Bayes Theory Essay, Research Paper

REVIEW OF RELEVANT LITERATURE AND RESEARCH

I first became interested in Bayes’ Theorem after reading Blind Man’s Bluff, Sontag (1998). The book made mention how Bayes’ Theorem was used to locate a missing thermonuclear bomb in Spain in 1966. Furthermore, it was again used by the military to locate the missing submarine USS Scorpion (Sontag, pg. 97) that had imploded when it sank several years later. I was intrigued by the nature of the theory and wanted to know more about it. When I was reading our textbook for the class, I came across Bayes’ Theorem again, and found an avenue to do more research.

There has been much study and many, many articles, papers and books devoted to Bayesian thought and statistics. My research involved literary search at the University of Memphis through Lexis-Nexis, ABI and many other electronic sources available at the University. I read many peer reviewed papers and reviewed several books about Bayed Theorem. I searched the Internet using several search engines and found much of the same literature found through the more conventional methods at the university. Additionally, as part of my research, I conducted an in depth telephone interview with the historian at the Atomic Museum in Albuquerque N.M..

I researched the development of the theorem and its criticism, and included my findings in this paper. Probably the most useful text in understanding the Theorem, and a definitive work supporting its use, is John Earman’s work, Bayes or Bust?: A Critical Examination of Bayesian Confirmation. This book examined the relevant literature and the development of Bayesian statistics as well as defended it from its critics.

LIST OF EQUATIONS AND ILLUSTRATIONS

page

Equation 1: Bayes Theorem A1

Equation 2: Bayes Theorem of Prior Probabilities A1

Equation 3: Bayes Theorem in the example of the caner test A1

Equation 4: Bayes Theorem in the example of the caner test, with

numbers applied A1

Illustration 1: Photo of B52 Bomber A2

Illustration 2: Photo of Lost bomb found off the coast of Spain A2

CHAPTER I THOMAS BAYES

Reverend Thomas Bayes was an English theologian and mathematician born in London England in 1702. His development of what is known today as Bayes’s Theorem contributed a powerful yet controversial tool for assessing how probable a specific event or outcome will be, based on quantitative reasoning. This form of reasoning known as conditional probabilities, has been the subject of much controversy and discussion. Many debate its usefulness as a valid scientific method. However, while it does have shortcomings as pointed but by Pearson who argues that,

It does not seem reasonable upon general grounds that we should be able on so little evidence to reach so certain a conclusion?.The method is much too powerful?it invests any positive conclusion, which it is employed to support, with far too high a degree of probability. Indeed, this is so foolish?that to entertain it is discreditable (1907).

Despite such criticism, it is still used today in all areas of study. Many different forms of this theory have evolved, but for the purposes of this paper, the way of looking at a problem and its solution from the Bayes point of view, can be referred to as Bayesian. “In a weak sense, any position on the foundations of probability which permits the wide or unrestricted use of Bayes’s theorem may be described as Bayesian (Logue 1995, pg. ix).”

Thomas Bayes’ father was one of six nonconformist ministers to be ordained in England in the 17th century. After a private education near his family home in Bunhill Field, he attended the University of Endinburgh, but never finished his degree. Like his father before him, Thomas Bayes was eventually ordained a nonconformist minister. After several years of serving with his father as a Presbyterian minister, he spent most of his career as a minister in Tunbridge Wells until his death in April 1761.

In addition to his position in the community as a minister, he also had the reputation of being “?a good mathematician.” (J.J. Oconnor and E.F. Robertson) In fact, he gained prominence in the field of mathematics by writing a pamphlet defending Sir Isaac Newton from critics of his work on fluxions. As result of the pamphlet, he was nominated and subsequently elected as a Fellow of the Royal Society in 1742.

The organization known as the Royal Society was a scholarly group formed to promote the natural sciences, including mathematics and all applied aspects such as engineering, and medicine. The society was founded in 1660 during the reign of King Charles II, and was incorporated by royal charter in 1662. The society is self-governed by a president and council, whose statutory responsibilities include making appointments to research councils, and it has representatives in the governing bodies of many organizations. The people that nominated Bayes described him as, “a gentleman of known merit, well skilled in Geometry and all parts of Mathematical and philosophical Learning (Norland, 2000 pg. 2)”. Bayes retired from the ministry in 1752 and died nine years later.

