Teaching Design and Research of Bayes Formula Based on Case Introduction for Probability Theory Course in Universities

Xin Wang[1, a *] and Jingshu Wang[2, b]

[1]Software Park Road No.8, A1 102a, Dalian Neusoft University of Information, Liao Ning, China
[2]Software Park Road No.8, Dalian Neusoft University of Information, Liao Ning, China

[a]wangxin@neusoft.edu.cn; [b]wangjingshu@neusoft.edu.cn

Keywords: Bayes formula; Case introduction; Medical diagnosis; Mathematical modeling.

Abstract. Conditional probability is error-prone content for beginners, especially facing the concept of disjoint events. This paper designs the application cases in each aspect of teaching, hoping to stimulate students' initiative and cultivate students' ability to ask and analyze questions. Further more, let us appreciate the fascinating charm of the mathematical world.

Introduction

The conditional probability formula is easy to get trapped for newcomers who learn probability theory. Some students confuse conditional probabilities with ordinary probabilities in the process of establishing mathematical models for practical problems. *Total Probability Theorem* and *Bayes Rule*, based on conditional formula, are more complex. Especially for Bayes formula, using backward conditional probability which examines reverse thinking and contrary to intuition, has become a difficult point in teaching. On the other hand, Bayes rule is one of the most widely used rules in all mathematical discoveries. [1] We will combine our teaching process with specific application cases close to production and daily life, aiming to train students'abilities of analyzing problems and mathematical modeling, at the same time, to enjoy the fun of mathematics.

Teaching Goal

Understand Bayes formula; use Bayes formula to solve practical problems; exercise the unique thinking of random mathematics; appreciate the fascinating charm of the mathematical world.

Teaching Preparation

Preview. According to our teaching experience and students'feedback, the class sometimes polarize in the process of learning Probability and Statistics. Due to the limitations of teaching arrangements and students'basic level, in order to enrich learning, we try to find the contents and ways that are suitable for them, both in class and under class. We arrange pre-training sessions before the class. They have learned the conditional probability formula, the multiplication formula and the total probability formula in the previous lesson. Therefore, we arrange two types of questions for students to preview before class: one is basic *conceptional* problem around the new knowledge, the other is a relatively complex *application* problem.

Question 1:

(a) Tell the difference among $P(A), P(A|B), P(B|A), P(AB)$. How do they derive from each other?

(b) What is the difference between the Total Probability Formula and the Bayes Formula? What problems can they solve?

The former can be used as a targeted review, and the latter can be used as a reference for examining the effects of the preview, which is also the teaching goal of this lesson.

Question 2:

  DOI: 10.25236/icess.2019.118

(a) A car crash occurred in a city. The city's car only has two colors, 15% is blue and 85% is green. A witness identified it a blue car. However, according to experts on-site analysis, the possibility that the conditions can be seen correctly at that time is 80%. So, what is the probability that the accident car is a blue car?

(b) A villa has had two thefts in the past 20 years. The owner of the villa has a dog. The dog barks three times a week. The probability of the dog barking during the thief's invasion is estimated to be 0.9. What is the probability of an invasion when the dog is barking?

Students can choose one of the provided cases for a brief explanation, analyze the meaning of the question, list the formula, and do not need to solve. Proper blackboard writing is required, and PPT or software aid can also be used according to their own needs. All contents are for reference only. Better cases are welcome, as well as other professional cases can be selected in combination with professional knowledge. Students who perform well will earn performance scores.

Teaching Processing

Case Introduction. The 2015 Nobel Prize in Diagnostic Medicine was presented to a group of researchers who studied appendicitis by studying the deceleration zone.

In the early stage of appendicitis, many people's performance does not seem to be characteristic, and the general auxiliary examination is not well diagnosed. Although there are a series of methods of examination, doctors hope to find more clues about the diagnosis of appendicitis. Some imaginative doctors turn their eyes to the speed bumps on the road.

People won't like the jolt in the speed bumps, and for patients with appendicitis, this jolt may aggravate their pain. Several researchers from Stoke Mandeville Hospital and Oxford University conducted a statistical study. In the questionnaire survey of patients with suspected appendicitis (64), people who felt that the pain was aggravated when they passed the speed bump were judged to be "positive with the deceleration zone". Those who did not feel pain or could not remember were classified as "negative with the deceleration zone".

Among these respondents, 54 were positive. It turned out that this diagnosis seems to have a good sensitivity (97%) as well as a specificity of 30%. If the symptom of the deceleration zone is positive, what is the probability of suffering from appendicitis? [3,4]

Bayes Formula. By the analogy of the case of "taking balls from the box" given in last lecture, this example is concise. The division of the Complete Event Set is relatively simple, and has a lot to do with the content of the previous lesson, which is easy for students to accept. The Bayes Formula finds a conditional probability, with the probability of result is already known, to find the probability of the reason. During the teaching, we focus on setting up disjoint events that form a partition of the sample space. In the past, students often pay more attention on calculations than analysis. The process of building models can draw inferences and innovative thinking. The ability to find and analyze problems is as important and valuable as the ability to solve problems.

