Advances in Economics, Business and Management Research, volume 215
Proceedings of the 2022 7th International Conference on Social Sciences and Economic Development (ICSSED 2022)

Application Strategy of Probability Theory and Mathematical Statistics in Big Data Analysis

Hechenjin Yu

Zhengzhou No.7 High School, Zhengzhou, Henan, 450000, China
**Corresponding author. Email: 2387218357@qq.com*

ABSTRACT
Since the 21st century, the rapid development and promotion of the Internet has caused data to show a geometric growth, which has ushered in the era of big data in my country. Because big data has the characteristics of large scale, fast growth, and sparseness, this also brings great difficulties to big data analysis. In the era of big data, it is a simple and efficient method to use probability theory and mathematical statistics to analyze and mine complex data. In people's production and life, probabilistic knowledge has been widely used in all aspects. It is an important tool for a deeper understanding of the world. Through the application of probability and mathematical statistics tools, people can make calm and scientific analysis of various complex problems and data, so that people's quality of life can be significantly improved, and accurately predict the evolution law and development trend of things based on the existing data. It is precisely because of these advantages that probability theory and mathematical statistics have become the guidelines for many complex problems. Nowadays, people's demand for big data analysis is becoming more and more urgent, which also makes people urgently need an effective method that can be applied to big data analysis to solve complex problems in actual production and life. In view of this, the following will discuss the relevant application strategies of probability theory and mathematical statistics in big data analysis, hoping to provide corresponding reference suggestions for people's big data analysis in production and life.

Keywords: *Probability theory, Mathematics, Statistics, Big data.*

1. INTRODUCTION

Probability theory and calculus are two different branches of mathematics. Probability theory is a mathematical science that studies the statistical laws of random phenomena. Calculus uses the limit as a tool to study functions well. The idea of calculus not only runs through The entire study of advanced mathematics is also the cornerstone of building the building of probability theory. For example, under the action of mapping, the set is simplified to random events, and then the set is simplified to real numbers. When the sample space is simplified to the set of real numbers, the probability is corresponding It is approximated by a set function to a real function. The distribution function F (ming) is measured from the viewpoint of function. In probability theory, the distribution function F(x) has very good properties, such as monotone bounded, integrable, almost continuous everywhere, and almost always derivative, etc. Therefore, the idea of calculus in advanced mathematics can be smoothly applied to the field of probability theory. The distribution function of continuous random variables, the representation of probability density, expectation, variance, etc. All borrow or use the existing results of differential integration more or less. Another example is the use of limit theory in probability theory, such as the related properties of distribution functions, the central limit theorem, the law of large numbers, and random processes. All in all, the idea of calculus has penetrated into all aspects of probability theory. It can be said that without the promotion of the idea ofcalculus to probability theory, there will be no axiomatic and systematic probability theory, and probability theory cannot form an important part of mathematics branch. However, probability theory, as a course opened after advanced mathematics, is not a simple successor to advanced mathematics, but an extension of the idea of calculus, which opens up a new world of mathematics. The development path of probability theory and advanced mathematics. There is a big difference. Probability theory is advancing in the direction of random mathematics and has become a typical representative of random mathematics. It has the

same status as advanced mathematics, injecting fresh blood into the ancient mathematics discipline, enabling mathematics to advance with the times and further moving forward.

2. THE MEANING OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS

In advanced mathematics, probability theory and mathematical statistics are a kind of analysis with distinctive characteristics. It has very unique thinking characteristics in the research object, and it has a very close relationship with other disciplines, especially economics. The content of probability theory and mathematical statistics is very rich, which also makes it an important part of mathematics. At this stage, the methods of probability theory and mathematical statistics have been widely used in various fields. From the current point of view, probability theory and mathematical statistics can be regarded as a relatively independent discipline, which plays a huge role in people's production and life, whether in the industrial field or in other fields, probability theory and mathematical statistics methods. The requirements for information technology are very strict, and the use of probability theory and mathematical statistics has unparalleled advantages in big data analysis. At the same time, it does not belong to an independent discipline. This is because it has a close internal connection with other disciplines and has a mutual penetration effect. It is precisely because the coverage and application scope of probability theory and mathematical statistics is very wide, which also makes people It is difficult to explain them one by one. Therefore, this article only discusses the application strategies of probability and mathematical statistics in several aspects in order to clarify the specific applications and functions of probability theory and mathematical statistics in big data.