CHAPTER II DEVELOPMENT OF BAYES THEOREM

After Bayes death in 1761, his family was left with most of his property, but he also left a small bequest to Richard Price, another minister and amateur mathematician. Among Bayes’ papers, Price found two essays on mathematical subjects. He was so impressed with them that he sent them to the Royal Society hoping they would be published. Bayes set out his theory of probability in one of the essay’s titled towards solving a problem in the doctrine of chances it was so well received by the Royal Society, that it was published in the Philosophical Transactions of the Royal Society of London in 1764.

Bayes essay would shape the nature of statistics. Bayes’ theory stated “Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named (Bayes 1763, pg. 376)”. This problem and the solution it entails have come to be referred to as inverse statistics.

Logue put it this was,

Thus Bayes, in his famous [1763] essay, having defined probability as ‘the ratio between the value at which the expectation depending upon the happening of the event ought to be computed, and the value of the thing expected upon its happening’, then equivocates upon ‘expect’, sometimes taking it as wholly relative to an individual’s mental state, sometimes as though expectations were externally fixed values. (1995, pg.95)

The theorem was eventually accepted by mathematicians of the time. The mathematician LaPlace later accepted it as a valid process as Jaynes (1995) points out, “In almost his first published work (1774), Laplace rediscovered Bayes’ principle in greater clarity and generality, and then for the next 40 years proceeded to apply it to problems of astronomy, geodesy, meteorology, populations statistics and even jurisprudence” (pg. 2). LaPlace generalized Bayes’ approach, which was later generalized further into what we now call Bayes’ theorem. Essentially, the theorem is supposed to quantify the value of a hunch, factor in the knowledge that exists in people beyond their conscious minds. You see, according to Bayes’ theorem, one can always start with a belief with regard to the probability of an outcome and use that in the equation. If one has no prior knowledge, the prior distribution would be diffuse (spread out).

CHAPTER III BAYES THEOREM EXPLAINED

Bayes theorem for conditional probabilities is described by equation 1:

where the marginal probability of A occurring given B has occurred is represented by P(A|B). Said another way, the theorem allows you, knowing little more than the probability of A given B to find the probability of B given A.

A second derivation of Bayes Theorem gives the rule for updating belief in a Hypothesis A (i.e. the probability of A) given additional evidence B. This is shown in equation 2

Where A\* is the false outcome of an argument. The left hand term, P(A|B) is called the posterior probability, and it gives the probability of the hypothesis A after considering the evidence B. P(A|B) is called the likelihood, and it gives the probability of evidence assuming the hypothesis B and the background information is true.

In many situations, predictions of outcomes involve probabilities, one theory might predict that a certain outcome has a twenty percent chance of happening; another may predict a sixty percent chance of the same outcome. In these types of situations, the actual outcome would tend to shift our degree of belief from one theory to the other. As previously noted, Bayes theorem gives a way to calculate this experience or “degree of belief” (Logue 1995, pg.95). To construct an example of Bayes’ theorem, one begins by designing a mutually exclusive and all-inclusive hypothesis. That is, an hypothesis that includes all out comes. Next, one needs to spread out the degree of belief among them by assigning a probability based on what we believe to be true, to each hypothesis. This assignment would be between zero and one to make it all-inclusive. Not the often misused probability such as a person saying, “I am behind you one hundred and ten percent”, for such a statement to be true is not only physically impossible but certainly falls outside the requirement of being all inclusive. If one has no prior basis in either experience, or observation of the hypothesis, one simply spreads out he probabilities evenly among the hypothesis.

The next step in setting up the equation is to construct a list of possible outcomes. The list of possible outcomes, like the hypothesis, should also be mutually exclusive and all-inclusive. Each hypothesis is then calculated with its assigned conditional probability (either based on prior knowledge or assigned randomly if no prior knowledge is present) of each of the possible outcomes. This step simply assigns the probability of observing each outcome if that particular hypothesis is true. The unique part of Bayes’ theorem, and what was new in the 18th century, is one then makes note of which outcome actually occurred and can then compute revised prior probabilities for the hypothesis, (See equation 2), based on the actual outcome.

CHAPTER IV MEDICAL USE OF BAYES THEOREM

Suppose you undergo a medical test for a relatively rare cancer. Your doctor tells you the cancer has an incidence of 1% in the general population. In other words, the chance of you having the cancer is one in one hundred, i.e. a probability of 0.01. The test is known to be 89% reliable. That is, the test will not fail to find cancer when present, but will give a positive result in 11% percent of the cases where no cancer is present, this is known as a false positive.

