Journal of Electronics and Information Science (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2024.090101
ISSN 2371-9524 Vol. 9 Num. 1

Application and Extension of Bayes Formula

Wang Jianxin

College of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao, 266061, China
wjxqd@163.com

Keywords: Bayes formula; Posterior probability; Signal processing; Software fault diagnosis; medical diagnosis

Abstract: Bayes formula is an important formula in probability theory and mathematical statistics. It has a wide range of applications. Bayes formula continuously adjusts prior information to a posterior probability under the condition of adding new information. This article deeply analyzes the Bayes formula and applies it to probability prediction in different fields such as signal processing, software fault diagnosis, medical diagnosis, etc. The Bayes formula provides an effective probability prediction method in practical applications.

1. Introduction

Bayes formula is an important formula in probability theory and mathematical statistics, as well as the theoretical foundation of Bayesian statistics. It is widely applied in various fields of life. And it is closely related to other scientific fields such as information theory, the Internet, artificial intelligence, queuing theory, control theory, biostatistics, medicine, and finance. [1,2,3] It plays a crucial role in probability calculation problems. The Bayes formula is mainly used to calculate the probability of finding the causal event from the results in complex events. It is essentially a comprehensive application of conditional probability, multiplication formula, and full probability formula, which calculates the probability of the causal event under the condition of known results. Due to the complexity, difficulty in memorizing, and widespread application of formulas, Bayes formulas are a key and difficult point in probability theory and mathematical statistics teaching. Many students' understanding of Bayes formulas is only superficial. And they only memorize formulas. It is difficult to truly understand the essence and connotation of Bayes formulas. It is difficult to be applied in practice. In order to deeply understand the theorem and proficiently apply it to solve practical problems, this article extends the application of Bayes formula.

2. Bayes Formula and Its Ideas

2.1 Bayes Formula

Let $A_1, A_2, \cdots, A_n$ be a partition of the sample space Ω, that is, $A_1, A_2, \cdots, A_n$ is incompatible with each other, and $A_1 \bigcup A_2 \bigcup \cdots \bigcup A_n = \Omega$. If $P(B) > 0$ $P(A_i) > 0$, $i = 1, 2, \cdots, n$, we have

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(A_j)P(B|A_j)},\quad i=1,2,\cdots,n$$

The above equation is called Bayes formula [4].

2.2 The Thought Implied in Bayes Formula

In the Bayes formula, event A_i can be understood as the "cause" that causes event B to occur. $P(A_i)$ represents the probability of various "causes" occurring. $P(A_i)$ is called the prior probability of event A_i. $P(A_i|B)$ reflects a new understanding of the probability of various causes after the occurrence of event B. $P(A_i|B)$ is called the posterior probability of the event A_i. The posterior probability refers to the probability of modifying and adjusting the prior probability after the occurrence of event B (new information), which is a reacquaint of the cause. The Bayes formula quantitatively depicts this change. It is the process of forming a new understanding of the cause event A_i after the occurrence of event B (new information). It is a process of continuously utilizing new information to form a scientific understanding.

When using Bayes formula to predict the probability of unknown event A_i, a prior probability $P(A_i)$ is given based on existing information and statistical data. Then, the probability of unknown event A_i is iterated and adjusted in a timely manner as new information (event B) accumulate and update continuously. When there is enough new information, the predicted probability of event A_i will approach the true value. In daily life, Bayes formulas are also unconsciously used for decision-making. For example, in a new environment, how to find the restaurant that best suits one's taste; When the rainy season arrives, we see cloudy weather and determine if it will rain; Judging whether grapes are sweet or not based on experience when buying grapes, and so on. All of these judgments in daily life contain the ideas of Bayes formula.

3. Exploration on The Application of Bayes Formula

3.1 Application in Signal Processing

In signal processing, it is necessary to study the encoding, transmission, and decoding processes of signals. During this process, decoding errors may occur. Bayes formulas are widely used to calculate the probability of errors in the signal encoding and decoding process [5].

