Proceedings of ICMMGH 2026 Symposium: Biomedical Imaging and AI Applications in Neurorehabilitation
DOI: 10.54254/2753-8818/2026.AU29753

Review of the Practical Application and Future Prospects of Bayesian Theory in Modern Medical Diagnosis

Weilin Chen

Quanzhou No.5 Middle School, Quanzhou, China
510044199@qq.com

Abstract. Bayes' theorem, originally posited by Thomas Bayes, centers on a fundamental principle: the iterative refinement of subjective probabilities regarding a specific hypothesis —such as the presence of a disease—through the incorporation of new empirical data, like diagnostic test outcomes. Laplace's promotion garnered significant attention. David Eddy's mid-20th-century work greatly popularized Bayesian methods in medicine, revealing discrepancies between doctors' intuitive probability estimates and objective Bayesian calculations. Physicians often overestimate disease probability given a positive test, neglecting base prevalence rates. This has emerged as a primary catalyst propelling the contemporary advancements in Bayesian theory. Since the 20th century, Bayesian methodologies have seen extensive application across diverse domains, including statistical analysis, machine learning, and artificial intelligence. As a mathematical method to infer posterior probabilities through prior conditional probabilities, this theory has played an essential role in various fields of life, such as promoting medical development and predicting financial markets. This paper introduces Bayesian theory, summarizes its practical applications in medical diagnosis with the help of Bayesian methods and probability theory, and looks forward to the promising development prospects of Bayesian theory in the future.

Keywords: Bayesian, Medical Diagnosis, Artificial Intelligence

1. Introduction

Up to now, China has developed rapidly in various fields, and people's living standards have steadily improved. Contemporary medicine is indispensable to daily existence, assuring well-being and longevity. On March 11, 2020, the World Health Organization declared the novel infectious disease a potential global pandemic [1, 2]. The emergence and spread of the COVID-19 virus have brought troubles to the international social economy and people's lives, and it is also a significant test for the modern medical field. Using statistical knowledge to solve the hidden dangers caused by the virus and reduce the number of infections is the primary task in the face of the virus crisis. Drawing upon Bayesian statistical and probabilistic frameworks, this paper scrutinizes and synthesizes the application methodologies at the intersection of contemporary medical practice and statistical analysis, with a particular focus on their deployment during epidemic events. For example, it uses Bayesian theory to predict the effectiveness of COVID-19 vaccines and infer the accuracy of positive test paper detection, to summarize the critical application of this theory in the medical field

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.AU29753

in the past few years, and make preliminary simple predictions and provide relatively practical suggestions for possible challenges in the medical field in the future, such as influenza viruses or new viruses.

2. Bayesian theory

Bayesian theory, or Bayesian inference, concerns the conditional and marginal probabilities of random variables, providing a formula to compute posterior probabilities from conditional probabilities [3]. Bayesian probability leverages prior knowledge and existing statistical data, employing probabilistic methods to estimate the likelihood of future events [4]. The basic Bayesian formula is:

$$P(A1|B) = \frac{P(A1)\cdot P(B|A1)}{P(A1)\cdot P(B|A1)+P(A2)\cdot P(B|A2)} \quad (1)$$

In the formula, letters A1, A2, and B represent three related events, and events A1 and A2 are mutually exclusive. P(A1 | B) is the posterior probability, and P(A1) is the prior probability. Compared with the prior probability, the posterior probability has an additional condition of event B, so it can be understood as the probability of event A1 occurring when event B is observed. In contrast, the prior probability is the probability of event A1 occurring when event B is not observed. To ascertain the posterior probability from the prior, suitable data sources are indispensable; within the medical domain, these encompass historical transmission and mortality rates of infectious viruses, such as SARS, or empirically-derived assumptions posited by seasoned researchers. P(B | A1) is called the likelihood, representing the degree of support of observed data B for category A1. Therefore, for multiple events A1,A2,A3...An, the general formula can be derived:

$$P(A1|B) = \frac{P(B|A1)\cdot P(A1)}{\sum_{k=1}^{n} P(B|Ak)\cdot P(Ak)} \quad (2)$$

Where A1, A2, A3...An are mutually exclusive and exhaustive events. The Bayesian formula is one of the essential formulas in probability theory and is widely used in real life [5].

3. Practical application of Bayesian theory in the modern medical field

3.1. Prediction of the effectiveness of drugs and vaccines

In recent years, the high incidence of various infectious diseases has seriously affected people's health, especially the health of children. Vaccines represent the most efficacious and cost-effective intervention for the prevention of infectious diseases [6]. However, vaccine research and development is a highly challenging and risky process, and the primary task is to minimize the number of infections and diseases while reducing vaccine side effects as much as possible. Although a qualified vaccine can control the number of infections and reduce the risk of disease, a group of vaccinated people will still be infected with the virus. Therefore, using statistical knowledge such as Bayesian theory to test the effectiveness of vaccines through clinical trials to maximisze the prevention rate of vaccines is an essential part of ensuring people's healthy lives. Taking clinical trials as an example, from the perspective of Bayesian theory, it can be divided into several steps. First, determine the prior probability. Since many new viral infectious diseases, such as COVID-19, are new viruses, the prior probability cannot be determined. Therefore, when data is lacking, the prior distribution of infection probability is uniform (non-informative prior), or it is set as a Beta

DOI: 10.54254/2753-8818/2026.AU29753

distribution based on historical data such as the SARS virus, which has some similarities with the COVID-19 virus in 2003.

