The Application of Bayesian Theorem

Yuexin Zhang *
The School of Economics, Xiamen University, Xiamen, Fujian, China, 361005
* Corresponding Author Email: 15220192202763@stu.xmu.edu.cn

Abstract. Bayesian theorem is an outstanding theorem in probability theory. With the development of technology and the progress of society, the advantages of Bayesian theorem gradually show, which let us better utilize existing resources to make more accurate judgments. Further understanding of the application of Bayesian theorem in different fields will help us to clarify the future development trend and deficiency of this theorem. Numerous studies have explored this theorem's advantages and developments in several heated areas. In this study, the link between Bayesian theorem and three main fields, which are finance, computer science, and medicine, is carefully studied. In finance, the distribution, $h(\theta|x) = \frac{\rho}{\sqrt{2\pi}} e^{-\frac{1}{2}\rho\left[\theta-\frac{1}{\rho}\left(\frac{\mu}{\tau^2}+\frac{x}{\sigma^2}\right)\right]^2}$, aiding investors to make decisions in terms of both prior information and sample information are studied, as its mean $\frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}x$ balances two kinds of information. In computer science, an algorithm based on the Naive Bayes model is used to effectively filter spam, and TAN. Besides, one of naive Bayes model's improvements is also discussed and compared. In medicine, the theorem upgrades the accuracy of diagnosis. In addition, a practical example is adopted to prove its function in breast lump disease. With abundant investigations and extensive applications, Bayesian theorem will be deeply rooted in human lives in various fields and greatly facilitates further research in these fields soon.

Keywords: Bayesian theorem, Finance, Computer science, Medicine.

1. Introduction

Bayesian theorem, discovered by Bayes, has attracted the attention of the mathematical circle. This outstanding achievement, which has an important influence on both modern probability theory and mathematical statistics, still gains high popularity after more than 200 years and has been widely used and studied in many fields, such as [1],[2],[3]. In 1774, Laplace published his paper, giving a comprehensive and popular exposition of Bayesian ideas [4]. Then, in 1781 and 1786, Laplace published two more papers on Bayesian theory, improving and expanding on the ideas in 1774. Through these articles, people can have a deeper and more direct understanding of Bayes' thoughts. Next, Finetti proposed and developed the subjective probability theory [5]. Besides, the modern utility theory is established by Neumann and Morgenstern. In particular, the most perfect theoretical form of Bayesian theory was proposed by Savage, which is called classical Bayesian theory. This theorem mainly describes how to calculate the probability of event A happening given that event B occurs. In mathematical terms, that means $P(A|B) = \frac{P(A \cap B)}{P(B)}$ [6]. Prior probability, the probability that people get from a subjective judgment of events, is represented by $P(A)$. Conditional probability is a modified probability based on the objective investigation, which is represented by $P(A|B)$. The discovery of Bayesian theorem greatly promotes the development of probabilistic statistics. Compared with traditional probability estimation that cannot be modified, Bayesian theorem can be constantly modified, which substantially improves the practicability of probability statistics [7]. Compared with traditional classical estimation, Bayesian theorem takes subjectivity as the starting point and shows great advantages. However, Bayesian theorem also has certain limitations, as it is based on subjective judgment and has strong subjectivity [8]. Due to each person's different interpretation of prior information, the prior probabilities obtained are different, and the posterior probabilities obtained are also varied, which lacks scientific objectivity [7].

Based on the above certain understanding of Bayesian theorem, application and promotion of Bayesian theorem are introduced and explained in the following three aspects, which are finance, computer science, and medicine, respectively. In finance, Bayesian theorem helps decision-makers to predict important evidence in the decision-making process based on Bayesian theorem, combining the information by existing data and subsequent sampling results. Therefore, investors will improve the effectiveness of decision-making and reduce the blindness of investment. When it comes to computer science, algorithms based on Bayesian theorem can correctly identify and filter spam, which greatly enhances the efficiency of the use of cyberspace and saves valuable time for businessmen. At last, Bayesian theorem also shows satisfying performance in medicine. Through it, doctors can combine expert opinions with previous cases to make a more reasonable and effective diagnosis for patients.

