The Power of Bayes

David F. Percy CMath FIMA CSci, Centre for Operational Research and Applied Statistics, University of Salford, Greater Manchester, M5 4WT.

Summary

The famous theorem attributed to eighteenth century cleric Thomas Bayes has far reaching consequences in many areas of modern life, including medical diagnosis, pre-natal care, safety monitoring, personnel recruitment, quality control, fault diagnosis, lie detection, forgery prevention, drug testing and criminal trials. All involve forms of imperfect screening, where errors have widespread implications. Bayes' theorem reveals all.

Introduction

Bayes' theorem was first published in 1763 shortly after its author's death and Laplace generalized it in 1774. It proved to be invaluable for calculating inverse probabilities, though the many benefits it offers statistical inference and decision theory were only appreciated in the twentieth century, when Jeffreys, Savage, de Finetti, Lindley and others developed the foundations of probability and statistics. This paper investigates the importance and benefits of Bayes' theorem and demonstrates some surprising results, with amazing consequences for practical applications.

For any two events, A and B, the simplest form of Bayes' theorem enables us to evaluate the conditional probability of B given that A has occurred, as

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)}. \tag{1}$$

In applications, the denominator is often determined using the law of total probability

$$P(A)=P(A|B)P(B)+P(A|B')P(B') \tag{2}$$

where B' represents the complement of the event B. Bayes' theorem can be interpreted as updating our prior probability for event B, $P(B)$, to give the posterior probability for event B based on observing event A, $P(B|A)$. Alternatively, we can view it as a formula for transposing the conditionality, or calculating $P(B|A)$ using knowledge of $P(A|B)$.

Medical diagnosis

One symptom of coronary heart disease is the experiencing of chest pains, though indigestion and other unrelated conditions can produce similar pains. In a specific population, the disease is present in about 10% of people. About 80% of people with the disease experience the symptom, whereas only about 25% of people without the disease experience the symptom. A person visits a doctor with chest pains. What is the probability that this person has coronary heart disease?

This is a classic diagnosis problem that is a simple form of medical screening. From the information above, experiencing chest pains appears to be a good indicator of coronary heart disease. Consequently, we might intuitively guess that 75% or 80% of people with the symptom actually have the disease. However, we shall see that this is not so. The solution is readily obtained by defining the events A = "person experiences chest pains" and B = "person has coronary heart disease".

The first paragraph informs us that $P(B)\approx 0\cdot 10$, $P(A|B)\approx 0\cdot 80$ and $P(A|B')\approx 0\cdot 25$ in terms of marginal and conditional probabilities. Using these estimated probabilities, Equation (2) gives

$$P(A)\approx 0\cdot 80\times 0\cdot 10+0\cdot 25\times 0\cdot 90=0\cdot 305$$

and so we can calculate the required probability from Equation (1) as

$$P(B|A)\approx\frac{0\cdot 80\times 0\cdot 10}{0\cdot 305}\approx 0\cdot 26.$$

This result is very worrying, because it means that only about a quarter of people with the chest pain symptom actually have coronary heart disease!

Clearly, this diagnostic indicator is not as good as it appears. What if a positive diagnosis immediately led to cardiac surgery? Then 74% of these traumatic operations would be unnecessary. By applying Bayes' theorem, we are at least aware that further diagnosis is needed before deciding to operate. However, the presence of chest pain symptom in a person increases the probability of coronary heart disease from $P(B)\approx 0\cdot 10$ to $P(B|A)\approx 0\cdot 26$ and so does provide useful diagnostic information.

Incidentally, the probability that a person without the chest pain symptom has coronary heart disease is

$$P(B|A')=\frac{P(A'|B)P(B)}{P(A')}\approx\frac{0\cdot 20\times 0\cdot 10}{0\cdot 695}\approx 0\cdot 03$$

so only 3% of diseased people slip through the net when using this symptom as a diagnostic indicator, which is much better than the 10% that would otherwise be missed.

Probability and odds

In the example above, we drew attention to the use of Bayes' theorem for modelling inverse probability by reversing the order of conditionality of events. The prior probability of disease also has a profound bearing on the accuracy of screening, as we demonstrate now by means of a similar medical application.