Case 1: Given that the signal transmitter encodes 0 and 1 with probabilities of 0.6 and 0.4 respectively. Errors may occur during transmission. When transmitting a signal encoding 0, the decoding may not necessarily be 0. But it is decoded as 0 or 1 with probabilities of 0.8 and 0.2 respectively. Similarly, when transmitting signal encoded as 1, it is decoded as 1 or 0 with probabilities of 0.9 and 0.1, respectively. If the decoding is 0, calculate the probability of decoding without errors.

Analysis: Assuming event A_0 represents “encoding as 0”, A_1 represents “encoding as 1”, and B represents “decoding as 0”. Then A_0, A_1 is a partition of the sample space. According to the question, it can be seen that

$$P(A_0)=0.6,\ P(A_1)=0.4,\ P(B|A_0)=0.8,\ P(B|A_1)=0.1.$$

From the full probability formula, it is obtained that

$$P(B) = P(A_0)P(B|A_0) + P(A_1)P(B|A_1)$$

$$= 0.6 \times 0.8 + 0.4 \times 0.1 = 0.52$$

If the decoding is 0, the probability of decoding without errors is the probability $P(A_0|B)$ of encoding to 0. According to Bayes formula, it can be concluded that

$$P(A_0|B) = \frac{P(A_0)P(B|A_0)}{P(B)} = \frac{0.6 \times 0.8}{0.52} \approx 0.923$$

Using the same method, if the decoding is 1, we can calculate the probability of decoding without errors. The probability is 0.75.

In the process of encoding, transmission, and decoding, the Bayes formula is used to solve the probability prediction problem of decoding errors. When applying Bayes formula to probability problems in signal processing, we can guide students to analyze layer by layer in specific problems. In the process of finding solutions to problems, we can discover the general problem-solving laws of Bayes formula. These promote students' understanding and application of Bayes formula. And these can cultivate the ability to conduct in-depth research and exploration.

3.2 Application in Software Fault Diagnosis

Case 2: When diagnosing a fault in computer software, it is assumed that there are three causes of the fault. According to previous statistical data [6], the probabilities of these three causes are 0.32, 0.21, and 0.47, respectively. To determine the cause of the fault, experiments need to be conducted. Each test result is represented by “normal” and “abnormal”. The probabilities of the test results being “normal” under the three fault causes are 0.75, 0.25, and 0.88, respectively. One day, this malfunction occurred and the engineer conducted a test. The test results showed “normal”. Please analyze which cause of the malfunction is most likely to occur. If the test results show “abnormal”, please analyze which cause of the fault is most likely to occur.

Analysis: Assuming that event A_i represents “three types of fault causes F_i, $i = 1, 2, 3$”, event B represents “normal test results”, and event $\overline{B}$ represents “abnormal test results”. A_1, A_2, A_3 forms a partition of the sample space. We learned from the question: $P(A_1) = 0.32$, $P(A_2) = 0.21$, $P(A_3) = 0.47$, $P(B|A_1) = 0.75$, $P(B|A_2) = 0.25$, $P(B|A_3) = 0.88$. It is easy to know that the probabilities of “abnormal” test results under the three fault causes are $P(\overline{B}|A_1) = 0.25$, $P(\overline{B}|A_2) = 0.75$, $P(\overline{B}|A_3) = 0.12$, respectively. From the full probability formula, it is obtained that

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)$$

$$= 0.32 \times 0.75 + 0.21 \times 0.25 + 0.47 \times 0.88 = 0.7061$$

Therefore, according to the Bayes formula, under the condition of “normal” test results, the probabilities of three fault causes occurring are:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{0.32 \times 0.75}{0.7061} = 0.34$$

$$P(A_2|B)=\frac{P(A_2)P(B|A_2)}{P(B)}=\frac{0.21\times 0.25}{0.7061}=0.074$$

$$P(A_3|B)=\frac{P(A_3)P(B|A_3)}{P(B)}=\frac{0.47\times 0.88}{0.7061}=0.586$$

By using the same method, when the test results show “abnormal”, the probability of three types of fault causes occurring can be determined as:

$$P(A_1|\overline{B})=0.272\,,\ P(A_2|\overline{B})=0.536\,,\ P(A_3|\overline{B})=0.192$$

