

Statistical species distribution and their respective development

Yinjie Liu

Department of mathematics, Oregon State University, Corvallis, United States

liuyinj@oregonstate.edu

Abstract. A distribution is a set of information on a variable. When these data are normally grouped in size order from least to largest, they may then be visually represented. In fact, it is fairly simple to comprehend statistical distributions in terms of functional relationships. Simply put, a data variable is thought of as being coupled with another data variable or several data variables into specific functional relationships, the majority of which can be reflected in the coordinate axes. Once the distribution function has been built, it may be swiftly used to define and compute significant variables, such as an observation's probability, as well as to show the relationships between observations in the domain. The distribution of statistics is very broad, it goes deep into various fields of study, with different specialties and models combined. The process of discovering them is also different, and again, different distributions apply to different scenarios. When faced with different mathematical models, Mathematicians should choose the most suitable distribution method to calculate the probability density function, cumulative distribution function, and calculate the probability and expected value, so as to correctly understand the model. The following article will focus on different definitions of statistical distributions and their origins.

Keywords: probability space, random variables.

1. Introduction

A mathematical function that explains the correlation between observations of various heights is known as a statistical distribution [1-17]. Simply said, a distribution is a group of data about a variable. These data can then be visually depicted when they are typically grouped in size order from smallest to largest. In fact, it is relatively easy to understand statistical distributions in terms of functional relationships, which can be simply thought of as a data variable with another data variable or several into certain functional relationships, most of which can be reflected in the coordinate axes. Once the distribution function is established, it may be quickly utilized to describe and compute important values, including the probability of an observation, as well as to depict the connection between observations in the domain. Different statistical distributions generally respond to different scenario models, for example, the most familiar and widely used is the normal distribution, which accurately describes the distribution of values of many natural phenomena. For example, the height of a population is normally distributed because most people in a given population are of average height, they are close to each other, those who are taller and shorter than the average are almost equally tall, and very few are either extremely tall or extremely

short. If the height of the surveyed population is plotted from low to high, the graph will eventually be found to show a bell-shaped curve. These distributions are ultimately what help researchers to analyze data structures such as probability distributions, so let's start by understanding what statistics are.

2. Probabilistic programming

Probabilistic programming is a technique for developing system that assist humans in making decisions in the face of uncertainty. Many daily choices involve using judgment to identify pertinent, non-observable aspects. Probabilistic programming identifies those unseen aspects that are crucial to a decision by fusing our understanding of a situation with the rules of probability. Probabilistic reasoning system is becoming simpler to design and more often utilized thanks to a unique technique called probabilistic programming.

Interested in knowing, for example, if it will sell well. You may think it will succeed because you think it is well designed and because your market research shows that there is a market for it, but you can't be certain [1-17]. Mathematicians can make these selections with the use of probability terminology. You can gauge the probability of a product's success when you introduce it by using prior experience with related products. The decision to introduce the product can then be made using that probability. You can be concerned not only with whether the product will succeed, but also with how much money it will make or how much money it will cost you if it fails. You can utilize the probability of various events to decide more wisely.

FACT: Knowledge and logic are used to make decisions.

The main goal of probabilistic programming is to offer techniques to express the knowledge and reasoning needed to respond to inquiries.

2.1. Critical definitions

What you know to be true about your area generally, without taking into consideration the particulars of a particular circumstance, is known as general knowledge.

A probabilistic model is a way to encode generic information about a subject in numerical terms.

The exact facts you have to back up a claim is called evidence.

Query: Describe a feature of the situation you're interested in.

The process of inference involves drawing conclusions from evidence using a probabilistic model.

In probabilistic reasoning, you develop a model that quantifies and probabilistically expresses all the pertinent general knowledge in your domain.

The rules of probability offer a clear mathematical definition of the link between the model, the data you supply, and the responses to questions. Probabilistic inference, or simply inference, is the method of applying the model to respond to inquiries based on the data. Fortunately, mathematical computations have been automated by computer programs, which do the arithmetic for you. Inference algorithms are the name given to these methods.