Cardiac fluoroscopy is used to diagnose coronary artery disease by detecting whether 0, 1, 2 or 3 arteries are calcified. Consider the events A_i= "i arteries calcified" and B = "disease present". Empirical evidence from medical studies gives the estimated conditional probabilities in Table 1.

Table 1: Estimated conditional probabilities for the diagnosis of coronary artery disease.

i	$P(A_i\|B)$	$P(A_i\|B')$
0	0·42	0·96
1	0·24	0·02
2	0·20	0·02
3	0·15	0·00

Firstly consider a woman aged 30-39 with non-anginal chest pain, for whom $P(B) \approx 0{\cdot}05$. If such a woman has no calcified arteries, Equations (1) and (2) give the posterior probability that she is diseased as

$$P(B|A_0) = \frac{P(A_0|B)P(B)}{P(A_0|B)P(B) + P(A_0|B')P(B')} \approx \frac{0{\cdot}42 \times 0{\cdot}05}{0{\cdot}42 \times 0{\cdot}05 + 0{\cdot}96 \times 0{\cdot}95} \approx 0{\cdot}02$$

For comparison, consider a man aged 50-59 with angina, for whom $P(B) \approx 0{\cdot}92$. If such a man has no calcified arteries, the posterior probability that he is diseased is given by

$$P(B|A_0) \approx \frac{0{\cdot}42 \times 0{\cdot}92}{0{\cdot}42 \times 0{\cdot}92 + 0{\cdot}96 \times 0{\cdot}08} \approx 0{\cdot}83$$

Clearly, the prior probability of disease has a major impact upon the posterior probability of disease.

Figure 1 illustrates the relationship between the prior and posterior probabilities for this example. For a person who presents but has no calcified arteries, the posterior probability of having coronary artery disease exceeds one half (more likely diseased than not) only if this person's prior probability exceeds 0·70.

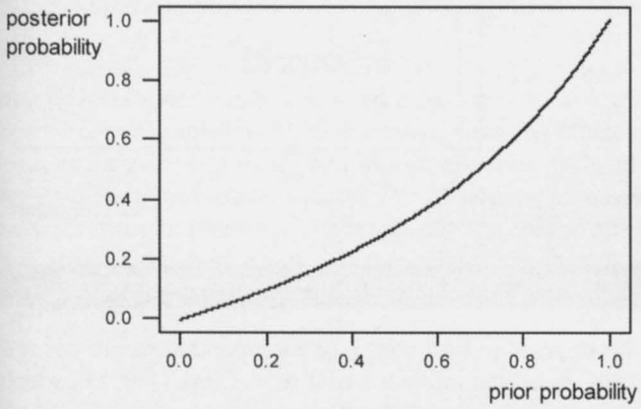

Figure 1: Plot of posterior against prior probabilities of coronary artery disease for a person who presents but has no calcified arteries.

In a general setting, we define the prior odds as $P(B) \,/\, P(B')$ and the posterior odds as $P(B|A) \,/\, P(B'|A)$. The ratio of posterior odds to prior odds is known as the Bayes factor and so can be written

$$BF = \frac{P(B|A)}{P(B'|A)} \div \frac{P(B)}{P(B')} = \frac{P(B|A)}{P(B)} \div \frac{P(B'|A)}{P(B')}.$$

Now, from Bayes' theorem in Equation (1), we see that $P(B|A)/P(B)=P(A|B)/P(A)$ and $P(B'|A)/P(B')=P(A|B') \,/\, P(A)$, so we can express the Bayes factor as

$$BF = \frac{P(A|B)}{P(A|B')}. \tag{3}$$

As such, it represents the strength of diagnosis, or weight of evidence, independently of the prior probability. For the above application relating to cardiac fluoroscopy and coronary artery disease, the Bayes factor corresponding to no calcified arteries is about $0{\cdot}42 \,/\, 0{\cdot}96 \approx 0{\cdot}44$, which is less than one because this indicates disease absent rather than present.