From the above calculation results, it can be seen that when the test result is “normal”, the probability of the fault being caused by F_3 is slightly higher than that caused by F_1. We make the diagnosis selection more difficult. But the fault F_2 can be eliminated. When the test result is “abnormal”, the probability of the fault being caused by F_2 is slightly higher than that caused by F_1. The diagnostic selection faced may be more difficult. But the fault F_3 can be eliminated. In this case, further experiments must be carried out. We can use Bayes formulas to correct and adjust probabilities based on new information. Through continuous iteration, we make the optimal diagnosis of the cause of the fault. Therefore, Bayes formulas play an important role in software fault diagnosis.

3.3 Application in Medical Testing

With the rapid development of computer technology, Bayes formulas are widely used for probability prediction of related problems such as disease screening, medical testing, new drug research and development, and medical experiments [7,8].

Case 3: According to the surveys, the prevalence rate of a certain disease is 0.2%, which can be screened through blood sampling tests. According to previous data, the accuracy of the diagnosis of this disease is 97%, which means that the probability of a positive test result for people with this disease is 97%. And the probability of a positive test result for people without illness is 3%. If a person's test results are positive, what is their true probability of getting sick?

Analysis: Supposing that event B represents “positive test results”, event A represents “the person being tested is sick”, and $\overline{A}$ represents “the person being tested is not sick”, then A and $\overline{A}$ form a partition of the sample space. According to the question, it can be seen that

$$P(A)=0.002\,,\ P(\overline{A})=0.998\,,\ P(B|A)=0.97\,,\ P(B|\overline{A})=1-0.97=0.03$$

According to Bayes formula, it can be concluded that

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

$$=\frac{0.002\times 0.97}{0.002\times 0.97+0.998\times 0.03}=0.061$$

From the above calculation results, it can be inferred that the probability of a person with a positive test result actually suffering from this disease is approximately 0.061. This result seems to be inconsistent with reality. Because the data shows that the accuracy of this test seems to be high. If a person tests positive, their likelihood of getting sick should be high. However, the calculated result is only 6.1%, which means that about 93.9% of the people who test positive are not sick. This is a result

that goes against intuition. The test data cannot accurately reflect the actual situation. This indicates that the test accuracy is not sufficient. In order to reduce the error rate, we adopt a review method. We will conduct another double check on those who test positive. At this time, the incidence rate of the disease is adjusted to $P(A) = 0.061$. Using the corrected incidence rate and the Bayes formula again, the following is obtained:

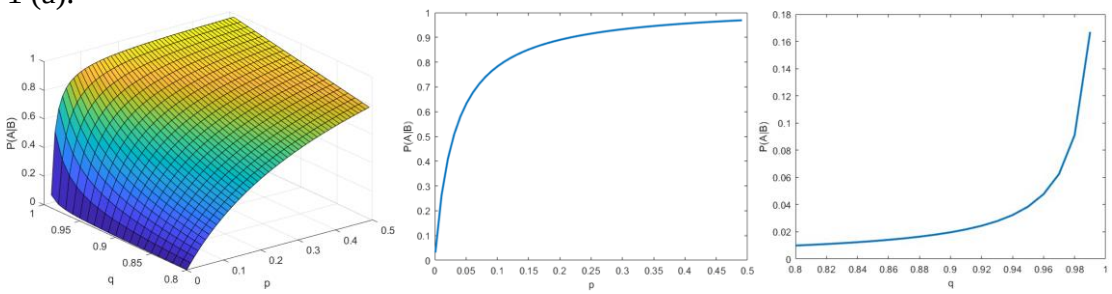
$$P(A|B) = \frac{0.061 \times 0.97}{0.061 \times 0.97 + 0.939 \times 0.03} = 0.677$$

After two consecutive calculations using the Bayes formula, the probability of a person with a positive test result actually getting sick is approximately 0.677. The accuracy of detection has significantly improved.

Extended thinking: Assuming the prevalence rate a of the disease is $P(A) = p$. The accuracy of disease diagnosis is $P(B \mid A) = q$. According to Bayes formula, a posterior probability can be concluded.