The wide application of Bayesian theorem benefits from its characteristics to develop rapidly, which brings countless novelty and convenience to our daily life. With the development of society, Bayesian theorem will continue to develop, resulting in better use in more emerging fields.

2. Three Aspects of Application

2.1. Finance: Project Investment

In investment decision-making, we often encounter the problem of making decisions in an uncertain state. Generally, people make rational decisions based on some reliable evidence, such as the distribution of sales of a product to further decide whether to sell a product. Unfortunately, those valid distributions of these clues are hard to acquire. As a result, in the discussion of this module, decision-makers, taking Bayesian theorem as the foundation, will combine the existing materials with the information provided by the follow-up sampling results to forecast the important evidence in the decision-making process. This is for the sake of improving the effectiveness of decision-making and reducing the blindness of investment.

In the discussion below, we focus on a continuous random variable, which is more commonly used in economics.

Theorem 2.1.1 [1] Suppose θ is crucial evidence in decision making and X is the random variable that represents the outcome of the sampling. Hence, we have prior function h(θ), $f(x|\theta)$ and prediction function $k(x)$ to calculate the posterior function $h(\theta|x)$.

$$h(\theta|x) = \frac{f\ (x|\theta)\ h\ (\theta)}{k\ (x)} \quad (1)$$

$$k(x) = \int_{\theta\epsilon\Omega} f\ (x\ |\theta)\ h\ (\theta)\ d\theta \quad (2)$$

In general, the use of Bayesian methods to calculate a posterior density is often encountered with some computational difficulties. To solve it, we decide to utilize a special class of prior probability distributions with no computational difficulties, called conjugate distributions. In the following, we will define this special distribution.

Theorem 2.1.2 [1] Suppose P represents the class of likelihood function $f(x|\theta)$ and Q represents the class of prior distribution of $h(\theta)$ which is called the conjugate of P. For all $f(x|\theta) \in P$ and $h(\theta) \in Q$, $h(\theta|x)$ is also in class Q.

In observation, many stochastic problems follow the normal distribution, so as economic activities. Therefore, to facilitate the application of Bayesian theorem to the normal distribution, several theorems about normal distribution will be given below.

Theorem 2.1.3 [1] $X \sim N(\mu, \sigma^2)$, suppose σ^2 is known. Then we have $X|\theta \sim N(\theta, \sigma^2)$.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\theta)^2}{2\sigma^2}} \quad (4)$$

Theorem 2.1.4 [1] if $\theta \sim N(\mu, \tau^2)$, then we have

$$h(\theta) = \frac{1}{\sqrt{2\pi}\tau} e^{\frac{(x-\theta)^2}{2\tau^2}} \tag{5}$$

In order to get the expression of $k(x)$, we utilize function (1) and function (2), combining the theorem 2.1.1. After calculation, the result is as follow.

$$f\ (x|\theta)\ h(\theta) = \frac{1}{2\pi\sigma\tau} e^{-\frac{1}{2}\rho\left[\theta-\frac{1}{\rho}\left(\frac{\mu}{\tau^2}+\frac{x}{\sigma^2}\right)\right]^2} e^{-\frac{(\mu-x)^2}{2(\sigma^2+\tau^2)}},$$

$$\text{where } \rho = \frac{1}{\tau^2} + \frac{1}{\sigma^2} = \frac{\tau^2+\sigma^2}{\tau^2\sigma^2}. \tag{6}$$

From the above equations (4), (5) and (6), the marginal density function of $x$can be calculated which is $k(x)$.

$$k(x) = \frac{1}{\sqrt{2\pi}\rho\sigma\tau} e^{-\frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}} \tag{7}$$