Note that the Bayes factor is the same for both, indeed any, of the people under consideration because it is independent of the prior odds. This is convenient, as it avoids the need to apply Bayes' theorem for each new patient. Rather, we simply evaluate the prior odds and use the result

$$\textit{posterior odds} = \textit{Bayes factor} \times \textit{prior odds} \tag{4}$$

before optionally converting the posterior odds (*odds*) to posterior probability (*prob*), using the formula

$$prob = \frac{odds}{odds+1}. \tag{5}$$

Further mathematical properties of symmetry and additivity can be introduced by taking logarithms of Equation (4). The algebraic simplifications and advantages are substantial, though the introduction of a transcendental function might have the opposite effect by confusing practitioners and deterring them from adopting this analytic procedure. Consequently, I somewhat reluctantly avoid taking logarithms and prefer to work with the structure represented by Equations (4) and (5).

If further diagnostic evidence becomes available, we can compute a new Bayes factor for this evidence and multiply it by the posterior odds obtained from Equation (4) to evaluate revised posterior odds, which can then be converted to give a revised posterior probability using Equation (5).

Safety and reliability

A favourite example to illustrate the power of Bayes' theorem is as follows. An aircraft warning light comes on if the landing gear on either side is faulty. Defining events W = "warning light comes on" and L = "landing gear faulty", suppose we know that $P(W|L)=0{\cdot}999$, $P(W|L')=0{\cdot}005$ and $P(L)=0{\cdot}004$. Our aim is to calculate the probability that the landing gear is faulty if the warning light comes on.

Using the law of total probability

$$P(W) = P(W|L)P(L) + P(W|L')P(L') = 0{\cdot}008976$$

and then Bayes' theorem gives

$$P(L|W) = \frac{P(W|L)P(L)}{P(W)} \approx 0{\cdot}45$$

so most (55%) of these warning lights are false alarms! Fortunately, the probability that the landing gear is faulty if the warning light does not come on is

$$P(L|W') = \frac{P(W'|L)P(L)}{P(W')} \approx 4{\cdot}0 \times 10^{-6}$$

so we need not be too concerned for our safety.

Similar anomalies arise in the process of quality control. Christer (1994) described a situation whereby chips for an integrated circuit are tested automatically. He also presented empirical observations, from which Percy (2003) determined Bayes estimates suggesting that 2·7% of all chips are faulty, 98·2% of defective chips are detected, and 98·2% of functional chips are identified as such.

We are interested in finding the probability that a chip declared faulty is actually sound, as this represents money lost due to wastage. Defining the events F = "chip faulty" and D = "chip declared faulty", we have $P(F) \approx 0{\cdot}027$, $P(D|F) \approx 0{\cdot}982$ and $P(D'|F') \approx 0{\cdot}982$. The law of total probability gives

$$P(D) = P(D|F)P(F) + P(D|F')P(F') \approx 0 \cdot 044028$$

and Bayes' theorem then gives

$$P(F'|D) = \frac{P(D|F')P(F')}{P(D)} \approx 0 \cdot 40.$$

Despite the apparent accuracy of this automatic testing procedure, our result implies that 40% of discarded chips are actually functional!

True or false?

Polygraph tests are used to detect whether a person is lying and are used in recruitment, security, espionage, interrogation, litigation and other areas. Define the events T = "positive test (indicates lying)" and L = "person is lying". From empirical studies, $P(T|L) \approx 0 \cdot 88$ and $P(T'|L') \approx 0 \cdot 86$ so polygraph testing appears to be reasonably accurate.

Suppose that a specific polygraph test is positive, indicating that the subject is lying. Given the observed accuracy of the test, many people expect the probability that the person is actually lying to be fairly high, perhaps around 0·87. However, in security screening most people are honest and $P(L) \approx 0 \cdot 01$, so the law of total probability gives

$$P(T) = P(T|L)P(L) + P(T|L')P(L') \approx 0 \cdot 1474,$$

from which Bayes' theorem gives

$$P(L|T) = \frac{P(T|L)P(L)}{P(T)} \approx 0 \cdot 06.$$

In words, 94% of positive polygraph readings are in error! Imagine the consequences if a positive polygraph test leads to accusations or arrests. This illustrates the danger of using screening procedures on general populations. If the test is used on suspected criminals with $P(L) \approx 0 \cdot 5$, we obtain $P(L|T) \approx 0 \cdot 86$ which is far more acceptable.

