Necessity of Multiple Rounds of Nucleic Acid Detection: Application of Bayesian Decision Theory in Real Life

Yuting Liu
Zhou Enlai School of Government, Nankai University, Tianjin, China.
2113531@mail.nankai.edu.cn

Abstract. Bayesian Decision is a risk-based decision that uses the Bayesian formula to combine prior probability with conditional probability, which helps people make more mathematically logical inferences in the continuous updating of data and facts. In this paper, Bayesian Decision will be applied to analyze the necessity of multiple rounds of nucleic acid detection in middle and high risk areas up against the epidemic. Besides, data processing will be conducted in positive patients and negative patients. Then Bayesian factor is introduced to simplify the calculation to draw the corresponding conclusions, and the practical application value of Bayesian Decision is prospected.

Keywords: Bayesian Decision; Probability; COVID-19 Virus; Multiple Rounds of Nucleic Acid Detection.

1. Introduction and Basic Idea of Bayesian Decision Theory

Bayesian Decision is a risk-based decision that uses the Bayesian formula to modify the prior probability based on the posterior probability. Prior probability is the occurrence probability of various natural states formed by summarized experience or subjective judgment, while posterior probability is to supplement new information through market investigation, experiment, and expert opinion, and then weigh prior probability to achieve improvement and correction. Thus, the Bayesian formula expresses the relationship between prior probability and posterior probability, aiming to utilize updated posterior probability to make the decision more in line with objective facts.

The Bayesian theorem formulated after abstraction is as follows:

$$P(H \mid E)=\frac{P(H)P(E \mid H)}{P(E)} \tag{1}$$

Where, $P(H)$ is a prior probability, $P(H \mid E)$ is a posterior probability, and $P(E \mid \mathrm{H})$ is a likelihood ratio known as the Bayesian factor. Their meaning will be further explained in subsequent practical applications.

The purpose of Bayesian Decision is reflected in the gap between prior probability and posterior probability. The greater the gap, the more obviously the posterior probability corrects the prior probability, the smaller the error between the posterior probability and the actual decision environment, and the greater the role of the Bayesian Decision. Up against the multiple rounds of nucleic acid detection in COVID-19, this paper will use Bayesian Decision to prove that correcting the results based on the known ones can improve the detection accuracy, which manifests the necessity of local governments to conduct multiple rounds of nucleic acid detection in some areas.

2. Application of Bayesian Decision Theory in Multiple Rounds of Nucleic Acid Detection

2.1 Background of Multiple Rounds of Nucleic Acid Detection

The COVID-19 pandemic has raged all over the world since 2019. As for areas with relatively more confirmed cases, China has classified them with different risk levels and implemented relevant control policies, one of which is to conduct multiple rounds of nucleic acid detection for all people. However, some residents questioned the necessity of high-frequency nucleic acid detection. According to the official response, no matter how accurate the detection is, it fails to guarantee that all positive samples are collected, because false negatives or false positives due to contaminated samples perhaps exist, and viruses in incubation may have not been detected. A few days ago, many positive cases that were confirmed to be negative occurred after several rounds of nucleic acid detection, which proved the necessity of multiple rounds of nucleic acid detection in some areas again. We will use the Bayesian theorem to analyze and demonstrate this topic mathematically.

2.2 Necessity Analysis of Multiple Rounds of Nucleic Acid Detection for Positive Patients

A is recorded as people infected with COVID-19 and B as those not infected. Meanwhile, the ratio of the number of people infected with COVID-19 to the total number in China is approximately regarded as the probability of COVID-19 infection. According to the current big data statistics, as of December 2, 2022, $P(A)=\frac{9118328}{1411780000}\approx 0.6458\%$ and $P(B)=1-P(A)=99.3542\%$ (data source: National, Provincial, and Municipal Health Commissions. All the following estimation results are kept to four decimal places). If the event C_i means positive for the i th nucleic acid detection $(i=1,2,\ldots)$, then A and B will lead to C_i and form a complete events group. As long as the individual participates in nucleic acid detection, the occurrence of any event in A or B is a necessary condition for C. According to the foreign clinical medical literature, the false positive rate of detection is about 0.1%-0.6% with 1.9%-5% as the highest. A study in British clinical medicine shows that the false positive rate of operational swabs in nucleic acid detection is initially estimated to be between 0.8% and 4.0%[3] Assuming that nucleic acid is detected under ideal conditions, the estimated false positive rate is 0.8%, that is, $P(C_1 \mid A)=99.2\%$ and $P(C_1 \mid \mathrm{B})=0.8\%$. According to the full probability formula:

$$P(C_1)=P(A)P(C_1 \mid A)+P(B)P(C_1 \mid \mathrm{B}) \tag{2}$$

The positive probability in the first round of nucleic acid detection is $P(C_1)=0.6458\%\times 99.2\%+99.3542\%\times 0.8\%\approx 1.4355\%$. Based on formula (1), when a person is detected to be positive, the infection probability is $P(A \mid C_1)=\frac{P(A)P(C_1 \mid A)}{P(C_1)}=\frac{0.6458\%\times 99.2\%}{1.4355\%}\approx 44.6300\%$. In the medical field, $P(A \mid C_1)$ is the positive predictive value (PPV), that is, the proportion of true positive cases in all detected positive cases reflects the possibility of patients to be truly positive in detection. The lower PPV indicates that the misdiagnosis rate of individuals in the first round of detection is high. When a person is detected to be positive in the first round, he has a nearly 50% possibility of not being infected.

The second round of detection is conducted for positive patients: the infection probability in positive patients is updated as $P(C_2)\approx 44.7160\%$ by the data of the first round of detection for all people, and other data are changed accordingly (the calculation method is the same as that of the first round). In this round, when a person is detected to be positive, the infection probability is $P(A \mid C)=\frac{P(A)P(C_2 \mid A)}{P(C_2)}=\frac{44.6300\%\times 99.2\%}{44.7160\%}\approx 99.0092\%$. This shows that when the patients detected to be positive in the first round are detected in the second round, 99% of the positive cases may reflect the infection, and the accuracy is greatly improved. It is necessary to take protective measures such as centralized isolation for the positive patients in time.

2.3 Necessity Analysis of Multiple Rounds of Nucleic Acid Detection in Negative Patients

In addition to positive patients, due to the false negatives, the necessity of multiple rounds of nucleic acid detection for negative patients is also analyzed. Data including $P(A)=0.6458\%$ and $P(B)=1-P(A)=99.3542\%$ in 2.2 are used. D is recorded as negative for nucleic acid detection. By the same token, A and B will lead to D, with A and B constituting a complete events group. As long as the individual participates in nucleic acid detection, the occurrence of any event in

A or B is a necessary condition for D. According to a systematic evaluation report, the false negative rate of repeated sample detection is between 2% and 33% [4]. Assuming that nucleic acid detection is conducted under ideal conditions, the false negative rate is 2% and the accuracy rate of negative detection is 98%, that is, $P(D|B)=98\%$ and $P(D|\mathrm{A})=2\%$. Based on formula (2), the negative probability is $P(D)=P(A)P(D|A)+P(B)P(D|\mathrm{B})=0.6458\%\times 2\%+99.3542\%\times 98\%\approx 97.3800\%$ According to formula (1), when a person is detected to be negative, the probability of not being infected is $P(B|D)=\dfrac{P(B)P(D|B)}{P(D)}=\dfrac{99.3542\%\times 98\%}{97.3800\%}\approx 99.9868\%$, which shows that the accuracy of the first round of detection is high. When a person is detected to be negative in the first round, he has a very small infection possibility.

The following is the second round of nucleic acid detection for negative patients. Now the probability of no infection is updated as $P(B)=99.9868\%$ by the data of the first round of detection for all people. In this round, when a person is detected to be negative, the probability that he is not infected is $P(B|D)=\dfrac{P(B)P(D|B)}{P(D)}=\dfrac{99.9868\%\times 98\%}{97.9871\%}\approx 100\%$. In other words, patients who are negative in both rounds of detection can be approximately regarded as patients not infected.