Three methods of reasoning are available to probabilistic systems:

1. Predict future occurrences, which refers to thinking forward in time and making predictions about the future based on what you know about the present.

2. Determine the origin of events by extrapolating past circumstances from the results of the present.

3. Use historical data to improve your ability to anticipate the future.

The three modes of reasoning provide ways of reasoning about specific situations given the evidence. A probabilistic reasoning system may also be used to learn from the past and develop common sense. In the third method of thinking, you may understand how you might draw lessons from certain previous events to better anticipate particular future circumstances. Enhancing the model itself is another approach to gain knowledge from the past. The purpose of the learning algorithm, which differs somewhat from an inference algorithm, is to create a new model rather than provide a response to queries. The old model is the starting point for the learning algorithm, which builds a new model by updating the old one in response to experience. The new model will eventually be able to respond to query. The

results produced by the new model should provide more accurate information than those produced by the previous model.

2.1.1. Systems of probabilistic inference and reliable predictions. The more data there are, the more accurate the probabilistic inference system will be, much like any machine learning system. The initial model's accuracy in reflecting reality and the quantity of data made available both affect how well the forecast turns out. In general, the initial model loses significance when more data are presented. This is so because the new model strikes a balance between the information in the data and the old model. The original model will predominate if there is very little data, thus it better be correct. When there is an abundance of data, the data will be considered more important, and the new model will frequently disregard the previous model.

To explain its probabilistic models, every system of probabilistic inference employs a representation language. These programs are referred to as probabilistic reasoning systems or probabilistic programming systems, and they are created in a programming language like Bayesian networks.

In contrast to mathematical constructs like a Bayesian network, models are written as programs in programming languages, which is the main distinction between probabilistic reasoning systems and probabilistic programming systems. This modification affects how evidence, questions, and responses relate to program variables. Program variable values may be specifically stated in evidence, whereas program variable values are requested in questions, and answers are probabilities of various query variable values. A probabilistic programming system also frequently includes a collection of inference methods. The language's programs can use these algorithms.

2.1.2. Key definitions. Representation language is a language that you may use to model your knowledge of a domain. Expressive power is a representation language's capacity to represent a variety of types of knowledge.

A language that can express every computation a machine is said to be Turing-complete.

Utilizing a Turing-complete programming language to express knowledge in a probabilistic representation language is known as probabilistic programming.

2.1.3. Probabilistic models represented as programs. Execution is a fundamental concept in programming languages. To produce output, you run a program. A probabilistic program is similar, but rather of having only one execution route, it may have several, each of which produces a distinct outcome. Random selections made throughout the program specify which execution route will be taken. The software encodes the probability of each conceivable outcome for each option made at random. Consequently, a probabilistic program maybe conceptualized as a program that run at random to produce a result.

Why probabilistic programming?

To anticipate the future, infer the past, and improve future predictions, one can utilize probabilistic reasoning. And probabilistic programming entails utilizing a Turing-complete programming language to describe probabilistic reasoning.

FACT Probabilistic reasoning + Turing-complete = probabilistic programming

3. Statistics

What are statistics? Statistics is the analysis of data. Traditionally, statistics is the formulation of hypotheses to questions and then analysis to answer judgmental hypotheses. What is a stochastic process? A stochastic or random process is a mathematical construct that is typically specified as a random variable in probability theory and other related theories.

Random variable is kind of like sometimes when you are doing survey or doing some experiment, you are not concerned in the specifics of the experiment's results. Since real numbers are a function of the sample space, random variables are defined as such. However, there are situations when random

variables can only have a distinct range of values (finite or countable). A "new" sample space is created when Mathematicians define a random variable, and it is based on the random variable's possible values.

As a result, Mathematicians should be able to use the probability measure defined in the original sample space, to compute the probabilities associated to the different values of the random variable. Now that Mathematicians mention random variable, Mathematicians have to mention the distribution. The probability distribution of a random variable describes the distribution of probabilities over the values of the random variable.