As a result, we can utilize function (6) and function (7) to obtain the expression of the posterior function $h(\theta|x)$, which is especially vital clue in decision-making process.

$$h(\theta|x) = \frac{\rho}{\sqrt{2\pi}} e^{-\frac{1}{2}\rho\left[\theta-\frac{1}{\rho}\left(\frac{\mu}{\tau^2}+\frac{x}{\sigma^2}\right)\right]^2} \tag{8}$$

Based on the above theorems, we know that the distribution of posterior function $h(\theta|x)$ follows the normal distribution.

$$h(\theta|x) \sim N(\mu(x), \frac{1}{\rho}), \tag{9}$$

Where $\mu(x) = \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right) = \frac{\sigma^2}{\sigma^2+\tau^2}\ \mu + \frac{\tau^2}{\sigma^2+\tau^2} x.$

Through the results presented, posterior information is the combination of prior information and sample information. To be more specific, $\mu(x)$ is exactly the weighted average of the mean μ of the prior distribution and the observed value x, whose coefficients are $\frac{\sigma^2}{\sigma^2+\tau^2}$ and $\frac{\tau^2}{\sigma^2+\tau^2}$, respectively. Hence, we can conclude that the application in the discussion makes good use of prior information and sample information, aiding the decision-makers to better master the detail of different situations.

If possible, the scientists in this field of work will spare no effort to test its efficiency by applying it in more practical cases and investigating the expression function of discrete random variables to broaden its usable range.

2.2. Computer Science: Spam Filtering

With the development of the Internet, E-mail as a fast and economical way of communication has been popularized. Mail is one of the most popular applications on the Internet [9]. Therefore, when there is a large number of unrecognized spam filling our mailboxes, the network broadband and server storage space will be misused, causing great trouble to the high-speed operation of the network. To decline this waste, many scientists have made detailed studies in this area. Returning to the mail itself, the problem of mail classification can be regarded as a binary classification problem, which divides the mail into legitimate mail and spam, and various text classification methods can be utilized to filter spam. Therefore, rule-based Ripper algorithm [10], decision tree C4.5 algorithm [11], Boosting method [12], Rough Set method [13], kNN algorithm [14], and Bayesian classification method all have been applied in learning this problem. Among them, Bayes' classification method has been widely studied and discussed due to its unique and outstanding performance [2].

In the following analysis, we decide to make assumptions and clearly explain this topic base on Navie Bayse. Hoping to better understand this problem, we must introduce the related concept of the theory first.

Definition 2.2.1 [2] Bayesian network is a binary $\mathrm{B} = < \mathrm{G}, \theta >$. Thereinto, G is a directed acyclic graph, in which nodes represent random a variableX_i, and directed edges between nodes represent conditional dependencies between random variables. θ is the parameter vector of nodes, and each component is a conditional probability table, which defines the local probability distribution of corresponding nodes.

The structure of Bayesian network indicates that a node X_i is independent of the non-descendant nodes in the network under the condition of the given parent node. A Bayesian classifier is a Bayesian network for classification tasks, which contains a node C representing categorical variable and a node X_i representing characteristic variables. Taking X (the realization of the characteristic variable is $(x, x_2, \dots x_n)$ for example, Bayesian network allows us to compute the probability of each possible categories c_k, $P(C = c_k|X = x)$, and the task of classification is to find out which c_k maximizes the $P(C = c_k|X = x)$.

Theorem 2.2.1 [2]

$$P(C = c_k|X = x) = \frac{P\ (X = x|C = c_k)P\ (C = c_k\ )}{P\ (X = x)} \tag{10}$$

In this formula, $P\ (X\ =\ x)$ is the same for every c_k, so we don't have to worry about it. The prior probability $P\ (C\ = c_k\ )$ can be expressed as the proportion of the total number of vectors belonging to category c_k to the total number of vectors in the sample space. In order to calculate $P(C = c_k|X = x)$, we adopt the Navie Bayse which assumes that each characteristic variable X_i is independent under a given category variable C. This classification takes the initial form of variable independence hypothesis, which is also the most restrictive form [15].