Because the second round of detection for negative patients nationwide failed to improve the accuracy, the role of Bayesian Decision could not be well demonstrated. Therefore, we doubt whether this accurate effect is diminished due to the large selection area. Thus, we will select Gaocheng District, Shijiazhuang City, Hebei Province, which has been rated as high-risk areas (as of December 2, 2022, there are more than 50 high-risk areas), for analysis. As of December 1, 2022, the total population of the Gaocheng District is 694,108, and the cumulative number of confirmed cases has exceeded 1,000 (data source: Gaocheng District Government of Shijiazhuang City). If we approximately take 1,000, $P(A)=\dfrac{1000}{694108}\approx 0.1441\%$ and $P(B)=1-P(A)=99.8559\%$ can be generated, so the negative probability is $P(D)=P(A)P(D|A)+P(B)P(D|\mathrm{B})=0.1441\%\times 2\%+99.8559\%\times 98\%=97.8617\%$ When a person is detected to be negative, the probability of no infection is $P(B|D)=\dfrac{P(B)P(D|B)}{P(D)}=\dfrac{99.8559\%\times 98\%}{97.8617\%}\approx 99.9970\%$. After the second round of detection for negative patients, when a person is detected to be negative, the probability that he is not infected is updated to $P(B|D)=\dfrac{P(B)P(D|B)}{P(D)}=\dfrac{99.8559\%\times 98\%}{97.9971\%}\approx 100\%$. Although the accuracy rate has been improved, it is still not significant, which may be related to the low overall case concentration rate in Gaocheng District.

Because the second round of detection for negative patients in Gaocheng District, Shijiazhuang City, Hebei Province failed to greatly improve the accuracy, we continued to narrow the scope of multiple rounds of detection from districts to streets. Zhongnan Street, Liwan District, Guangzhou City, Guangdong Province, as a high-risk intensive area is selected for analysis. As of December 1, 2022, the total population of Zhongnan Street is about 8,300, and the cumulative number of confirmed cases has exceeded 50 (data source: National Municipal Health Commission). 50 is taken as an approximate, $P(A)=\dfrac{50}{8300}\approx 6.0241\%$ and $P(B)=1-P(A)=93.9759\%$ can be generated. Then the negative probability is $P(D)=P(A)P(D|A)+P(B)P(D|B)=6.0241\%\times 2\%+93.9759\%\times 98\%\approx 92.2169\%$ When a person is detected to be negative, the probability that he is not infected is $P(B|D)=\dfrac{P(B)P(D|B)}{P(D)}=\dfrac{93.9759\%\times 98\%}{92.2169\%}\approx 99.8693\%$. After two rounds of detection for negative patients, the probability of a person detected to be negative but not infected is updated to

$P(B \mid D)=\frac{P(B)P(D \mid B)}{P(D)}=\frac{99.8693\% \times 98\%}{97.8745\%} \approx 100\%$. Compared with multiple rounds of nucleic acid detection in the whole country and Gaocheng District of Shijiazhuang City, the accuracy of the second round in Zhongnan Street is promoted.

2.4 Introduction of Bayesian Factor

Bayesian Decision is used in the above practical calculation, which proves that multiple rounds of nucleic acid detection can improve the diagnosis accuracy. However, there are a lot of operations with high similarity in the process. To speed up the transformation of prior probability into posterior probability and simplify the calculation, we introduce the Bayesian factor. Bayesian factors in the practical application of nucleic acid detection will be defined.

The probability of COVID-19 infection is expressed in the form of the COVID-19 occurrence ratio α. If the number of patients is m and the number of healthy people is n, then

$$\alpha=\frac{m}{n} \tag{3}$$

and the transformation relationship between the COVID-19 prevalence rate $P(A)$ and α is as follows:

$$\alpha=\frac{P(A)}{1-P(A)} \tag{4}$$

In the medical field, the sum of the true positive rate (sensitivity) β and the false negative rate $1-\beta$ is 1, and the sum of the true negative rate (specificity) γ and false positive rate $1-\gamma$ is 1. The priori original occurrence ratio α is the number of sick people divided by healthy people. After the first nucleic acid detection, the posterior occurrence ratio $\dot{\alpha}$ becomes the number of true positive people divided by false positive people, that is,

$$\dot{\alpha}=\frac{\beta m}{(1-\gamma)n}=\alpha\frac{\beta}{1-\gamma} \tag{5}$$

$\frac{\beta}{1-\gamma}$ is defined as a Bayesian factor, that is, true positive rate divided by false positive rate, and posterior COVID-19 occurrence ratio is a priori COVID-19 occurrence ratio × Bayesian factor.

