the sample will be. Implementing the sampling involves selecting the sample according to the chosen method, followed by data collection and analysis. Effective sampling helps researchers use sample data and inference reliably from the population, which is a scientific basis for decision-making. However, ensuring the representativeness and randomness of the sample is crucial to getting accurate results; a small sample may lead to unreliable conclusions, while a large sample may waste resources.

The mean and sample variance (also standard deviation) are crucial statistics for describing central tendency and how data is dispersed. The mean, also called the average, is calculated by the formula:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i,$$

Where the mean is represented by $\bar{X}$, the number of observations are represented by n and X_i denotes an individual observation. The mean indicates the data's central location, but outliers that do not fully represent the overall distribution can affect it.

Sample variance is used to measure the dispersion of sampled data and it measures the average squared deviation of every data point's position in relation to that of the mean. The formula for sample variance is:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2,$$

Where the variance is represented by s^2 is the variance, $\bar{x}$ is the mean, x_i represents each data point, and the number of observations is represented by n . Data points widely distributed around the mean stands for larger variance indicates, reflecting higher dispersion. However, since the unit of variance is the square of the original data unit, it is not always easy to interpret directly. Therefore, the standard deviation-simplifies the interpretation of the dispersion measure by bringing the unit back to the data's initial scale. The standard deviation has the following formula:

$$s = \sqrt{s^2},$$

Where the standard deviation is represented by s and the variance is represented by s^2 . Data points close to the mean are represented by a smaller standard deviation, while data points spread from the mean are represented by a larger standard deviation. Collectively, data distributions are analyzed and understood using the average of the sample, variance and standard deviation.

3.2. Law of large numbers and central limit theorem

The Law of Large Numbers (LLN) is fundamental. It shows what scholars get from repeating an experiment. According to the law, the expected value is close to the average from several attempts and ends up becoming closer with more trials. The law is in two main forms: the Strong Law of Large Numbers (SLLN) and the Weak Law of Large Numbers (WLLN). According to the SLLN, the sample's average almost surely approaches the expected value, which of the ways it converges is stronger, while according to the WLLN, the averages of the sample have a probability toward the expected value as more samples are included[11].

From SLLN, it can be observed that the probability that the sample's average gets close to the population's average as more samples are included that it becomes infinitely large and then equal to 1. This can be formulated as:

$$P(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$$

where:

$\bar{X}$ denotes n i.i.d. random variables' sample mean. μ stands for the expected value (mean) of each of these random variables.

The key difference between these two laws is the mode of convergence. The SLLN ensures certainly converges, which implies that the sequence of sample means will almost certainly get close to the mean of the population mean as more samples are included indefinitely. The WLLN states convergence in probability, meaning that while the mean of the sample may not approach the population mean on every sample sequence, the sample mean is more likely to be near the population mean as more samples are included.

The Law of Large Numbers is widely applied in theoretical statistics and other practical fields. It supports the idea that increasing sample size used in experiments will result in the mean of the sample accurately estimating the overall mean of the population. This theorem justifies approximating expected values empirically in situations where the actual distribution of an outcome is unknown, which is commonly the case in fields such as economics, insurance, and physics.

For instance, in the financial sector, specifically in risk management and portfolio optimization, the LLN plays a crucial role. Investment decisions in financial markets often rely on predicting future

returns and risks based on historical data analysis. LLN provides a theoretical foundation that ensures as the sample size increases, the statistical estimates (such as average returns) more accurately reflect the overall characteristics, thus aiding financial analysts in making more informed decisions. In portfolio management, the risks and expected returns of various assets are assessed by investors. Investors use historical data to work out the mean of the sample and variance of the sample of returns by analyzing historical data. LLN guarantees that growth of the size of the sample results in the mean of the sample converging to the mean of the population, and the variance of the sample will converge to the variance of the population [12]. This allows investors to more accurately estimate the long-term expected returns and risks of assets, thereby optimizing their portfolios to maximize returns and minimize risk.

The Central Limit Theorem (CLT) sheds light on the normal appearance of many distributions under certain conditions. From the theorem, it can be understood that, when the sample size is large enough, the independent, identically distributed variables that have a finite mean and variance will have a distribution of the sample mean that is close a normal distribution, regardless of the underlying distribution shape[13]. When the size of the sample is increased, the approximation also improves. In statistics, it is an important theorem because it allows for the simplification of problems and the normal distribution can be used to approximate results of various random processes.

Mathematically, the following is a representation of the CLT: Suppose $X, X, \dots, X$ are n independent random variables that are identically distributed with expected value μ and have the finite variance σ^2 . According to The Central Limit Theorem, as n approaches infinity, the normalized sum:

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}},$$

approaches a standard normal distribution $N(0,1)$, where $\bar{X}_n$ stands for n variables of the sample mean [14]. This implies that $\bar{X}_n$ itself approaches the normal distribution $N(\mu, \sigma^2/n)$.

