The Bayes Formula Case Design and Practice under "Research Teaching"*

Y.LI
Basic Experimental Center for Natural Science, University of Science and Technology Beijing, Beijing, China
L.T. ZHAO, Y.X. FAN, B.X. LI & Q. YU
School of Mathematics and Physics, University of Science and Technology Beijing, Beijing, China

ABSTRACT: The purpose of Research teaching is to train students in learning and capacity of innovation. This paper, based on the Bayes formula, discusses the design and practice of the case under the research-oriented teaching. Combining knowledge, we designed cancer diagnosis, automatic text correction, business decisions in three cases. We combined classroom teaching with real life and integrated the modern teaching methods, such as animation and computer simulation. Creating a teaching route, which is starting from problems then analyzing for solutions, and further extended to advance the level of knowledge, we aim to arouse students' enthusiasm and initiative, improve the effect of teaching, and provide a reference for the development of research teaching.
KEYWORDS: research teaching; the Bayes formula; case

1 INTRODUCTION

Society put forward higher requirements for quality of talent with the advent of technological development and the information age. As the talents training base-university, it needs to reform the traditional teaching model and establish new teaching ideas in order to train innovative talents to adapt to the world. Research teaching is proposed precisely conform to the requirements of the times. Compared with traditional teaching, research teaching can greatly stimulate students' interest in knowledge, improve students' enthusiasm and initiative to broaden their horizons and develop students' awareness of innovation as well as develop a scientific spirit and attitude.[1-2] Research teaching has already been the tendency of universities to improve the quality of talents. Research teaching is diverse from formally teaching. Case method is an effective method.[3] The purpose of this method is to encourage students to think, to enable students to grasp the initiative in the classroom and improve learning motivation. Therefore, how can the case be successfully applied to teaching practice? Case selection is critical and the quality of case directly affects teaching effectiveness. The cases which are selected covering knowledge at the same time, not only help students understand the knowledge but also stimulate students' interest.

Bayesian formula is an important part of the probability and statistics and is also the core and theory cornerstone in Bayesian statistics. It occupies a very important position in the whole study. The Bayesian formula is applied widely. It has a value of practical application in the financial, internet, medical, engineering and other fields, such as risk assessment, failure diagnosis, forecast and alarm. In traditional teaching, there are some drawbacks: formula complex, difficult to understand, the role of fuzzy, leading students don't know how to apply what they have learned. Therefore, this paper, based on the Bayes formula, discusses the design and practice of the case under the research teaching.

2 THE CASE DESIGN AND PRACTICE OF BAYES FORMULA

2.1 *Cancer Diagnosis*

2.1.1 *Problem*

Cancer diagnosis is a common example in the textbook, which some scholars have done research on this problem.[4] It spent much time for us on the problem how we have our own characteristic in the teaching design. In order to arouse students' interests and improve the efficiency of learning, we use

* This work is supported by Beijing Higher Education Young Elite Teacher Project (No.YETP0386).

computer simulation and animation demonstration in the case as below.

Case 1: The morbidity of a certain cancer in some district is 0.005. If someone has cancer, he has a 95% chance of testing positive for cancer, and if one doesn’t have it, the test will be positive 4%. Now we select one person in the random sample inspection, and the test is positive. Question: How much is the probability of that people suffering this cancer?

2.1.2 *Analyzing for solutions*

Assuming that *A* denotes the positive reaction to the test, and *C* denotes the person with this cancer in random sample inspection, we have $P(C)$=0.005, $P(A|C)$=0.95, $P(\bar{C})=0.995$, $P(A|\bar{C})=0.04$ in this case. Using the Bayes formula, we can obtain $P(C|A)$ =0.1066.

2.1.3 *Extension*

From the above analysis, the probability of the person who having cancer given that one tests positive is just 0.1066. That means only less than 11 persons in every 1000 person are cancer patients. Therefore, the meaning of the test may be questioned. Before the test, the morbidity of this district is 0.005 according to the clinical statistic which is prior probability. But after testing the positive result, the probability of people’s getting cancer is 0.1066 which is posterior probability. Compared to the former, the latter increases 21 times. The result shows that the test is meaningful for cancer findings. It adds the risk of cancer if the test is positive.

In order to study the statistical regularity, a large number of repetitive experiments are needed. Obviously, this cannot be achieved in the limited class hours. So we simulated the cancer diagnosis problem by using Matlab, and compared it with the theoretical result, as shown in Figure 1.