The Bayesian factor is applied to multiple rounds of nucleic acid detection for 2.2 positive patients. According to the known data, the prior COVID-19 occurrence ratio α is $\frac{13}{2000}$ and the Bayesian factor is 49.6. Thus, after one round of detection, the posterior COVID-19 occurrence ratio is $\dot{\alpha}=\frac{13}{2000}\times 49.6=\frac{403}{1250}$ with 24.3799% as the transition probability. Then, the second round of detection was carried out, the posterior COVID-19 occurrence ratio was $\ddot{\alpha}=\frac{403}{1250}\times 49.6 \approx 16$ with 94.1176% as the transition probability, so there was a high probability to confirm the positive results of detection twice. If the third round of detection is to be conducted, the posterior COVID-19 occurrence ratio is $\dddot{\alpha}=16\times 49.6 \approx 793.6$ with 99.8742% as the transition probability. In other words, three rounds of positive results can confirm that an individual has been infected, and subsequent detection does not play a significant role in verifying the identity of positive patients.

If the COVID-19 occurrence ratio after the i th detection is updated to α_i, then

$$\alpha_i=\frac{\beta}{1-\gamma}\alpha_{i-1}=\left(\frac{\beta}{1-\gamma}\right)^i \alpha \qquad (i=1,2,\ldots) \tag{6}$$

If α is known, the greater the Bayesian factor, the smaller the COVID-19 occurrence ratio can obtain ideal and accurate results. Thus, the COVID-19 infection can be accurately judged while minimizing the number of detection.

According to the formula (6), the higher the value of the Bayesian factor, the more reliable it is. Thus, it is significant to improve the true positive (numerator) and reduce the false positive (denominator) simultaneously. First of all, regarding the numerator, the probability of true positive has exceeded 99%, but if the virus concentration is too low, the possibility of patients being detected negative is greatly increased. This may be because COVID-19 has not been replicated in large quantities in vivo, which also means that the patient's chance of infecting others is very limited at present. The way to rule out this possibility is to carry out the second round of detection after two or three days and repeat this process until it passes the virus incubation period. This is why the domestic isolation regulations for the entry of overseas personnel are conducted (14 days centralized isolation +7 days home isolation; a total of 6 nucleic acid detection).

Secondly, regarding the denominator, even though the probability of false positives is low in theory, it is still impossible to avoid their occurrences when detecting many people. For example, although the false positive rate assumed in this paper is only 0.8%, detecting 1 million people will still produce about 8,000 false positive cases. Therefore, it is necessary to consider other means (such as imaging, etc.) to judge the diagnosis or adopt multiple rounds of repeated detection analyzed in this paper, especially in middle and high risked areas with high case concentration.

2.5 Results Analysis and Rationalization Suggestions

According to the analysis and reasoning of the Bayesian theorem, when the results of the first round of nucleic acid detection are positive, it is necessary to carry out two or even multiple rounds of detection for positive patients due to factors such as "false positive" and manual operation errors. The infection probability after being detected to be positive will increase exponentially, which can greatly improve the pertinence of COVID-19 pandemic prevention and treatment. It not only reduces unnecessary treatment, but also avoids economic and stigmatized psychological losses caused by ineffective isolation. This improvement is not just effective in some middle and high risked areas, but can be popularized nationwide.

However, when the results of the first round of detection are negative, the multiple rounds of nucleic acid detection in the whole country and the middle and high risked areas do not contribute significantly to the detection accuracy. From the perspective of political science and sociology, multiple rounds of nucleic acid detection are an effective guarantee for people's life safety and healthy life. Considering China's large population, a small increase in accuracy will also avoid some social "false negatives". However, from the perspective of mathematics and economics, if it is difficult to get benefits corresponding to the economic investment in multiple rounds of detection nationwide, especially for negative patients, they will inevitably become a waste of social resources and a burden of national finance.

Under the leadership of the Party, governments at all levels made precise decisions to promote multiple rounds of nucleic acid detection in some middle and high risked areas instead of the whole country, so as to avoid the potential risks of COVID-19 transmission as much as possible. Because it will greatly increase $P(A)$ and decrease $P(B)$ in the short term in the middle and high risked areas, the original detection accuracy is also to be updated, which will be higher than the improvement rate of large-scale detection, proving the absolute necessity to conduct multiple rounds of nucleic acid detection for all people in some key risk areas. This not only reflects the national emphasis on people's lives and safety, but also demonstrates the governance capacity of socialism with Chinese characteristics, which makes scientific decisions and implements precise policies to fight against the virus and manage epidemic prevention and control.