3. THE CLOSE CONNECTION AND COMMON METHODS OF PROBABILITY THEORY, MATHEMATICAL STATISTICS AND BIG DATA ANALYSIS

3.1. Probability theory is closely related to mathematical statistics and big data analysis

The advent of the era of big data has enabled people to use probability theory and mathematical statistics to analyze big data. This also makes it closely related to big data analysis. The connection is mainly concentrated in the following four aspects. First, probability theory. The research goals of mathematical statistics and big data analysis are the same, both are to explore and clarify the data structure in order to find out the internal connections and laws of big data[1]. Secondly, the continuous development of big data has enabled big data analysis to open up a new application space for statistics, which also provides a brand-new topic for the research of probability theory and mathematical statistics. Through the analysis of big data, it can be extremely promote the development of probability theory and mathematical statistics to a large extent. Third, big data analysis does not belong to a branch of statistics. Big data analysis can also be widely used in other fields, and can provide new ideas, tools and methods for other fields. For example, using big data analysis can make machines perform learn, and be able to realize data storage, etc. Finally, probability theory and mathematical statistics are a very widely used and mature problem-solving method and technology in DM, and they occupy an extremely important position in DM[2].

3.2. Commonly used methods of probability theory and mathematical statistics in big data analysis

There are two main methods commonly used in big data analysis for probability theory and mathematical statistics. One is the analytic hierarchy process, and the other is the Monte Carlo method. The so-called analytic hierarchy process refers to the evolution of certain uncertain factors. When conducting research on laws and development trends, it is necessary to comprehensively consider the influence and interrelationship of these factors[3]. Because these uncertain factors in the evaluation index can be divided according to levels, at the same time, the uncertainty in each level Sexual factors also contain several elements, which makes the structure of the entire complex problem appear to be a multi-level hierarchical structure. When solving such problems, the analytic hierarchy process can be used to analyze the differences in these levels. The deterministic factor judges the relative importance of the whole problem, and this produces probability. When applying the analytic hierarchy process, the mathematical model should be established through four steps. The first step is to clarify the various factors in the problem first, and then divide these factors into layers, so that the structure of the entire problem belongs to a kind of hierarchical hierarchy. Sub-structure, and then the elements of the upper level are used as criteria to compare the elements of the lower level in pairs, and determine the importance of the elements of the lower level to the elements of the upper level according to the evaluation scale, and construct the corresponding judgment Matrix, and then calculate the relative importance of each element in the problem, and calculate the comprehensive importance of the problem at the same time, and then bring reliable decision support guarantee to the decision maker. The Monte Carlo rule is to repeatedly and randomly sample the uncertain factors in the problem on the basis of probability theory and mathematical statistics, so as to simulate the degree of influence of the uncertain factors' own changes on the problem. The influence of all uncertain factors in the

problem on the problem is calculated and analyzed, and then the scientific analysis result is obtained. Monte Carlo method can simulate the actual process of the problem, which also makes it have a very significant effect in solving the actual problem[4]. The mathematical expression of Monte Carlo method isZ=k(x1,x2,x3,...,xn) , In this mathematical expression, xi(i=1,2,3,...,n) represents the existence of n mutually independent random variables in the complex problem, for example, in all uncertain factors that affect the problem , These uncertain factors are variables and have the characteristics of probability distribution, and the function of n variables is Z, and this is exactly the goal to be solved[5].

4. APPLICATION STRATEGIES OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS IN BIG DATA ANALYSIS

4.1. Application strategies of probability theory and mathematical statistics in economic data analysis

In the era of big data, the role of data in the economy is self-evident. Among various types of data, economic data is the most common type. The analysis of these economic data is very important for promoting social and economic development. significance. Since economic data exists in a low-density form on the Internet, this also brings greater difficulty to people's analysis of economic data. The use of probability theory and mathematical statistics to analyze economic data is a simple and effective method. For example, the normal probability distribution method is used to analyze economic data. This method can predict and describe the probability of continuous random variables. This probability method is also widely used in the field of economic and financial management. Using this method can enable people to quickly and efficiently analyze all relevant information of probability through probability theory and mathematical statistics, and to grasp the market economy in real time according to the analysis results, so that people can understand the laws of market economy, and Analyze more economic information from it, and use the help of this information to flexibly formulate and adjust subsequent decisions and plans. The economic market is unpredictable, but the changes are not too outrageous. The analysis of economic data must not only predict the law of change and development trend of the economic market, but also consider the risks in the economic market. Existence is the acquisition of benefits is not always static, but through the analysis of economic data, corresponding countermeasures can be found to avoid these problems. For economic risks, in order to avoid the occurrence of economic risks, the use of probability theory and mathematical statistics can effectively reduce the probability of occurrence of economic risks, and this is also the most commonly used response method. Taking stock investment as an example, using probability theory and mathematical statistics to analyze economic data, it can be clearly seen that the greater the number of invested stocks, the higher the probability of profit generation than the smaller the number of invested stocks. This is precisely obtained through the methods of probability theory and mathematical statistics. Therefore, in investment decision-making, more investors tend to diversify their funds into more stocks to reduce risk, and this allows investors to obtain a probability of profit It can be seen that probability theory and mathematical statistics play a significant role in economic data analysis[6][7].