Next, trial data will be collected, and this step requires sample classification based on specific data. Assume the clinical trial results are as follows:

Vaccinated group: total number of people A and infected people a.

Unvaccinated group: total number of people B and infected people b.

The next step is to calculate the posterior probability. The posterior distribution of the infection probability in the vaccinated group can be listed through Bayesian theory:

P(infection | vaccinated) ~ Beta(a+α, A−a+β), where α and β are prior parameters (if there is a non-informative prior, α=β=1 can be set).

The posterior distribution of the probability in the unvaccinated group: P(infection | unvaccinated) ~ Beta(b+α, B−b+β). The last step is to estimate the vaccine effectiveness (VE). Samples are drawn from the posterior distribution to calculate VE, VE= (1−Beta(b+α,B−b+β)Beta(a+α,A−a+β))×100%. In actual calculation, the posterior mean or median is usually used as the point estimate. Here is an example to illustrate this method:

Vaccinated group: 1000 people, 10 infected → P(infection | vaccinated)=10/1000=1%.

Unvaccinated group: 1000 people, 50 infected → P(infection | unvaccinated)=5%.

Using the Bayesian method, assuming the prior is Beta(1,1), the posterior distribution is:

Vaccinated group: Beta(10+1,990+1), mean ≈ 11/1002 ≈ 1.1%.

Unvaccinated group: Beta(50+1,950+1), mean ≈ 51/1002 ≈ 5.1%.

Then VE≈(1−5.1%1.1%)×100%≈78.4%.

This experiment only plays a role in preliminarily estimating the VE value. The accurate value should also consider factors such as age, gender, congenital diseases, and living habits. Moreover, when a new virus arrives and data is lacking, a relatively accurate prior distribution cannot be obtained, and there is a certain gap between the obtained probability and the actual precise probability. Therefore, considering the role of uncertain factors mentioned above is a deficiency of this experiment, which needs to be solved by using regression or Bayesian hierarchical models in the future. To ascertain the projected vaccine efficacy, researchers can employ a controlled variable methodology, establishing multiple experimental cohorts to modulate the constituents and categories of diverse vaccine components. By leveraging Bayesian statistical methods, the vaccine effectiveness within each cohort can be computed. Following iterative experimentation, the cohort exhibiting the apex of effectiveness is selected for subsequent investigation and refinement, with the aim of formulating highly efficacious and cost-effective vaccines for widespread dissemination, thereby furnishing a robust safeguard against the COVID-19 pandemic.

3.2. Estimation of the accuracy of viral antigen detection

In recent years, China has included antigen detection in COVID-19 screening methods and allowed people to self-test, attracting much attention. COVID-19 antigen detection has been applied earlier abroad. Antigen detection cannot replace the "gold standard" nucleic acid detection for diagnosis, but it has the advantages of convenience, low cost, and easy popularization. In light of the rapid dissemination of variant strains, early detection of the epidemic is advantageous [7]. However, the antibody detection of the new coronavirus also has errors, namely "false positives". The accuracy of antibody detection is affected by many factors, such as samples, equipment, reagents, and environment, which may lead to deviations in the final results. For example, unreasonable primer design in the kit can lead to primer dimers, or laboratory pollution can cause false positives. Therefore, before an antibody detection reagent is promoted, it is important to control the spread of

DOI: 10.54254/2753-8818/2026.AU29753

the virus by using statistical knowledge to calculate the false positive rate and actual efficiency of the reagent, and on this basis, adjust the component content to maximise the actual detection efficiency. For example, "Assume the prevalence rate of a certain disease is 1%, and a reagent is developed for detection. If a person is sick, the probability of testing positive is 99%; if not sick, the probability of testing positive is 3% (false positive). When a person tests positive, what is the probability of being sick?"

From the perspective of Bayesian theory, let C1 be the case of being sick, C2 be the case of not being sick, and D be the event of positive test results. Then the required probability can be expressed as: P(C1 | D). According to the Bayesian formula:

$$P(C1|D) = \frac{P(C1)\cdot P(D|C1)}{P(C1)\cdot P(D|C1)+P(C2)\cdot P(D|C2)} \tag{3}$$

Finally, substituting the data into the formula, the result is 25%. This helps adjust testing policies and interpret results, contributing to the timely detection and prevention of the virus.