4. Statistical distributions

4.1. The Bernoulli distribution

The Bernoulli distribution is very closely connected with like the logic operators, closely connected with logical operators. So the Bernoulli distribution can only have two values, whether it is 0 or whether it is 1. For probability of P , let $X=0$ and $X=1$, the value can take on one, but for the probability function, the sum of the probability is equal to one. So the probability of $X=0$ is equal to one minus P , another probability of $X=1$ is P . Mathematicians can define one event if does not occur, then the probability is 0. If the event occurs, the probability is 1, this is Bernoulli distribution.

4.2. the binomial distribution

A Binomial experiment is defined by the following conditions:

- There are n identical trials total in the experiment.
- Either success (S) or failure (F) is the outcome of each experiment.
- A constant number p , which represents the probability of win on a single trial, determines the probability of success. The probability of unsuccess is also $q = (1 - p)$.
- Each trial is independent.
- The total number of successes noted in the n trials is defined as X , the random variable of interests.

Observe that a Binomial experiment consists of a sequence of n independent repetitions of a Bernoulli distribution trial with success probability p .

Next, the total numbers of success are a binomial random variable with two parameters n and P , where character X can have values of $0, 1, \dots, n$ and

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Each one configuration with k successes has the same probability of success $p^k (1 - p)^{n-k}$ and there is $\binom{n}{k}$ configurations.

Then Mathematicians will talk about the continuous random variable

A random variable that can have continuous values is shown to be a continuous random variable.

Mapping $X: \Omega \rightarrow R, \omega \rightarrow X(\omega)$, where X is a random variable.

- Cumulative distribution function (abbreviation CDF) : $F(x) = P(X \leq x)$
- As a continuous random variable, X is said to, if f exists, which satisfies $P(x \in A) = \int_A f(x)dx$ for the density function, denoted by f , is known for any "reasonable" set A .

In statistics, the binomial distribution is a typical distribution function for discrete processes in which each value that is independently created has a set probability.

Originally studied in pure probability games, in practically all areas of human research, data analysis is now often done using the binomial distribution. It applies to any fixed number (n) of iterations of an independent process, each with the same probability (p) of producing a particular outcome. For instance, it gives a formula to determine the probability of getting 10 sixes out of 50 dice rolls. The Swiss mathematician Jakob Bernoulli proved that the probability of k such occurrences from n repetitions is equal to the k th term of the expansion of the binomial expression in a paper that was published after his death in 1713. $(p + q)^n$ (k starts at 0), where $q = 1 - p$ (because of the name

binomial distribution). In the dice rolling example, the probability that any certain number will appear on each roll is $\frac{1}{6}$ (the number of faces on the dice). Then the probability of 10 sixes occurring in 50 rolls is equal to $(5/6 + 1/6)^{50}$, the tenth term in a 50-roll (beginning with the 0th term), or 0.115586.

4.3. The normal distribution

The most significant distribution in statistics is the normal distribution.

Definition: A random variable Y is shown that, Y have the normal distribution with two parameters μ and σ^2 , if and only if it has a probability density function that has the following the form

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad -\infty < y < \infty, \text{ where } -\infty < \mu < \infty, \text{ and } \sigma^2 > 0$$

From the definition, Mathematicians observe that the normal distribution is symmetric around μ and that σ^2 determines the spread of distributions.

The French mathematician Abraham de Moivre, who had a scientific interest in chance games and frequently assisted gamblers in calculating probabilities, is credited with developing the normal distribution. De Moivre is credited with working on the coin toss probability distribution. He sought to come up with a formula to describe, for example, the probability of getting 60 or more crosses in a set of 100 coin flips. He came at what is known as a "normal curve," a bell-shaped distribution, to respond to this query. This is an important finding since many phenomena exhibit a roughly normal distribution. For instance, a normal distribution describes variables (phenomena) like height, weight, and strength. Because of this, a Z-score table may be used to compare a standing person's weight or height against others. The first person to observe the relationship between the distribution of weight and height and the normal curve was a Belgian astronomer named Lambert Quetelet. Initially, measurement errors in astronomical data were examined using normal curves. These mistakes happened as a result of the observers' inherent bias and the inaccurate equipment.