4.2 Application strategies of probability theory and mathematical statistics in business data analysis

In the big data environment, the importance of business data to enterprises is self-evident. There is a certain connection between business data and economic data. Business data is a type of economic data, but economic data is not necessarily business data. When companies analyze business data, probability theory and mathematical statistics are the most commonly used methods. Take the probability of losing big customers in business data as an example to discuss the application strategies of probability theory and mathematical statistics in business data. First, a research model needs to be established[8]. When the model is established, the following conditions need to be ensured. One is that the basic attributes of major customers should be similar, and the churn data can satisfy the same churn functionf0(t). The second is that the distribution conditions of the churn data are all derived from the churn function exponential termexp(c,zi)T. Then find out which factors have a greater impact on the churn probability of major customers, logarithmic h0(t)According to the classification and set a specific time period, and then summarize the churn situation of large customers in a specific situation, and obtain the churn situation trend chart, and then calculate the standard function of the trend chart, namely F(t,)=f0(t)·exp(c,) , And then obtain a certain customer’s churn probability in a certain time and the position in the churn function, the customer’s churn probability in the [0,T] period is p=exp(-(t,)dt). The vector of p-dimensional regression parameters is c, and the vector of p-dimensional covariates is. And define the covariate as an influencing factor, and then complete the construction of the research model. Secondly, after the research model is established, the parameters and covariates must be selected, and then these selected regression parameters are calculated through the maximum partial likelihood function. Since there are many factors affecting key customer churn in the calculation process, if all the factors are defined as

covariates, the model will have more dimensions, which will greatly increase the difficulty of parameter estimation, which also makes the parameter estimation accuracy rate. Cannot be guaranteed. Therefore, these factors need to be selectively used. In order to determine the number of covariates, they should be filtered according to the results of mathematical statistics, so as to avoid errors[9].

5. CONCLUSION

In summary, the role of probability theory and mathematical statistics in big data analysis is very obvious. Nowadays, probability theory and mathematical statistics are no longer an auxiliary analysis tool in big data analysis, but a simple and simple tool. And efficient analysis method. Through the application of probability theory and mathematical statistics, the process, trend, and effect of various types of data in big data have become the object of analysis when people conduct data analysis. Facing the rapid growth trend of big data, the application of probability theory and mathematical statistics for big data analysis will be more helpful to promote the development of people's production and life, and promote the rapid growth of my country's economy.

ACKNOWLEDGMENT

I'm very happy to be able to put the paper to the end. I have unknowingly been writing at the desk for many days and nights, and I have to lament that there is no end to learning and time is like an arrow. Summarizing the studies, the first thing to be grateful to is my supervisor, Mr. Dr. Pierre Clare. During the course he provided, this tutor has a vain personality, serious academic attitude, and vision and insights, which have benefited me a lot.

Also, I would like to thank my parents and friends who have been silently caring about and helping my own on the road of learning. Your ardent expectation and selfless love have given me the spiritual strength to continue to fight against difficulties and to be aggressive, thank you!

REFERENCES

[1] Zhang Sisheng. On the application of probability theory in higher mathematics[J]. Mathematics Learning and Research, 2021(10):10-11.

[2] Pan Xiaofeng, Hu Kun. Talking about mathematical expectation from the history of probability theory [J]. Middle School Mathematics Monthly, 2021(8): 65-65+F0004.

[3] Deng Jinhong. Research on the Teaching Reform of Probability Theory Based on "Internet+"[J]. Journal of Kaifeng Vocational College of Culture and Art, 2021, 41(7): 109-111.

[4] Xiao Min, Xu Jing, Tang Yeyun. Thoughts on Teaching Reform of "Probability Theory and Mathematical Statistics"[J]. Educational Progress, 2021, 11(4): 1090-1094.

[5] Liu Bin, Yuan Yunyun, Niu Xiangyang. The main characteristics and implementation strategies of probability theory methods[J]. Journal of Jilin Institute of Chemical Technology, 2021, 38(3): 91-96+100.

[6] Chen Chen, Dong Huimin, Liao Beishui. Semi-structured Probability Argument[J]. Logic Research, 2021, 14(2):1-21.

[7] Yang Han, Zeng Lang, Chen Bo, Li Wenbao, Liang Shaofeng. Applied Teaching Reform of "Probability Theory" [J]. Market Research Information: Comprehensive Edition, 2021(4): 0169-0170.

[8] Xue Li. Discussion on the Application of Probability Theory and Mathematical Statistics in Data Mining[J]. China Broadband, 2021(12):117-118.

[9] Tao Yuyu. Counter-intuitive method in the teaching of probability theory[J]. Journal of Jiangxi Vocational and Technical College of Electric Power, 2021, 34(9): 75-76.