A practical application of the CLT is found in polling and survey analysis. When a pollster collects data from a sample of people, the CLT enables them to make inferences about the entire population. For instance, if different samples are taken from a population to estimate the average height, as more samples are included, the sample mean height's distribution will closely resemble a normal distribution. Pollsters use this attribute to apply techniques that assume normality, such as confidence intervals, even if the distribution of individual heights is not normal. Another significant application of the CLT is in the field of quality control. In manufacturing processes, quality assurance often involves measuring attributes such as thickness, strength, or resistance across a range of products. By the CLT, regardless of the non- normal distribution of these attributes, the distribution of the average of these measurements, taken over many products, will be approximately normal. This enables engineers to use properties of the normal distribution to set acceptable ranges and tolerance limits, helping to maintain consistent product quality and predict manufacturing defects.

3.3. Confidence intervals

In statistics, Confidence Interval (CI) is estimates the range of values likely containing a true population parameter, expressed with a certain degree of confidence. In most cases, it is expressed in terms of a percentage (for example, 90%, 95%, or 99%), this interval estimates the likely range of the unknown parameter. The confidence level indicates the frequency of intervals. A certain percentage of the intervals calculated using different samples taken from the same population, will include the unknown parameter being calculated for. However, it is not a guarantee that any specific interval will contain the exact parameter.

Confidence interval selection involves selecting a sample from the population. The next steps involve computing the mean sample's standard error. Next, the confidence interval is calculated by:

$$CI = \bar{X} \pm z^* \left(\frac{s}{\sqrt{n}} \right)$$

$\bar{x}$ represents the the sample's mean in the formula, while s represents standard deviation of the sample, n is the size of the sample, and z^* stands for the corresponding critical value to get the wanted confidence level (for example, 1.96 for 95%).

The size of the sample, the population's standard deviation, and the confidence level determine how wide the confidence interval will be. Narrower intervals indicate estimates that are more precise but require larger sample sizes or reduced confidence levels to maintain accuracy. This balance between precision and confidence is a fundamental consideration in statistical analysis and reflects the inherent trade-offs in empirical research.

4. Conclusion

In conclusion, the study of statistics and probability provides a fundamental framework for breaking down the complexities of data across a myriad of fields. By exploring probability theories and statistical inferences, we gain essential tools to guide decision-making and predictions from empirical data. Applying statistics and probability concepts extends from scientific research and public policy to finance and everyday decision-making, further demonstrating their importance in both theoretical and practical contexts.

As we look to the future, statistics and probability will evolve greatly. The increasing availability of big data offers potential and drawbacks. Therefore, sophisticated statistical methods and algorithms need to be developed. Machine learning and artificial intelligence innovations are expected to enhance further our ability to analyze complex datasets and derive actionable insights. Additionally, the integration of statistical techniques with emerging technologies such as quantum computing may open new frontiers in data analysis. As these advancements unfold, they will undoubtedly drive new research directions and applications, shaping the future of how we understand and leverage data.

References

- [1] Bertsekas, D., & Tsitsiklis, J. N. (2008). *Introduction to probability (Vol. 1)*. Athena Scientific.
- [2] Breiman, L. (1992). *Probability*. Society for Industrial and Applied Mathematics.
- [3] Ross, S. M. (2014). *Introduction to probability models*. Academic press.
- [4] Jeffreys, H. (1998). *The theory of probability*. OuP Oxford.
- [5] Koopman, B. O. (1940). *The bases of probability*.
- [6] Berrar, D. (2019). *Bayes' theorem and naive Bayes classifier*.
- [7] Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R., & Mykytka, E. F. (1979). *A probability distribution and its uses in fitting data*. *Technometrics*, 21(2), 201-214.
- [8] Shao, J. (2003). *Mathematical statistics*. Springer Science & Business Media.
- [9] Schervish, M. J. (2012). *Theory of statistics*. Springer Science & Business Media.
- [10] Ostle, B. (1963). *Statistics in research*.
- [11] Révész, P. (2014). *The laws of large numbers (Vol. 4)*. Academic Press.
- [12] Baum, L. E., & Katz, M. (1965). *Convergence rates in the law of large numbers*. *Transactions of the American Mathematical Society*, 120(1), 108-123.
- [13] Rosenblatt, M. (1956). *A central limit theorem and a strong mixing condition*. *Proceedings of the national Academy of Sciences*, 42(1), 43-47.
- [14] Kwak, S. G., & Kim, J. H. (2017). *Central limit theorem: the cornerstone of modern statistics*. *Korean journal of anesthesiology*, 70(2), 144-156.