In the Bayesian network shown in Figure 1, there are only arcs from category variables to feature variables, while there are no arcs between feature variables X_i.

As a result, $P\ (X = x\ |C = c_k) = \prod_i P\ (X_i = x_i\ |C = c_k)$.

In the calculation process, the maximum likelihood estimation of x_i which is taken from the sample is treated as $P(X_i = x_i)$ under given c_k. Therefore, $P(X_i = x_i|C = c_k) = \frac{n_{xi\_in\_ck}}{n_{c_k}}$ [2].

n_{c_k} is the sample number in the category c_k, and $n_{xi\_in\_ck}$ is the number of samples whose characteristic variable X_i is equal to x_i in category c_k [2].

The efficiency of naive Bayes is relatively high. Let m be the number of classification variables and N be the number of training samples, the time of learning the naive Bayes model is $O(mN)$ and the time of classification at runtime is $O(m)$ [2]. Besides, through this method, many classification applications, including spam filtering, have achieved surprising results. Extension of this assumption to further improve performance has been widely studied, resulting in several variants of the naive Bayes. TAN (Tree Augmented Naive Bayes) is one of the variants, which loosens the independence hypothesis in naive Bayes and extends the structure of naive Baye, as it allows variables other than category variables to have a tree structure in Figure 2.

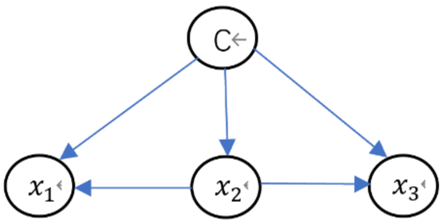

Figure 1 Network of naive Bayes

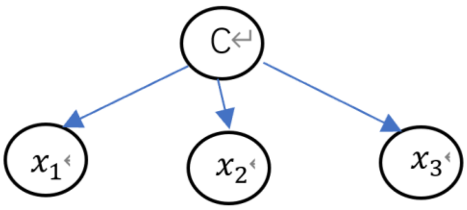

Figure 2 Network of tree augmented naive Bayes

2.3. Medicine: Disease Diagnosis

The application of various laboratory tests and examinations of patients using medical instruments is to diagnose disease or test for disease screening [16]. Bayesian theorem, based on its properties, can be extended to multiple studies, and predict the probability of having a disease or not for patients with certain identifying characteristics. Such applications not only deepen the research and understanding of the diseases, but also help clinicians improve their diagnostic skills.

Bayesian theorem used in medical diagnosis, can reflect the probability in the diagnosis of masculine and feminine according to the collected diagnostic information (i.e., the probabilities of events). The formula is as shown.

Theorem 2.3.1 [18]

$$P(D^+|T^+) = \frac{P(D^+)\,P(T^+|D^+)}{P(D^+)\,P(T^+|D^+)+P(D^-)\,P(T^+|D^-)} \tag{11}$$

$P(D^+|T^+)$ is the probability of positive (illness) in the diagnostic test, which is called the positive prediction value. The prevalence can be interpreted as $P(D^+)$, which is the frequency of illness in the subject population. $P(T^+|D^+)$ is the probability of positive test results of patients, which is called sensitivity. $P(T^-|D^-)$ is the probability of negative test results for patients without disease, namely the specific degrees.

Such applications have yielded good feedback in practical results, such as the specificity of antigen and antibody detection often reaching more than 99%. However, we should not ignore that we can only use this model to make some inferences about the results of disease diagnosis only when we have mass reliable data. That is definitely a disappointing outcome to research rare diseases, which may even greatly influent the patient's daily life, due to the lack of study samples.

