
The Teaching Research of Probability and Mathematical Statistics Under the Background of Big Data

Rui Chen^{1*} and Liang Fang²

^{1,2} College of Mathematics and Statistics, Taishan University, Tai'an, Shandong, China.

*Corresponding author email id: chenruimengting@163.com

Date of publication (dd/mm/yyyy): 30/04/2021

Abstract – The goal of this paper is to develop a new teaching system of probability theory and mathematical statistics, which is suitable for cultivating talents of big data analysis, based on students' own development in the background of big data era, using the method of organically combining case teaching, computer visual drilling teaching and practical feedback teaching, the students can finally understand the abstract mathematical knowledge in practical application, enhance the practical ability of students.

Keywords – Big Data, Probability Theory and Mathematical Statistics, Teaching Reform, Teaching Strategies.

I. INTRODUCTION

A. Big Data is a Product of the New Era

Big data is a data set with a large scale that is far beyond the capability of traditional database software tools in terms of acquisition, storage, management and analysis. It is characterized by massive data scale, rapid data flow, diverse data types and low value density [1]. Big data requires new processing models with greater decision-making power, insight discovery power, and process optimization capabilities to accommodate large volumes, high growth rates, and diverse information assets. The strategic significance of big data lies not in the mastery of huge data information, but in the professional processing of these meaningful data. In other words, if big data is compared to an industry, the key to the profitability of this industry lies in improving the ability to analyze and process data and realizing the value of data through analysis and processing.

B. The Era of Big Data has Higher Requirements for Talents

The times are progressing and technology is developing. With the emergence of big data in people's life, the data structure is becoming more and more complex and the data is developing towards the direction of mass. All walks of life are involved in the application of big data, which affects people's production and life. In the era of big data, data resources are the greatest value, so the government, scientific research institutions, enterprises, etc., attach great importance to the research on big data [2]. In order to make students more in line with the needs of the development of the times, it is necessary to let students master certain data processing ability. In this process, the importance of probability theory and mathematical statistics is more prominent. First, probability theory and mathematical statistics are widely used. The research object of probability theory and mathematical statistics is the random phenomenon, which can be seen everywhere in our daily life. The mathematical ideas and methods reflected in the random phenomenon can be seen in almost all departments of science and technology, industrial and agricultural production and national economy. Second, probability theory and mathematical statistics are the theoretical basis of statistics and data mining. With the advent of the era of big data, data analysis technology becomes particularly important. The value of big data lies in the laws it contains. Big for data processing, analysis, statistics and data mining, and its principle and method based on

probability theory and mathematical statistics, using the principle of probability and mathematical statistics of the attributes in the data statistics, analysis, and find out the law contains data, to create value, is the student must master the basic theory and knowledge, most of the students is an essential skill for future work and life. Thirdly, the combination with mathematical software makes the application of probability theory and mathematical statistics more convenient and extensive. Therefore, let the students master the operation of statistical software is also called a particularly important link. Therefore, in the era of big data, students can better improve their data processing ability by mastering the relevant contents of probability and statistics courses, so as to find the research direction in the mass data, accurately analyze and process the data, and achieve the desired effect.

C. The Existing Teaching Content of Probability theory and Mathematical Statistics is too Theoretical

Probability theory and mathematical statistics are disciplines that study random phenomena and objective laws in many fields such as industry, agriculture, medicine, society and economy [3]. Compared with other mathematics courses, the thinking methods needed by this subject in dealing with related problems are obviously different. When studying this course, students should change the thinking mode established in other mathematics courses and give full play to their analytical ability on the basis of their own understanding of the teaching contents of probability theory and mathematical statistics. It can be seen that when students learn probability theory and mathematical statistics, the abstractedness and complexity of the course content will be highlighted, which makes students feel that it is difficult to understand and apply the existing problem solving methods. When teachers explain the knowledge of probability theory and mathematical statistics to students, as long as they focus on the explanation of theoretical knowledge and ignore the explanation of thinking methods and application examples, students will find it very difficult to apply the theoretical knowledge of probability theory and mathematical statistics in colleges and universities in their future work practice.