Galileo first noticed that these mistakes were symmetrical. Additionally, he noticed that minor mistakes had a larger frequency. Inferences concerning the distribution of mistakes were made as a result of this. But it wasn't until the 19th century that it was realized how frequently these blunders occurred. Adrian and Gauss, two mathematicians, independently created a formula for the normal distribution. The calculation revealed that a normal curve provided a good approximation of the inaccuracies.

It is important to remember that Laplace, who developed the central limit theorem, also identified the same distribution near the end of the 18th century. According to this theorem, even when the sample mean is drawn from a non-normal distribution, its distribution still conforms to the normal distribution. The distribution approaches the normal distribution more closely as sample size increases. Numerous disciplines, ranging from the social sciences to medicine, use normal distributions and the accompanying Z-score computation. Z-scores are a helpful standardized number for comparing populations based on traits like height, weight, test scores, wealth, and a variety of other factors.

4.4. The uniform distribution

Definition: A random variable Y is shown to have a uniform distribution on the interval (θ_1, θ_2) if and only if it has the probability density function of the form $f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 < y < \theta_2 \\ 0, & \text{otherwise} \end{cases}$

Mathematicians write $Y \sim \text{Uniform}(\theta_1, \theta_2)$

The standard uniform distribution has the shape of a square with a height of one since it is the specific case where a is 0 and b is 1. The curve's area under any interval in a probability distribution denotes the probability that the distribution denoted by a number happens in that interval. The probability of seeing a value between (0, 0.5) and the probability of seeing a value between (0.5, 1) for the standard uniform distribution are both 50%. All values in the finite set are equally likely to occur in the discrete version of the uniform distribution. If the set of potential values is 0 and 1, for instance, there is 50% probability

of obtaining 0 and a 50% probability of getting 1, like how a fair coin has a 50% chance of coming up heads and a 50% chance of coming up tails.

Because uniform probabilities imply unpredictability whereas the society is defined by regularity, or a sort of randomness, uniform distributions are uncommon. The uniform distribution is still useful to social scientists, though. In fact, after the normal distribution, it is the distribution that is employed the most frequently. Finding random variables from other probability distributions, such the normal distribution, is the uniform distribution's most well-known use. A number picked at random from a group of numbers with the distribution's probability is referred to as a random variable. Any distribution's random variables are created by feeding a uniform normal variable through the distribution's inverse cumulative probability function.

4.5. Expected values

For a discrete random variable such as X , the expectation or expected value $E(X)$ is defined as

$$E(X) = \sum_x x \cdot p(x) \text{ where } x \text{ 's possible values are all added together to get the total.}$$

When the expected value exists (i.e., the sum is well defined), the population mean is another name for the expected value, as the expectation is just a kind of weighted average, where more probable values of x are given more weight $p(x)$.

Note that the expected value $E(X)$ of a discrete random variable can be equal to a value that is not a possible result.

The expected value represents what Mathematicians anticipate seeing on average over a long series of observations, not what Mathematicians will see in a single observation.

For example, Expected value of Binomial distribution $X(n, p)$

$$E(X) = np, \text{ i.e. which is the mean value of the random variable } np.$$

Expectations of functions of the random variable

Theorem: Let X be a discrete random variable, and with the probability mass function $p(x)$, also let $g(X)$ be a real valued function of X . Next, the expected value of $g(X)$ is given by $E[g(X)] = \sum_x g(x) \cdot p(x)$. Observe that if X is a random variable, Mathematicians will say $Y = g(X)$ is also a kind of random variable.

Therefore, this theorem allows us to obtain the expected value of the random variable Y without having to obtain the probability mass function of Y .