Here is a clinical example to show how Bayesian theorem can be used in a real case [3]:

In order to solve several vital problems in breast lump disease, a retrospective analysis was performed in a hospital to collect the diagnosis of breast lump disease in the past 12 months. After analysis, the hospital finally diagnosed 240 cases of fibro tumor, 16 cases of breast disease, and 50 cases of breast cancer. The basic age, disease status, and lump surface area of these patients are shown in Table 1.

Table 1. Clinical features of 450 cases of breast lump

		Fibro tumor (D_1)	Breast disease (D_2)	Breast cancer (D_3)
Age (T_1)	<40	192 (80.0%)	133(83.1%)	7(14.0%)
	>40	48(20.0%)	27(16.9%)	43(86.0%)
Surface (T_2)	Regular	117(48.8%)	74(46.3%)	4(8.0%)
	Irregular	123(51.3%)	86(53.8%)	46(92%)

Table 2. Bayesian calculation for patients with a single symptom of irregular lump surface

	Fibro tumor (D_1)	Breast disease (D_2)	Breast cancer (D_3)
Prior probability $P(D_i)$	0.5333	0.3556	0.1111
Conditional probability $P(T\|D_i)$	0.5125	0.5375	0.9200
Posterior probability $P(D_i\|T)$	0.4824	0.3373	0.1804

As shown in Table 2, using a single measure of mass surface irregularities as a criterion, the patient had a 48% chance of fibroadenoma, a 34% chance of breast disease, and an 18% chance of breast cancer, according to the Bayesian formula.

Generally, it is very difficult to use multiple indicators as conditions for further probability calculation, so we assume that the multiple indicators are independent. For this example, this is a reasonable assumption. Hence, we have the formula of m indicators (I_i):

$$P(D_i|I_1 I_2 \dots I_m) = \frac{P(D_i)P(I_1|\ D_i)P(I_2|\ D_i)\dots P(I_m|\ D_i)}{\sum_I P(D_i)P(I_1|\ D_i)P(I_2|\ D_i)\dots P(I_m|\ D_i)} \tag{12}$$

When the two indicators of irregular surface of lump and age >40 years are used simultaneously, the posterior probability is: Suppose$A = 0.5333 \times 0.2 \times 0.513, B = 0.3556 \times 0.169 \times 0.538, C = 0.1111 \times 0.86 \times 0.92$.

$$P(D_1|A_2 S_2) = \frac{A}{A+B+C} = 0.3127 \tag{13}$$

$$P(D_2|A_2 S_2) = \frac{B}{A+B+C} = 0.1845 \tag{14}$$

$$P(D_3|A_2 S_2) = \frac{C}{A+B+C} = 0.5028 \tag{15}$$

The combined application of multiple indicators is an important link in clinical diagnosis. At present, there is more and more diagnostic information for different diseases and increasing epidemiological investigations. In the future, the model can be used to calculate the probability of various diseases based on the diagnosis information, which greatly improves the effectiveness of diagnosis.

3. Conclusion

The content and development of Bayesian theorem are briefly introduced to provide a clear picture of its strengths and development prospect at the beginning. In the following, the applications of Bayesian theorem in three main fields will be discussed, which are finance, computer science, and medicine respectively. In the financial field, investors often encounter the problem of making decisions under uncertainty. Given this risk, the method to improve the correctness of the decision is given by us, that is, to conduct sampling experiments at first. According to the information provided by the sampling results, decision-makers can deepen their understanding of various natural states that affect decision-making before making decisions, which greatly reduces the blindness and risk of investment. When it comes to computer science, the principle of Bayesian classifier is clearly explained. In addition, in order to effectively identify spam, the Naive Bayes classification algorithm and one of its improvements are introduced and compared. In medicine, doctors can utilize Bayesian theorem to gather information related to the diagnosis that has occurred, based on their existing experience. As a result, the efficiency of clinical medicine and the correct rate of disease diagnosis are improved. In addition, the advantages of Bayesian theorem in this field are verified by the simulation of real cases.