D. The Existing Teaching forms of Probability Theory and Mathematical Statistics are Simple

The traditional teaching is based on the teacher's classroom teaching, which consists of three teaching modules: the explanation of new knowledge, the review of existing knowledge, and the practice and consolidation of new knowledge. In the process of teaching knowledge, students are passive recipients of knowledge. Mathematics course learning itself is very boring, and teachers focus on theoretical knowledge when imparting relevant knowledge, which leads to the lack of characteristics and unique thinking methods of the course itself, increasing the boredom of students, who lose their independent thinking of knowledge and do not understand the knowledge framework and connotation of the course [4]. In a word, the single teaching form and defective teaching methods make students lose interest in learning and enhance their aversion to the course.

E. The Existing Assessment and Evaluation Methods of Probability Theory and Mathematical Statistics are Backward

The current assessment method of the course does not achieve the real purpose of promoting learning by examination, nor does it reflect students' learning results comprehensively, objectively and truly. Curriculum examination is mainly adopt the method of the test conclusion, pass the examination in the form of a check students as well as the grasp of the relevant knowledge of teaching content, in addition, although the teacher to student's class performance and learning process to make the corresponding examination, but did not make it

according to certain proportion into the course grade, it will leave some of the students, they think that such courses as long as the assault learning is in right before the exam, and meet performance requirements, so that it is bad for students to deeply understand the course. In addition, this course is an applied course with strong practicality. However, the current course assessment and evaluation does not examine the practical content of students, so that the assessment and evaluation method of this course does not comprehensively evaluate and examine the students' mastery of the knowledge and flexible application. Therefore, it is still an important task for college teachers to optimize the assessment methods of probability theory and mathematical statistics courses and carry out corresponding teaching reform.

II. TEACHING STRATEGIES OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS IN THE ENVIRONMENT OF BIG DATA

A. *Mathematics Software Assisted Teaching*

At present, the statistical software mainly includes SAS, SPSS, S-plus and Mini-Tab, etc., each of which has its own characteristics. SAS and SPSS are the most commonly used in teaching. Teachers should choose mathematical software scientifically. The choice of teaching software should not only make the software meet the teaching needs of the course, but also make the software meet the learning needs of students in the future, so that the great role of mathematical software can be truly played. By giving full play to the powerful function of mathematical software, including data analysis, error analysis and so on, to train students' scientific calculation ability and practical ability. In the process of teaching probability theory, teachers should make students master the use of mathematical software to simulate some probability experiments. In mathematical statistics, it is necessary to calculate a large number of data, although know how to calculate, but caused by the data volume is too big can't specific operation, which requires teachers to guide students to use mathematical software to calculate, so as to solve some practical problems in production and scientific research, so that the students can make full use of modern tools to solve problems, improve their learning confidence, promote them more proactive in learning [5]. This is very important to cultivate application-oriented talents. To improve students' practical operation ability of statistical software, we can start from the following aspects: First, on the premise of ensuring the realization of course objectives, we can appropriately reduce the difficulty of theory and improve the interest of theoretical knowledge, so that students can accept it more easily; Second, in order to improve the application ability of statistical knowledge, students can be taught more descriptive statistics content, so that students can better conduct data frequency analysis, dispersion degree analysis, etc.; Third, statistical modeling and mathematical modeling should be integrated into the teaching process to make students more team oriented and improve their ability to solve practical problems through modeling. Cases in production and life should be introduced into the case base to better enable students to understand how to solve practical problems with the existing knowledge of probability and statistics.

B. *Adjust the Teaching Content to Meet the Requirements of the Big Data Era*

Teachers can integrate and adjust the contents of probability theory and take them as review content, so as to spare more time to teach students statistical content such as regression analysis and variance analysis, and constantly improve students' ability to deal with data, making this course more practical. For example, in the class, students are given practical cases of enterprises, so that they know how to apply these statistical methods

to solve problems in the actual process, such as improving old products, trialling new products, and improving the ratio of raw materials. It can also solve problems related to continuous production through statistical analysis of data, such as process control, reliability analysis of components produced in large quantities, etc.

C. Use Network Resources to Enrich Teaching Means

With the advent of the era of big data, a large number of systematic high-quality online teaching resources have emerged, and online education has also gained development. The ways for students to acquire knowledge are more diversified, which brings challenges as well as opportunities to traditional classroom teaching. Teachers need to integrate classroom teaching and network resources, open up the channel of online learning and classroom teaching, optimize the classroom teaching content, and form a model with classroom teaching as the priority and online learning as the supplement. Teachers teach teaching difficulties and key points in class to improve classroom efficiency; Part of the content will be given to students to use network resources for independent learning. Students will choose appropriate resources for independent learning according to their interests and levels. The development of technology has made it possible to reform the traditional teaching mode. Modern teaching methods such as computers, projectors and even smart classrooms are used to introduce and display knowledge points through videos, GIFs and images, so as to transform abstract into concrete and static into dynamic, enrich classroom contents and improve students' interest and effect in learning. For example, for the joint probability density function of two-dimensional normal distribution, we can use mathematical software to draw three-dimensional graphs. During the introduction, we can show them from different angles, so that students can have a direct understanding of the two-dimensional normal distribution and deepen their understanding of it.