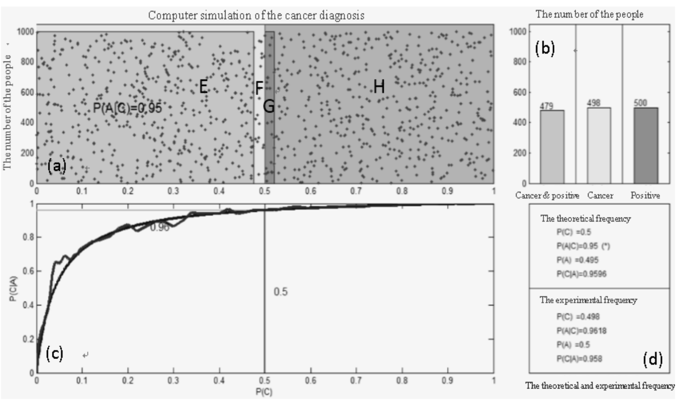

Figure 1 Computer simulation of the cancer diagnosis

The experiment simulates the morbidity dynamically from 0 to 1, and use scatter diagram, the frequency of cylindrical bar and probability value of experiment reflect the changes. In Figure 1, the 1000 random spots represent 1000 people. In Figure 1a, the area E shows the cancer patients who react positive to the test, H shows normal people; F shows the cancer patients who react negative to the test, while G shows the normal people who react positive to the test. Figure 1b shows the frequency of three groups: the cancer patients who have a positive result, cancer patients, and people who have a positive result. Figure 1d shows the comparison of the experimental frequency and the theoretical probability. The dependence of the conditional probability of having cancer given that one test positive $P(C|A)$ on the morbidity $P(C)$ is shown in Figure 1c. The relatively smooth curve is the theoretical probability, the other one is the experimental frequency. Obviously, there is a good agreement between the two curves.

Figure 1c illustrates that lower morbidity vs lower suffering probability when positive reaction. Just like the question above, when the general morbidity is 0.005, the suffering probability is only 0.1066 if the test is positive, because cancer is very rare. If the general morbidity was 10 times reduced, the suffering probability given the positive reaction would be 10 times reduced too. It means, for a rare disease, even if your test presents positive, the suffering risk is very little. As a result, the rarer is the disease, the more difficultly for you making definite diagnosis. At that moment, further examination is necessary. However, if the general morbidity raise up to 0.3, the suffering probability under the positive reaction will increase to 0.91. Therefore, for some diseases with higher morbidity, such as flu, there is much possibility that you will suffer this disease when you react positive to the test and to some extent that you are diagnosed surely.

2.2 *Text automatic error correction*

2.2.1 *Problem*

Case 2: Play a video, raising the question: in a word document, when you input a wrong word, why document will give tips, and the order of the tip is how to be determined, as shown in Figure 2. If the user accidentally mistyped, entering theu, then the user may input the, they, them, then, thru, but what is the automatic error correcting sorting based on?

It’s familiar to us about what the example concerned, now we explain this problem by Bayes formula. Let's input *theu*, and the original plan is to input one of the five words we mentioned above, then we want to find out that, under the condition of that the result is *theu*, how much is the possibility of each word input respectively? At the same time, sort for them.

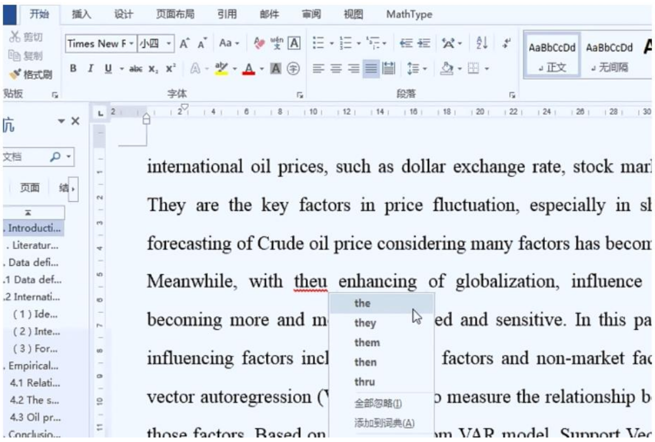

Figure 2 the description of the text error correction

2.2.2 *Analyzing for solutions*

Assuming that inputting *theu* is the event A, event B_i (i=1,2,…,5) represents the five words. According to the word frequency statistics, we can get the probability $P(B_i)$ (i=1,2,…,5) of events $B_1 \sim B_5$, and the probability of A under the conditions B_i, namely $P(A|B_i)$,i=1,2,3,4,5. Then using Bayes formula to get the probability of B_i under the condition A, namely $P(B_i|A)$,i=1,2,3,4,5, then determine the automatic error correction order according to the value of the probability.

2.2.3 *Extension*

Mathematical knowledge comes from life, for life. Actually, there are a lot of the applications of mathematical knowledge around us, while students are accustomed to accepting, not to discover it. So how to dig out of the applications of Bayes formula from daily life? This is one of the bright spots of this subject. The example is beneficial to develop the students' subjective initiative, to cultivate students' ability of finding the problem. At the same time, the students have a certain understanding of the application of Bayes formula in artificial intelligence.

2.3 *Business decision case*

2.3.1 *Problem*

Case 3: In order to improve the quality of products, a company decision makers consider increasing the investment to upgrade the production equipment. There are two kind of opinions: department θ_1 thinks that after upgrading equipment, the probability of the superior products can reach 90%, while department θ_2 thinks that the probability of the superior products can reach 70%. According to the situation that recommendations were adopted in the past, policymakers think that the two departments' credibility are 0.4, 0.6 respectively. Then how should the decision makers decide?