Case Review. Let A be event that "suffering from appendicitis", so event $\bar{A}=$ "not suffering from appendicitis", given $P(A)=p$, event $B=$ "positive with the deceleration zone". According to the statistical results, $P(B\mid A)=0.97, P(B\mid \bar{A})=0.3, P(B)=\frac{54}{64}\doteq 0.84.$

Using Total probability formula:

$$P(B)=P(A)P(B\mid A)+P(\bar{A})P(B\mid \bar{A})$$

$$=0.97p+0.3(1-p)=0.67p+0.3 \tag{1}$$

Get the solution that $p\doteq 0.81$.

Therefore, if the diagnosis is positive, the probability of having appendicitis is:

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

$$=\frac{0.81\times 0.97}{0.84}\doteq 0.94 \tag{2}$$

This kind of diagnosis seems to be sensitive. While it must be noted that the prevalence rate of the respondents is high (up to 0.81), relatively the sample size is small. The patient may also have memory deviation, etc. so the data is referenced. More experimentations are needed. At the same time, even if the diagnosis is negative, there is still the possibility of suffering from appendicitis, the probability is

$$P(A|\bar{B})=\frac{P(A\bar{B})}{P(\bar{B})}=\frac{P(A)-P(AB)}{P(\bar{B})}$$

$$=\frac{0.81-0.81\times 0.97}{1-0.84}\doteq 0.16 \tag{3}$$

There are 16% of false negatives, which is, there are real patients who are mistakenly excluded.

Case Promotion. If there is a rare disease (1% in the population), assuming no clinical symptoms. The accuracy of the test is 95%, that is, the true patient has a 95% probability of being positive; and the unaffected person has a 95% probability being negative. Once the test result is positive, what is the probability of being sick? [1]

Similar to the previous case, let event $C=$"sick", $\bar{C}=$"not sick", meanwhile $D=$"being positive", $\bar{D}=$"being negative", $P(C)=0.01, P(D|C)=P(\bar{D}|\bar{C})=0.95.$

In the same way,

$$P(C|D)=\frac{P(CD)}{P(D)}=\frac{P(C)P(D|C)}{P(C)P(D|C)+P(\bar{C})P(D|\bar{C})}$$

$$=\frac{0.01\times 0.95}{0.01\times 0.95+0.99\times 0.05}\doteq 0.16 \tag{4}$$

Although the accuracy of the test is as high as 95%, why is the probability of illness 16%? That is the *counterintuitive* situation we mentioned. The reason is that the basic probability of illness is extremely low (0.01), the lower the basic probability, the more likely the diagnosis is misdiagnosed, called a *false positive*. In this case, further check is necessary for the person who is positive. Conversely, if it is a disease with a high incidence, as long as the test results are positive, the possibility of being sick is greater, such as the example of appendicitis. The Bayes formula can help us sort out these problems, especially the basic probabilities are different and often contrary to our intuition.

Exercises. We will apply Bayes formula to complete some more cases shown in the preview section, to improve the sensitivity of random thoughts, exercise modeling ability and innovative thinking.

Course Summary. The key point of this lesson is the Bayes formula, which combines the conditional probability formula, the multiplication formula, and the total probability formula. The related problems are more abundant, and the use of the Bayes rule is not only this. We hope that when facing complicated problems, Bayes formula can help you build mathematical models, optimize the analysis process effectively and flexibly.

At the same time, we also hope that everyone can feel the fun of mathematics and apply it in daily life. For instance, in the era of *information fragmentation*, whether we are often referred to the "headline" case from the network, instantly reverse the understanding of a person or even a field, but neglect the tendency of statistical information. Just because the former is more vivid and more impactful, the overall statistical information is relatively abstract. In the era of big data, what kind of vision should we choose, facing massive data, please think a lot.

Summary

Probability and Statistics is related to life closely. Appropriate use of cases in teaching can not only enhance students' interest, but also motivate everyone's learning initiative. It is not only in terms of knowledge, but more importantly, the education of ability has a longer term meaning.

References

[1] P. Olofsson: *Probabilities: The Little Numbers That Rule Our Lives* (Wiley, U.S. 2015).

[2] Z. Sheng, S. Q. Xie and C. Y. Pan: *Probability and Statistics* (Higher Education Press, China 2008 In Chinese)

[3] H. F. Ashdown: *Pain over speed bumps in diagnosis of acute appendicitis: diagnostic accuracy study* (BMJ 2012; 345: e 8012)

[4] https://www.guokr.com/article/440736/

[5] W. J. Yan: *Applied Probability Statistics* (Neusoft Electronic Press, China 2015 In Chinese)