D. Case Analysis Teaching

In teaching, we should highlight the idea of big data, introduce relevant knowledge points in the form of cases, and finally conclude and apply the knowledge points with case solutions, which can achieve twice the result with half the effort. When teaching the central limit theorem, we can introduce cases about crowded boiled water houses. In case 1, it is assumed that there are 16,000 students in the dormitory building, and only one boiled water room with 90 faucets is equipped. The boiled water room is open from 12:00 to 13:00 noon and 18:00 to 19:00 pm every day. Assuming that the average time for each student to turn on the tap every day is 1.2 minutes, any tap can be used randomly to turn on the water. Due to the large number of people opening water at the same time and the small number of faucets, there is often a phenomenon of crowded long queues opening water. If you are a school logistics manager, please try to solve the following two problems: (1) what is the probability that students will have to wait in line to turn on the water without increasing the faucet? (2) If there are more faucets, how many new faucets should be installed, which can alleviate the queuing phenomenon of students opening water and avoid the waste caused by installing too many faucets?

When studying binomial distribution, we can put forward such a case in advance: Case 2, 100 questions are all single-choice questions, and each question has four options. Can you pass this kind of exam by luck? When learning normal distribution, we can think about the following questions: Case 3: How is TOEFL score calculated? Exam score is a reflection of the examinee's level. At present, the percentile scoring method is widely used in China, that is, the full score is designed to be 100 points. The main disadvantage of this scoring method is that the score is greatly affected by the difficulty of the topic. If the exam is easy, it is likely that most

of the examinee's scores are above 85 points. This score of 85 is not necessarily a good result. From this point of view, the percentile system cannot fully reflect the actual level of candidates. It is a good way to evaluate the relative scores of candidates by ranking, or rank method. Rank method also has its shortcomings. Because the rank is related to the number of candidates, it is difficult to compare the third among 1000 people and the third among 10 people. In order to overcome the shortcomings of percentile system and rank method, the percentile score or rank can be changed to percentile. However, percentile also has its shortcomings, that is, the original test score cannot be determined according to percentile. A reasonable and internationally popular scoring method is the standard scoring method, from which TOEFEL score is calculated. Please explain the calculation method and principle of TOEFEL score in detail. Mathematical statistics is a branch with strong application. Its theoretical derivation and proof are difficult to understand, and the data are huge. Therefore, students often find it difficult to learn. If we put forward a case that is very close to the students' reality at the beginning and solve this problem through in-depth study of the course, we can greatly deepen students' understanding of knowledge and truly apply what they have learned [6]. At the beginning of the part of mathematical statistics, we can give a survey data about the current consumption situation of college students, and ask students to make analysis according to the data: (1) try to estimate the consumption level interval of boys and girls with 95% confidence level; (2) Five years ago, the average consumption level of college students was 1200 Yuan per month. Is this level significantly improved now?

E. Interesting Classroom Teaching

Students are often interested in some mathematical historical facts and anecdotes. Integrating the knowledge points of probability theory and mathematical statistics into relevant anecdotes can not only stimulate students' interest in learning, broaden their horizons, but also make students know the ins and outs of knowledge and get good results [6].

For example, when introducing the T distribution, you can introduce the relevant historical facts and tell the students that Cassette first discovered the T distribution. At that time, he was an employee of Guinness Brewery in Dublin, Ireland. When doing quality monitoring, he found that the Z-test had a big error when the sample size was small, so he improved the Z- test with T- test. Because the brewery did not allow employees to publish all wine-related results, however, he was allowed to publish the discovery of T distribution under the pseudonym without mentioning winemaking. Cassette was modest and thought he was not a math professional, but he was a student, so he published the relevant results under the student's pseudonym.

For another example, when introducing Bayesian formula, we can incidentally introduce Bayesian School and the debate between Bayesian School and frequency school, so that students can realize that determining prior probability is not as easy as imagined. More importantly, students can understand that academic standards are not unique, and there can be different schools, debates and different academic ideas coexisting.