2.3.2 *Analyzing for solutions*

By the subject, we know $P(\theta_1)$=0.4 and $P(\theta_2)$=0.6. But only based on past experience, the prior probability is subjective, and credibility is not high, so we need the further test.

1) test A, create 5 products in accordance with new equipment plan, and all are superior products.

By the subject, we know that, $P(A|\theta_1)$=0.4, $P(A|\theta_2)$=0.6. Then Using the Bayes formula we obtain: $P(\theta_1|A)$=0.700, $P(\theta_2|A)$=0.300.

Adjusted by the experimental results, the posteriori probability is more inclined to the first opinion (θ_1), but still not enough to make policymakers make a decision. In order to increase the credibility of the conclusions, we need do one more again.

2) test B, create 10 products and including 9 superior products.

At this moment, don't throw away all the information of test A, then regard the posteriori probability of the first test as the prior probability of the second test. Repeat the above process, and we can get $P(\theta_1|B)$=0.883, $P(\theta_2|B)$=0.117 by using the Bayes formula.

According to the two kinds of results, policymakers adjust the reliability of the two departments from the original 0.4 and 0.6 to 0.883 and 0.117, respectively. So the policymakers decide to introduce new equipment vigorously in order to expect that the probability of the superior product reach 90%.

2.3.3 *Extension*

The design purpose of this case includes two aspects, as below:

1) Students have a further understanding of the prior probability, posterior probability and their relations. Prior probability is the probability that based on past experience and analysis, and is often subjective. While the posterior probability is the probability that based on new information and more close to the practical situation after correcting the original prior probability. The prior probability and posterior probability is relative. If they introduce new information, update what is now called the posterior probability, and get a new probability values, then the new probability value is called the posterior probability.

2) Lead to Bayes decision and Bayes statistics, expand aspects of knowledge, and promote knowledge level.

As mentioned before, the prior probability in this case with obvious subjectivity, in order to make more accurate judgments, just need to introduce new information, and then correct by Bayes formula. At this point, the posteriori probability combine the prior probability of the original sample with new information, the results obtained are close to the reality. The development of this idea is Bayes decision and Bayes statistics.

Bayes decision theory is an important part of the subjective Bayes theory of induction. Bayes decision is the decision that under the incomplete information situation, for some unknown states, policymakers estimate with subjective probability estimation, then using the Bayes formula to modify, and finally using the expected value and the amended probability to make optimal decision.

Statistics is based on the sample data collection (equivalent to the event A here) to find the answer to the problem of interest. This is a process of 'find out the reason from the result'. Therefore, the Bayes formula has great application value. In fact, according to the thought of this formula, people developed a set of statistical inference method, called 'Bayes statistics'. Now, the Bayes statistical occupies half of the country in the field of mathematical statistics.

3 CONCLUSIONS

This paper discusses Bayesian formula for research-oriented teaching mode design and practice, three examples have different emphases and achieved good effective.

1) Carefully designed cancer diagnosis problems. The design purpose of this example, on the one hand, takes into account it's typical. In concrete teaching process, it can inspire students to think independently, active participation, further summarized the general solution to this problem and make them know by analogy through interaction. On the other hand, it forms a lively and interesting teaching atmosphere by taking modern teaching methods, such as computer simulation, and stimulates student interest.

2) Animated text correction problem. This case close to the daily life and easy to accept and understand. We dig out the Bayesian formula from familiar problem, which is a highlight of this title. This example cultivate students' initiative and the ability of finding problems. What's more, in the design process, we used multimedia assisted teaching methods through dynamic video to stimulate the sense of students to improve students' interest.

3) Develop and improve --- business decision making problem. Through the example of solving, first of all, it deepen students' understanding of difficult understanding---prior probability and posteriori probability. Secondly, it can lead to cutting-edge science---Bayesian statistics and Bayesian decision so that it can enhance the level of knowledge. The introduction of the case help the students to understand the Bayes formula disciplines involved in the latest development, which laid the foundation for further study.

From the practice, students have high motivation and a greater emotional investment in the teaching process of the unit and we access to the ideal teaching effect. Meanwhile, the implementation of research teaching methods help students really appreciate the practical value of Bayesian formula, fully mobilize the students initiative, cultivate students' innovative consciousness and improved students' ability about finding problem, applying probability knowledge to solve problem.

REFERENCES

[1] Lin, H., Li. H., Yuan, J.M. & He, Y. 2013. On the Theory and Practice of Research Teaching in the Universities. Journal of University of Science and technology Beijing (Social Sciences Edition) 29(1):94-99.

[2] Zhao, H.Q., Zhao, S.K. & Yu, P.M. 2007. On Research-Oriented Teaching Model in Research-Oriented Universities. Higher Education of Sciences 3:101-104.

[3] Chen, H.J., Sha, R.F. & Liu, M.H. 2012. Trough Case Analysis Upgrading Teaching effect of Probability Theory and Mathematic Statistics. Journal of Northeast Agricultural University (Social Science Edition) 10(1):98-100.

[4] Yan, J.L. 2012. Expound of Test Results Based on Bayesian Formula. College Mathematics 28(3):111-113.