Linearity of the $E(x)$ expectation

X_1, \dots, X_n random variable's with expectation $E(X_i)$

$$Y = c_1 X_1 + \dots + c_n X_n \text{ Then } E(Y) = c_1 E(X_1) + \dots + c_n E(X_n)$$

Joint distribution

Two random variables characters, X and Y , which interested in their joint outcome. $(X, Y) = (x, y) \Leftrightarrow X = x, Y = y$.

The joint frequency functions

The values that discrete random variables X, Y take on x_1, x_2, \dots and y_1, y, \dots resp.

Joint frequency function: $p(x_i, y_i) = P(X = x_i, Y = y_i)$

Marginal probabilities $p(X = x_i) = \sum_j P(X = x_i, Y = y_j)$

The random variables:

similar story random variable's X_1, \dots, X_n are defined on the same sample space $p(X_1, \dots, X_n) = P(X_1 = x_1, \dots, X_n = x_n)$

$$p_{X_1}(x_1) = \sum_{x_2, \dots, x_n} p(x_1, x_2, \dots, x_n), p_{X_1, X_2}(x_1, x_2) = \sum_{x_3, \dots, x_n} p(x_1, x_2, x_3, \dots, x_n)$$

4.6. The geometric distribution

Suppose there is an experiment that repeats independent identical Bernoulli trials until Mathematicians observe the first success s . For example, flip a coin until Mathematicians observe the first heads. Since the first success may occur on the first trial, or at any time after the first trial, Mathematicians get that the sample space S of this experiment consists of an infinite but countable number of sample points (or outcomes) s_i of the f

$s_1 : S$ (a success on the first trial) $s_k : F F F \dots F s$ (success on k th trial). In this case, if Mathematicians define the random variable character X as the number of trials in which the first success is observed, Mathematicians note that the events $(X = 1)$, $(X = 2)$, and $(X = k)$ all contain only one sample point, s_1 , s_2 , and s_3 , respectively. In general, the event $(X = k)$ contains only the sample point s_k , and the possible values of the random variable X are $X = 1, 2, 3 \dots$. Now, since the trials are independent, if Mathematicians assume that the probability of trial success for several trial is equal to p , Mathematicians have this result for any x $p(x) = P(X = x) = P(F F F \dots F S) = p q q q \dots q p = q^{x-1} p$ where $q = 1 - p$. This observation leads to the following definition.

Definition: A random variable X is shown to have a Geometric distribution with success probability p , if and only if $p(x) = q^{x-1} p$, where $x = 1, 2, 3, \dots$ and $0 \leq p \leq 1$. Mathematicians write $X \sim \text{geometric}(p)$.

The mean of the geometric distribution

Suppose that $X \sim \text{geometric}(p)$. From the definition of the expected value Mathematicians obtain $E(X) = \sum_{x=1}^{\infty} x q^{x-1} p = p \sum_{x=1}^{\infty} x q^{x-1}$, where $q = 1 - p$. Now, because x is an integer, Mathematicians observe that for each term in the sum above, Mathematicians have $\frac{d}{dx} (q^x) = x q^{x-1}$,

and therefore, interchanging the order of the derivative and the sum, Mathematicians obtain $\frac{d}{dq} (\sum_{x=1}^{\infty} q^x) = \sum_{x=1}^{\infty} \frac{d}{dq} (q^x) = \sum_{x=1}^{\infty} x q^{x-1}$.

Now, Mathematicians observe that $\sum_{x=1}^{\infty} q^x = q + q^2 + q^3 + \dots$

corresponds to the geometric series, and from calculus Mathematicians know that $\sum_{x=1}^{\infty} q^x = q + q^2 + q^3 + \dots = \frac{q}{1-q}$. Then, Mathematicians obtain that $E(x) = p \sum_{x=1}^{\infty} x q^{x-1} = p \frac{d}{dq} (\sum_{x=1}^{\infty} q^x) = p \frac{d}{dq} (\frac{q}{1-q}) = p \left[\frac{1}{(1-q)^2} \right] = \frac{p}{p^2} = \frac{1}{p}$.