In a word, it is very important to enhance the interest of probability theory and mathematical statistics. Besides the methods mentioned above, there are many methods, such as introducing related frontier topics and adding discussion links.

F. Establish Scientific and Reasonable Assessment Standards

Under the big data environment, in order to promote the teaching quality of probability theory and

mathematical statistics, scientific and reasonable assessment standards should be established. Specifically, students' final assessment scores can be composed of the following four parts, including test scores, thesis scores, practical scores and peacetime scores. There are many advantages in adding thesis scores to students' final examination scores. The process of students' thesis writing is also a process of students' reviewing their previous knowledge. It can also improve students' ability to consult and sort out data, improve their ability to analyze and solve problems, and improve their ability of language organization. Practical achievement is based on students' performance in experimental class and their participation in ordinary practical activities. The usual grades include many contents, including students' attendance, performance in class, and completion of homework after class, etc. Assessing students by these methods is conducive to obtaining comprehensive and objective assessment results, better understanding of students' learning situation, and helping teachers to correct their own teaching methods, etc., which meets the needs of society for cultivating talents in colleges and universities under the background of big data era, and is also conducive to cultivating high-quality mathematics talents for society.

III. CONCLUSION

With the progress of science and technology, big data has gradually penetrated into various fields, and its analysis and application serve more and more fields of society. Under this background, training applied talents is a very important teaching goal of schools. The flexible application of probability theory and mathematical statistics will bring new development and effective utilization to the research of big data. In short, teaching includes teaching and learning. In the process of teaching and learning, it is necessary to constantly change the teaching methods and contents according to different students and different courses, teach students in accordance with their aptitude, and stimulate students' learning enthusiasm. The rapid development of the times will also prompt us to constantly improve teaching methods and methods, improve assessment methods, and strive to improve students' comprehensive quality [7]. In the future, we intend to apply more information technologies such as micro-courses and network-based PBL teaching to the teaching of probability theory and mathematical statistics to further improve. In the era of big data, data grows in geometric multiples, such as large scale, fast growth and sparsity, which will hinder the development of big data, in this paper; we guide the use of probability theory and mathematical statistical analysis methods to solve the problems encountered in the development of big data [8].

ACKNOWLEDGMENT

The work is supported by Tai'an Science and Technology Innovation Development Project (policy guidance): case-based teaching strategies for probability and Statistics in secondary schools (2020ZC355) and the 12th teaching reform and research project of Taishan University (201914).

REFERENCES

- [1] X. Zhai. Research and analysis of probability theory and mathematical statistics based on big data. *Science and Technology Economics Guide*, 2016(25):141-142.
- [2] Z.W. Chen, S.R. Hui, Y. Zheng. Exploration and practice of teaching reform of probability theory in colleges and universities]. *Journal of Shenyang Agricultural University*, 2011 (13): 331-334.
- [3] Y.H. Gu. Analysis of Curriculum Reform of Probability Theory and Mathematical Statistics in Big Data Era. *Curriculum Education Research*, 2017(2):243.
- [4] L.D. Guo, L.B. Wu. research and practice on teaching reform of probability and mathematical statistics in the era of big data. *education and teaching forum*, 2018(16):149-150.
- [5] Y. Wan, K. Peng, S.J. Peng, et al. Exploration of mixed teaching of probability theory and mathematical statistics from the perspective of informationization. *University Education*, 2019(10):95-97.
- [6] X.F. Zhao. Exploring the teaching reform of probability theory and mathematical statistics in independent colleges under the

- background of transformation. Education and Teaching Forum, 2019(39):106-107.
- [7] X.M.Wang. Probe into the teaching reform of probability theory and mathematical statistics in private universities. Mathematics Learning and Research, 2016(07):24+26.
- [8] Z.N. Li. Application of probability theory and mathematical statistics in big data analysis. Digital World, 2021(03): 65-66.

AUTHOR'S PROFILE



First Author

Rui Chen, is a associate professor at Taishan University. She obtained her master's degree from Shandong University in December, 2009. Her research interests are in the areas of application of probability theory, and applied statistics in recent years.



Second Author

Liang Fang, is a professor at Taishan University. He obtained his PhD from Shanghai Jiaotong University in June, 2010. His research interests are in the areas of cone optimization, and complementarity problems. email id: fangliang3@163.com