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})} = \frac{pq}{pq + (1-p)(1-q)}$$

Using MATLAB software, three-dimension curved surface can be drawn, which could show the relationship between posterior probability, prevalence rate and diagnostic accuracy. It is shown in Figure 1 (a).



(a) surface plot (b) screenshot at $p = 0.002$ (c) screenshot at $q = 0.97$

Figure 1: Relationship between posterior probability, prevalence rate, and diagnostic accuracy

If the prevalence rate a of the disease is $p = 0.002$, the cross-section line obtained by using $p = 0.002$ to cut surface is shown in Figure 1 (b). It can be seen that the posterior probability $P(A|B)$ is directly proportional to the diagnostic accuracy q. For diseases with lower prevalence rates, the posterior probability value is lower. And a higher diagnostic accuracy is necessary to have a higher posterior probability. If the diagnostic accuracy is $q = 0.97$, the cross-section line obtained by using $q = 0.97$ to cut surface is shown in Figure 1 (c). It can be seen that the posterior probability $P(A|B)$ is directly proportional to the prevalence rate p. Even if the diagnostic accuracy is high at 0.97, the posterior probability is very small for diseases with lower prevalence rates. When the prevalence rate increases from 0.001 to 0.1, the posterior probability increases rapidly. Therefore, those who test positive should undergo a reexamination to improve the posterior probability. Recheck provides a simple and effective method to improve the diagnostic accuracy for diseases with low prevalence rate. In this process, Bayesian formula constantly modifies the prior probability to make the diagnostic accuracy accurate and effective.

4. Conclusion

Bayes formula is the foundation and important application tool of probability statistics theory, and has important applications in fields such as Bayesian regression analysis, naive Bayesian, Bayesian risk decision-making, neural networks, machine learning, etc. This article analyzes and explores the application of Bayes formula in signal processing, software fault diagnosis, and medical detection. When using Bayes formula to solve problems, the prior probability is continuously modified by adding new information. Finally, the posterior probability is obtained. Fuzzy inference is made using a small amount of information. And the previous inference is continuously modified based on the inference results until the optimal probability is obtained. Bayes formula has extensive application value in daily life. Especially with the development of computer technology, its application is becoming increasingly widespread. The key algorithm core of many artificial intelligence phenomena such as spam recognition, smartphone automatic translation, and speech recognition is the self-learning function of Bayes formula. With the arrival of the big data era, Bayes formulas will have an increasingly widespread development.

Acknowledgments

This work was supported by Qingdao University of Science and Technology Teaching Reform Research Project. The author would like to thanks the members of the project for their help.

References

[1] Xuan Zuxing, Zhang Lixin, Yuan Anfeng. Course Ideological and Political Teaching Design in Bayes formula [J]. University Mathematics, 2022, 38 (2): 104-111

[2] Chen Zhongming. Discussion on Heuristic Teaching Design of Total Probability Formula and Bayes formula [J]. Education and Teaching Forum, 2019, 25: 202-203

[3] Murphy K. Inference and learning in hybrid Bayesian networks [J].Technical Report, 1998, 7(4):37-46

[4] Zhang Jufang. Probability Theory and Mathematical Statistics [M]. 2nd Edition. Beijing: Ocean Publishing House, 2016

[5] Wang Lei. On the Application of Bayes Method in Computer Intelligence [J]. Small and Medium sized Enterprise Management and Technology: Second Issue, 2011, (05): 281-282

[6] Zhou Guiru. Application of Bayesian statistical method in software fault diagnosis [J]. Journal of Anhui Electronic Information Vocational and Technical College, 2010, 9 (6): 17-18

[7] Feng Guangqing, Han Chunyang. Analysis of the Application of Full Probability Formula and Bayes Formula [J]. Journal of Jiaozuo Normal College, 2022, 38 (3): 73-76

[8] Fang Hongyan, Wang Rui, Yang Wenzhi, Hu Futao. Deep Mining of Bayesian Formula Teaching [J]. Journal of Qufu Normal University, 2018, 44(4):1-4