As for the application of the Bayesian Decision, some data put forward suggestions for multiple rounds of nucleic acid detection in China. Firstly, the second or even multiple rounds of nucleic acid detection for positive patients should be paid enough attention. Secondly, the number of detection for negative patients should be controlled at 1-2 times as far as possible. Thirdly, the multiple rounds of nucleic acid detection for all people should be concentrated in a small range of middle and high risked areas. Besides, the more accurate the positioning of the middle and high risked areas, the more remarkable the effect of improving the detection accuracy.

3. Practical Application of Bayesian Decision

Scientists use the Bayesian theorem to analyze the extent to which new data can verify or deny their existing models, and programmers use it to build artificial intelligence. In real life, the Bayesian theorem tells us that complicated facts should not directly determine our views but constantly update our views. Just as in the analysis of the COVID-19 pandemic, we should keep pace with the times, update the corresponding data, and adjust policies in time, instead of sticking to the rules and implementing the measures at the early stage of the outbreak. Bayesian theorem makes our judgment in real life more quantitative and systematic, which can even reshape our views on thought itself and correct our intuition in some way. It also makes us avoid blindly following the majority when facing uncertainty in life, shedding the light of mathematical rationality into reality.

Bayesian Decision helps us to judge the value of each additional new information, and quantitatively evaluate the possibility of survey results by combining prior probability and posterior probability. There is no doubt that Bayesian Decision also has some disadvantages such as too much data to be processed, the subjectivity of prior probability, etc., but generally speaking, the Bayesian theorem makes our decision more perfect and scientific.

"Bayesian Theorem, which was born more than 200 years ago, has milestone significance in solving the calculation of correlation probability under uncertain conditions. It has important applications from making economic and social decisions to daily life, which is also the basis for the development of current artificial intelligence and machine learning." As a mathematical formula and scientific tool on the surface, the Bayesian theorem can not only better guide the research of mathematics, especially probability theory, but also instruct our practical life. For example, the Bayesian theorem teaches us how to combine all kinds of data and update our views according to new evidence, which is also the foundation of the Bayesian theorem. The new evidence in real life cannot determine our views out of thin air, but refresh our transcendental views or our existing experiences. In other words, consider the proportion after seeing all the evidence and limiting the probability space. The application of mathematical rationality in decision-making and judgment fully demonstrates the charm of mathematical tools and probability theory.

In this paper, Bayesian Decision is used to prove the mathematical necessity for multiple rounds of detection in middle and high risk areas. In fact, we are also trying to explain a paradox with probability theory tools, that is, the paradox between overall high accuracy and poor personal predictability in medical detection. Through the Bayesian theorem, it is tangible to manifest that their deviation is a paradox, which can coexist in mathematical logic, while the fallacy comes from the unacceptable reality of human psychology. "Real paradox does not mean the real contradiction". At the moment when the new regulations are released, most areas have liberalized control and nucleic acid is no longer a necessity. However, we still hope to unshackle the logical misunderstanding of multiple rounds of nucleic acid detection through the rational analysis of Bayesian Decision in this paper. Besides, in the reflection on the 3-year epidemic, we should enhance the tolerance of national policies and realize unremitting efforts made by the country to protect people's rights and life safety.

References

[1] Yao, X. T. (2017). Analysis of Bayesian theorem and its application. Education Modernization, 2(6), 254-256.

[2] Liu, L. P. & Gao, L. (2013). Mystery of Bayesian life experience—On the 250th anniversary of the publication of Bayesian theorem. Statistical Research, (12), 2-9.

[3] Surkova, E., Nikolayevskyy, V. & Drobniewski, F. (2020). False-positive COVID-19 results: hidden problems and costs. The Lancet Respiratory Medicine, 8(12),1167-1168.

[4] Arevalo-Rodriguez, I., Buitrago-Garcia, D., Simancas-Racines, D., et al. (2020). False negative results of initial RT-PCR assays for COVID-19: a systematic review. medRxiv, 15(12), 13-17.

[5] Cook, M. (2020). Sleight of Mind. The MIT Press, 87-93.