3.3. Correct interpretation of cancer screening results

With the significant improvement of people's living standards and the apparent extension of life expectancy, chronic diseases such as cardiovascular and cerebrovascular diseases, cancer, diabetes, and Alzheimer's disease have become the leading causes of death. Of the chronic conditions linked to an aging demographic, cancer precipitates the most substantial decrement in healthy life expectancy [8]. Taking mammography as an example, a 50-year-old asymptomatic woman undergoes routine breast cancer screening, and the result is positive. From a Bayesian standpoint, her a priori risk of developing breast cancer, perhaps 0.5%, can be initially estimated using data such as her race, age, and family history to establish the prior probability. This method offers greater precision compared to relying on the overall population's prevalence rate. Then determine the likelihood. It is known that the sensitivity of mammography is, for example, 90%, and the specificity is, for example, 91%. Finally, the Bayesian formula calculates the posterior probability, which is the positive predictive value. The final calculation result shows that the positive predictive value is only about 5%. That is to say, among 100 patients initially tested positive for breast cancer, only about 5 actually have cancer. This largely avoids unnecessary anxiety and panic, and guides the next rational diagnosis steps, reflecting "value-based healthcare".

4. Future development prospects of Bayesian theory

Currently, Bayesian statistics rivals classical statistics in prominence, and Bayesian theory, serving as its theoretical foundation, undeniably holds a pivotal role in contemporary statistical analysis [3]. In recent years, big data Bayesian learning has become the focus of attention. The pressing concern is how to bolster the adaptability of Bayesian learning and accelerate its inferential capabilities, thereby enabling it to effectively address the demands of the big data era [9]. In the medical field, Bayesian theory will promote the development of personalized medicine. Hierarchical Bayesian models using patients' historical data will infer population laws and individual differences to predict drug doses accurately. Bayesian theory is also one of the main contents in artificial intelligence. Based on tumor mutational burden and prior treatment response data, the Bayesian Multi-armed Bandit algorithm rapidly identifies the optimal therapeutic strategy for patients. The "arms" of the slot machine represent different experimental cancer treatment plans (such as new drug A, new drug B, standard therapy C, combinations of different doses, etc.). "Pulling the arm" means choosing a

DOI: 10.54254/2753-8818/2026.AU29753

treatment plan for a newly enrolled patient. "Reward" represents the outcome of the treatment, usually binary (e.g., tumor shrinkage reaching the predetermined standard = success/reward 1; otherwise = failure/reward 0) or continuous (e.g., duration of progression-free survival). Following treatment administration and subsequent observation of patient outcomes, the resulting evidence informs an iterative refinement of the treatment arm's posterior probability distribution. If an arm shows good results early, its posterior distribution will quickly move towards high efficiency, increasing the probability of being selected. Conversely, the probability of selecting an arm with poor results will decrease rapidly. This technology has shortened the trial-and-error cycle for cancer immunotherapy and provided a strong guarantee for the fight against cancer. In conclusion, combining this theory with artificial intelligence has become a trend in the development of the medical field.

5. Conclusion

This paper explains Bayesian theory in detail, introduces its principles and basic formulas, and then extends the general formula for n cases. It then enumerates the empirical implementations of Bayesian inference in medicine during the recent epidemic, specifically vaccine efficacy assessment and serological testing, and elucidates the crucial function of Bayesian approaches in enhancing vaccine effectiveness and diminishing false-negative rates in serological assays within the medical domain. Finally, it puts forward the future development direction of this theory, that is, combining it with artificial intelligence to significantly improve people's medical and health levels. At the same time, Bayesian theory has the characteristics of flexibility and handling uncertainty, and can make reasonable inferences for various future medical scenarios. It is hoped that the above review of Bayesian theory can play a role in preliminarily calculating possible diseases in the future to help people better prevent those diseases, and also expect this paper to help scholars better understand Bayesian theory and its important role in medical diagnosis.

References

[1] Beijing News. (2020, March 12). WHO declares COVID-19 a global pandemic, Guterres calls for action. http: //www.bjnews.com.cn/world/2020/03/12/702593.html
[2] National Health Emergency Office. (2020, June 30). The latest situation of new coronavirus pneumonia is as of 24: 00 on June 30. http: //www.nhc.gov.cn/yjb/s7860/202007/a98e49570be24eaf88de98e6e6217fc8.shtml
[3] Cheng Qihang. (2019). Review of the development of Bayesian theory and its application in classification problems. China New Telecommunications, 21(01), 27-28.
[4] Zheng Wei, Shen Wen, & Zhang Yingpeng. (2010). Research on spam filter based on an improved Naive Bayes algorithm. Journal of Northwestern Polytechnical University, 28(04), 622-627.
[5] Yu Shihang. (2011). Teaching methods of the Bayesian formula. Journal of Science of Teachers' Colleges and Universities, 31(06), 87-89.
[6] Li Zhen. (2024). The importance of children's vaccination. Health Must-read, (14), 69.
[7] Dai Xuan. (2022). Positive = infected? Should ordinary people take COVID-19 antigen tests? Grand Garden of Science, (07), 54-57. in China.
[8] Cao Guangwen. (2025). Epidemic characteristics, prevention and control status and future coping strategies of cancer Academic Journal of 279-290. https: //doi.org/10.16781/j.CN31-2187/R.20250050. Naval Medical University, 46(03),
[9] Zhu Jun & Hu Wenbo. (2015). Review of frontier progress in Bayesian machine learning. Journal of Computer Research and Development, 52(01), 16-26