References

[1] N. Zhang, J. Li, Risk Decision of Project Investment Based on Bayes Method, Optimization of Capital Construction, vol. 27, 2006, pp. 54-56. DOI: CNKI: SUN: JJYH.0.2006-03-017

[2] M. Zhang, Y. Li, W. Li, Survey of Application of Bayesian Classifying Method to Spam Filtering, Application Research of Computers, vol. 22, 2005, pp. 14-19. DOI: 10.3969/j.issn.1001-3695. 2005. 08.004

[3] B. Tang, The Application of Bayesian Theorem in Disease Diagnosis, Science and Technology & Innovation, 2019, pp. 152-153. DOI: 10.15913/j.cnki.kjycx.2019.01.152

[4] Z. Yu, Who Established Bayesian School? Bayes or Laplace? Journal of Statistics and Information, 2008, pp. 85-90. DOI: 10.3969/j.issn.1007-3116.2008.01.017

[5] X. Ren, Z. Li, A General Survey of the Development and Research Trend of Bayesian Decision Theory, Studies in Philosophy of Science and Technology, vol. 30, 2013, pp. 1-7. DOI: CNKI: SUN: KXBZ.0.2013-02-002

[6] X. Yao, Analysis of Bayes theorem and its application, Education Modernization, vol.4, 2017, pp. 254-256. DOI: 10.16541/j.cnki.2095-8420.2017.06.093

[7] C. Liao; A brief analysis of Bayes theorem and its application, Motherland, 2019, pp. 63-64. DOI: CNKI: SUN: ZUGU.0.2019-12-026

[8] J. Yang, D. Chen, X. Cheng, Several applications of Bayes' formula, College Mathematics, vol.27, 2011, pp. 166-169. DOI: 10.3969/j.issn.1672-1454.2011.02.037

[9] W. Luo, X. Gao, X. Ou, Y. Liu, D. Li, The Data Security and Secure Content of E-mail, Journal of Computer Applications, 2002, pp. 22-24. DOI: CNKI: SUN: JSJY.0.2002-03-006

[10] W. Cohen, Fast Effective Rule Induction, in Machine Learning, Proceedings of the 12th International Conference, Lake Taho, California, Mongan Kanfmann, 1995, pp. 115-123. DOI: 10.1016/b978-1-55860-377-6.50023-2

[11] X. Carreras, L. Marquez, Boosting Trees for Anti-Spam E-mail Filtering, Proceedings of Euro Conference Recent Advances in NLP (RANLP-2001), 2001, pp. 58-64. DOI: 10.1016/0375-6505(85)90011-2

[12] I. Androutsopoulos, G. Paliouras, E. Michelakis, Learning to Filter Unsolicited Commercial E-mail, Technical Report 2004/2, NC SR Demokritos, vol. 46, 2004, pp. 153-154. DOI: 10.1109/MAP.2004.1388867

[13] Y. Liu, X. Du, X. Huang, Z. Hou, C. Guo, E. Zhou, H. Luo, The Design of Spam Intelligent Filtering System, Computer Technology and Development, 2003, pp. 1-3. DOI: CNKI: SUN: WJFZ.0.2003-04-001

[14] H. Drucker, D. Wu, V. N. Vapnik, Support Vector Machines for Spam Categorization, IEEE Transactions on Neural Networks, vol. 10, 1999, pp. 1048-1054. DOI: 10.1109/72.788645

[15] P. Langley, W. Iba, K. Thompson. An Analysis of Bayesian Classifiers, Proc. of the 10th National Con.fon Artificial Intelligence, Menlo Park, AAAI Press, 1992, pp. 223-228.

[16] S. Gao, Application of Bayesian Theorem in Disease Diagnosis, Mathematics Learning and Research, 2008, pp. 116. DOI: CNKI: SUN: SXYG.0.2008-12-103