Litigation and conviction

Aitken (1996) discusses the use of Bayes' theorem in court. Consider the following situation. A crime was committed in a town of 80,000 people and the offender left blood at the scene, of a type present in 1% of the population. A suspect is found to have this blood type. There are two common mistakes:

- ***Prosecutor's Fallacy.*** "There is a 1% chance that the defendant would have this blood type if he were innocent, so there is a 99% chance that he is guilty."
- ***Defendant's Fallacy.*** "Roughly 800 people in the town have this blood type, so there is only 1 chance in 800 that he is guilty and the evidence is irrelevant."

Defining the events G = "guilty (suspect committed the crime)" and E = "evidence (suspect's blood type matches)", we want to determine $P(G|E)$.

The prosecutor's fallacy correctly states that $P(E|G') = \frac{1}{100}$ but wrongly interprets this as $P(G'|E)$ to evaluate $P(G|E)$. The defendant's fallacy wrongly quotes $P(G|E)$ without regard to $P(G)$ and so ignores the true impact of the evidence. If any of the town's residents could have committed the crime, Bayes' theorem gives

$$P(G|E) = \frac{P(E|G)P(G)}{P(E|G)P(G) + P(E|G')P(G')} = \frac{1 \times \frac{1}{80{,}000}}{1 \times \frac{1}{80{,}000} + \frac{1}{100} \times \frac{79{,}999}{80{,}000}} \approx 0 \cdot 00125,$$

so the evidence is highly relevant, increasing the probability of guilt from $1 \cdot 25 \times 10^{-5}$ to $1 \cdot 25 \times 10^{-3}$. Suppose eyewitness reports imply that only 50 of the 80,000 residents could have committed the crime. In this case, Bayes' theorem shows that the evidence changes the probability of guilt from 0·02 to 0·67. In both cases, the Bayes factor of Equation (3) is 100.

Figure 2 presents a plot of the posterior probability of guilt given evidence, $P(G|E)$, against the prior probability of guilt, $P(G)$, for this analysis. It demonstrates that the evidence is really quite compelling.

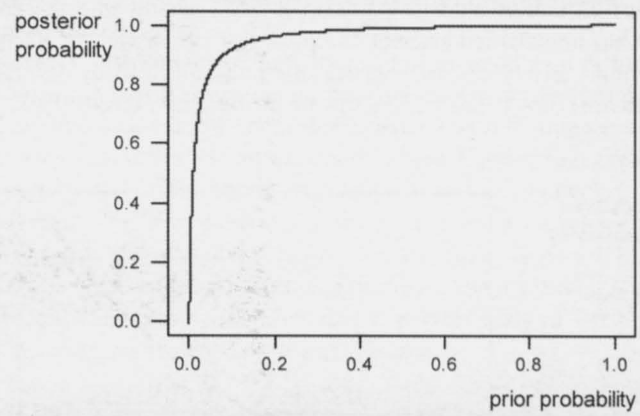

Figure 2: Plot of posterior against prior probabilities of guilt.

If this appears to be a contrived example that would not really happen, recall the infamous trial R. v. Clark (1999) in which a mother was jailed for life, having been found guilty of killing her two children. She protested her innocence and was finally released when her conviction was overturned on second appeal in 2003, her family having suffered considerably in the meantime. This case and others have been well documented by mathematicians including Joyce (2002) and Hill (2005), as well as provoking a rare public statement by the Royal Statistical Society.

At the trial, the jury was told by an expert witness that the chance of a mother's having two natural cot deaths was 1:73,000,000. Quite apart from the fact that this chance should be about 1:130,300 (Joyce, 2002), I suspect that the vast majority of people in the court concluded that such an event was so unlikely that she must have been guilty. They all fell victim to the prosecutor's fallacy, by interpreting the probability of evidence (E) given innocence (G') as the probability of innocence given evidence!