4.7. The Negative Binomial distribution

Consider the same setup for the geometric distribution. But Mathematicians are now interested not in the first success, but in the number of second, fifth, or, in general, eighth successes. If Mathematicians denote by X the number of r th successes, the first thing Mathematicians notice is that the previous $r - 1$ successes may occur in any order in the sequence of trials, so Mathematicians need to consider their possible permutations. Accordingly, Mathematicians can derive the random variable X 's probability mass function, which yields $p(x) = \binom{x-1}{r-1} p^r q^{x-r}$, $x = r, r + 1, r + 2, \dots$, which leads to the following definition If and only if, a random variable X is shown to have a negative binomial distribution $p(x) = \binom{x-1}{r-1} p^r q^{x-r}$, $x = r, r + 1, r + 2, \dots$, for a given integer r and $0 \leq p \leq 1$. Mathematicians write $X \sim \text{NegBin}(r, p)$.

4.8. The Hypergeometric distribution

One way to motivate the hypergeometric distribution is to assume that Mathematicians have a large urn filled with N balls that are identical in all respects except that r balls are red and $N - r$ balls are blue. If Mathematicians take n balls from the urn at random, what is the probability that x balls are red? Mathematicians observe that the all number of samples of size n that can be taken from the n balls in the urn is $\binom{N}{n}$.

On the other hand, Mathematicians can choose the x red balls in $\binom{r}{x}$ different ways, and the reminder $n - x$ blue balls in $\binom{N-r}{n-x}$ different ways.

Combining these results, Mathematicians obtain $P(x \text{ balls are red}) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$

which leads to the following definition.

Definition: A random variable character X is shown to have a hypergeometric distribution, if and only if $P(x \text{ balls are red}) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$ where $x = 0, 1, 2, \dots, r, n \leq N$ and $n - x \leq N - r$.

4.9. The Poisson distribution

Let's say Mathematicians want to determine the probability of a collision occurring at a specific crossroads over the course of a week. Divide the time period (in this example, a week) into n subintervals that are so tiny that Mathematicians can see at most one accident with a non-zero probability for each of them. This is one technique to tackle the problem. The probability of finding an accident in any given subinterval, denoted by p , then Mathematicians can get

$$P(\text{There are no accidents throughout the subinterval}) = 1 - p$$

$$P(\text{One accident happened in the subinterval}) = p$$

$$P(\text{more than one accidents occur within the subinterval}) = 0$$

In this instance, Mathematicians discover that the total number of accidents that take place in a week equals the total number of subintervals that Mathematicians record incidents. If Mathematicians assume that accidents occur independently in each interval, the all number of accidents obeys like one of distribution mentioned before, binomial distribution. The problem Mathematicians encounter here is that Mathematicians are unsure of how to determine the subintervals and therefore Mathematicians don't really know the values of n and p for the binomial distribution. However, according to intuition, as n rises, there are more subintervals., the probability that an accident will occur in one of these shorter intervals, p , falls. Mathematicians can formalize this idea by assuming that $\lambda = np$ remains constant as n increases. Then, if Mathematicians write the binomial Probability mass function in terms of the coefficient λ , Mathematicians obtain $\binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$, where $\lambda = np$,

which is the expected number of accidents for the entire week. Now, taking the limit as $n \rightarrow \infty$, Mathematicians have $\lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} = \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\dots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} = \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{x-1}{n}\right) = \frac{\lambda^x}{x!} e^{-\lambda}$. Because $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$, and all other terms in the product go to 1. This result leads to the following definition. Definition: A random variable X is shown to have a Poisson distribution if and only if $p(x) = \frac{\lambda^x}{x!} e^{-\lambda}$,

where $x = 0, 1, 2, \dots$ and $\lambda > 0$. Mathematicians write $X \sim \text{Poisson}(\lambda)$.

One of the most useful statistical distributions in the world is likely the Poisson distribution today for answering many questions. It has been in use for over a century. Use cases can cover a wide variety of problems in insurance, business, medicine, banking, risk management, and science.