Using a realistic value for the prior probability of guilt (Joyce, 2002), Bayes' theorem shows us that the posterior probability of guilt given the evidence is

$$P(G|E) = \frac{P(E|G)P(G)}{P(E|G)P(G) + P(E|G')P(G')} \approx \frac{1 \times \frac{1}{216{,}667}}{1 \times \frac{1}{216{,}667} + \frac{1}{130{,}300} \times \frac{216{,}666}{216{,}667}} \approx 0 \cdot 38.$$

It would appear that the mother is almost twice as likely to be innocent of the crime as she is guilty! Even then, this ignores much other evidence in favour of her innocence, which would decrease this probability further.

To be convicted as guilty beyond reasonable doubt, the probability of guilt given the evidence should be very large, perhaps 0·99 or more. Without further evidence, it is clear that the mother should never have been arrested at all. One cannot overstate the importance of Bayes' theorem in this context. However, its complexity is such that the form in Equation (4) is far better suited to use in court and elsewhere. The court should specify the prior odds and expert witnesses should specify the Bayes factor of Equation (3). Then the posterior odds can be calculated using Equation (4) and the posterior probability of the defendant's guilt can be calculated using Equation (5).

The Bayes factor for this application is $BF \approx 130{,}300$ and increases the odds of guilt from $P(G)/P(G') \approx 4\cdot 6\times 10^{-6}$ to $P(G|E)/P(G'|E) \approx 0\cdot 60$ which is less than one, implying that the mother is more likely to be innocent than guilty. This is clearly important evidence though: the odds that a mother has murdered two of her children must increase dramatically if we know that two of her children have died. However, for the posterior probability of guilt to exceed the nominal value of 0·99 mentioned above, the posterior odds must be at least 99 and it is very clear that this evidence is woefully short of incriminating her.

Discussion

So far, we have barely mentioned an even greater consequence of Bayes' theorem, which is the decision theoretic philosophy known as Bayesian inference. By specifying a prior probability density function for unknown model parameters and expressing observed data in the form of a likelihood function, we can update the prior given a set of observed data using Bayes' theorem thus:

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}. \quad (6)$$

This enables us to answer any questions we wish about the parameters without estimation, unlike the frequentist approach which is based upon hypothesis tests and confidence intervals.

O'Hagan (1994) argues that the frequentist approach "*...suffers from some philosophical flaws, has a restrictive range of inferences with rather indirect meanings and ignores prior information...*". In contrast, the Bayesian approach "*...is fundamentally sound, very flexible, produces clear and direct inferences and makes use of all the available information...*". These advantages come at a price. The specification of prior distributions remains unclear and the computational effort required is substantial. However, these difficulties are gradually being resolved; see Bernardo and Smith (1993) for example.❑

REFERENCES

1 Aitken, C. (1996). Lies, damned lies and expert witnesses. *Mathematics Today*, **32**, 76-80.
2 Bernardo, J.M. and Smith, A.F.M. (1993). *Bayesian theory*. Chichester: Wiley.
3 Christer, A.H. (1994). Modelling the quality of automatic quality checks. *Journal of the Operational Research Society*, **45**, 806-816.
4 Hill, R. (2005). Reflections on the cot death cases. *Significance*, **2**, 13-15.
5 Joyce, H. (2002). Beyond reasonable doubt. *MathPlus*, **21**, http://plus.maths.org/issue21/features/clark/
6 O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics: Bayesian Inference*, London: Arnold.
7 Percy, D.F. (2003). Subjective reliability analysis using predictive elicitation. In *Mathematical and statistical methods in reliability* (eds. Lindqvist, B.H. and Doksum, K.A.). New Jersey: World Scientific.

d.f.percy@salford.ac.uk

Dr. Dave Percy is a senior lecturer in mathematics, with research interests in applications of Bayesian inference, stochastic processes and multivariate analysis. He is chairman of the Lancashire and North-west Branch of the IMA and lives in the village of Newburgh with wife Mandy and cavalier King Charles spaniels Louis and Daisy!