Siméon-Denis Poisson, a French mathematician, created his function in 1830 to quantify the number of times with which a player succeeds on successive bets in a game of chance with few winners. If p denotes the probability of winning in any particular trial, the average number of wins in n trials (λ) is represent by $\lambda = np$. Using the binomial distribution of Jakob Bernoulli, a Swiss mathematician, and Poisson proved that the probability of receiving k victories is around $\frac{\lambda^k}{k!} e^{-\lambda}$.

Nowadays, the Poisson distribution is regarded as a highly significant distribution and has had very representative applications throughout history.

The British statistician R.D. Clarke presented his investigation of the distribution of flying bomb strikes in London during World War II in his 1946 publication, "An Applications of the Poisson Distribution." Certain sections were more exposed than others. The British Army was interested in learning if the Germans deliberately targeted certain regions, whether the strikes showed excellent technical precision, or whether the distribution was accidental. In reality, if the missiles were just randomly targeted, the British could easily disperse critical infrastructure to lessen the probability that it would be struck.

Clark split the region into thousands of pieces of the same size. There could not have been even one strike in any of these cases, let alone more. Additionally, if missiles fall at random, the probability that any plot will be struck is the same for all plots. In a lot of repetitions of a game of chance with a very tiny probability of winning, the total number of hits would therefore be fairly comparable to the total

number of victories. This justification prompted Clark to formalize the model's derivation using the Poisson distribution. The observed hit rate and the expected Poisson rate are fairly similar. According to Clark, the observed variance appears to have developed solely by chance.

5. Conclusion

When the statistical distribution is correctly applied to different scenes, Mathematicians can get the probability or law through calculation and analysis, so as to help us quickly analyze and apply the corresponding scenes in the future. It seems that they all calculate the same probability, probability density function, cumulative distribution function, etc., but different distributions are also found according to their own characteristics. Mathematicians need to combine different fields and societies with statistics, analyze the accumulated data, and then find patterns and model future data.

References

- [1] Anderson, W. J. (1991). Continuous-Time Markov Chains. Springer-Verlag, New York.
- [2] Adler, R. J. (1981). The Geometry of Random Fields., Wiley, New York.
- [3] Baldi, P., Mazliak, L. and P. Priouret. (2002). Martingales and Markov Chains: Solved Exercises and Elements of Theory. Chapman-Hall/CRC, Boca Raton.
- [4] Basawa, I. V. and B. L. S. Prakasa Rao. Statistical Inference for Stochastic Processes. Academic Press, London.
- [5] Billingsley, P. (2000). Convergence of Probability Measures, 2nd Ed., Wiley, New York.
- [6] Chung, K. L. (1967). Markov Chains with Stationary Transition Probabilities, 2nd Ed. Springer-Verlag, New York.
- [7] Chung, K. L. and R. J. Williams (1990). Introduction to Stochastic Integration, 2nd Ed. Birkhauser, Boston.
- [8] Embrechts, P. and Maejima, M. (2002). Selfsimilar Processes. Princeton University Press, Princeton.
- [9] Hida, T. and M. Hitsuda. (1991). Gaussian Processes. American Mathematical Society, Providence, R.I.
- [10] Kallenberg, O. (1976). Random Measures. Academic Press, London.
- [11] Karatzas, I. and S. E. Shreve (1991). Brownian Motion and Stochastic Calculus. Springer-Verlag, New York.
- [12] Lukacs, E. (1970). Characteristic Functions, 2nd Ed. Griffin.
- [13] Meyn, S. P. and Tweedie, R. L. (1993). Markov Chains and Stochastic Stability. Springer-Verlag, New York.
- [14] Petrov, V. V. (1995). Limit Theorems of Probability Theory. Oxford University Press, Oxford.
- [15] Samorodnitsky, G. and M. S. Taqqu. (1994). Stable Non-Gaussian Random Processes. Chapman-Hall/CRC, Boca Raton.
- [16] Shorack, G. R. (2000). Probability for Statisticians. Springer-Verlag, New York.
- [17] Durrett, R. (2019). Probability: theory and examples (Vol. 49). Cambridge university press.