When you are tested, the test yields a positive result. The question is, given the result of the test, what is the probability that you have cancer. It is easy to assume that if the test is nearly 75% accurate, and you test positive, then the likelihood you have the cancer is about 75%. That assumption is way off. The actual likelihood you have cancer is merely 3.9% (i.e., the probability is 0.039). Three point nine percent is still something to worry about with cancer but hardly as daunting as 75%. The problem is, that the 75% reliability factor for the test, has to be balanced against the fact that only 1% of the entire population has the cancer. Using Bayes’ method ensures you make proper use of the information available.

As I have discussed, Bayes’ method allows you to calculate the probability of a certain event C (in the above event, having the cancer), based on evidence (e.g. the result of the test), when you know (or can estimate):

(1) The probability of C in the absence of any evidence:

(2) The evidence of C

(3) The reliability of the evidence (i.e., the probability that the evidence is correct.

In this example, the probability in (1) is 0.01, the evidence in (2) is that the test came out positive, and the probability in (3) has to be computed from the 75% figure given. All three pieces of information are highly relevant, and to evaluate the probability that you have the cancer you have to combine them I the right manner. Bayes’ theorem allows us to do this.

To simplify the illustration, assume a population of 10,000, since we are only interested in percentages, the reduction in population size will not affect the outcome. Thus, in a population of 10,000, 100 will have cancer and 9,900 will not. Bayes method, as previously mentioned, is about improving an initial estimate after you have obtained the new evidence. In the absence of the test, all you could say about the likelihood of you having the cancer is there is a 1% chance that you do. Then you take the test, it shows positive. How do you revise the probability that you have the cancer?

There are, we know, 100 people in the population that do have the cancer, and for all of them, the test will show a positive result. But what of the 9,900 people that do not have the cancer, for 25% of them, the test will incorrectly give a positive result, thereby identifying 9900 X .025 = 2,475 people as having the cancer when they actually do not. Thus, overall, the test identifies a total of 100 + 2,475 = 2,575 people as having the cancer. Having tested positive, you are among that group (evidence tells you this). The question is, are you in the group that has the cancer or the group that does not but tested as if they did (false positive). Of the 2,575 that tested positive, 100 of them really do have cancer. To calculate the probability that you really have cancer, you take the number that really does (100) and divide it by the number that would have tested positive but do not have it (2,475) and you get the probability that you really do have the cancer. 100/2,475 = 0.039. In other words, there is a 3.9% chance that you have the cancer (see equation 4).

This calculation shows why it is important to take account of the overall incidence of the cancer in the population. This, in the Bayesian way of thinking, is known as prior probability. Being able to calculate results based on this prior probability, either known, speculated or not known, is the advantage of using Bayes’ theorem. In our case, in a population of 10,000 with a cancer having an incidence rate of 1%, a test reliability of 75% will produce 2,475 false positives. Thus far outweighs the number of actual cases, which are only 100. As a result, if your test comes back positive, the chances are overwhelming that you are in the false positive group. This data, as it is placed in Bayes’ theorem flows like this: Let P(H) represent the probability that the hypothesis is correct in the absence of any evidence – the prior probability. Therefore, H is the hypotheses that you have the cancer and P(H) is 0.01. You then take the test and get a positive result, this evidence of cancer we will call C. Let P(H|C) be the probability that H is correct given the evidence C. This is the revised estimate we want Bayes’ theorem to calculate. Let P(C|H) be the probability that C would be found if H did occur. In our example, the test always detects cancer when it is present so P(C|H) = 1. To find the new estimate, you have to calculate P(H-wrong), the probability that H does not occur, which is .099 in this case. Finally, you have to calculate P(C|H-wrong), the probability that the Cancer C would be found (i.e., a positive test) even though H did not actually occur (i.e., you do not have the cancer), which is 0.25 in the example. In equation 3, Bayes’ theorem states:

Using the formula for our example in equation 4:

The quantity such as P(H|C) is known as a conditional probability. That is, the conditional probability of H occurring given the evidence C.

CHAPTER V BAYES AND THE LAW

Bayes theorem is used in mathematics and, as I previously mentioned, in many professions. Amongst them is the practice of law. There have been instances where lawyers have taken advantage of the lack of mathematical sophistication among judges and juries by deliberately confusing the two conditional probabilities P(G|E), the probability that the defendant is guilty given the evidence, and P(E|G), the conditional probability that the evidence would be found assuming the defendant would be guilty. Intentional misuse of probabilities has been known to occur where scientific evidence such as DNA testing is involved, such as paternity suits and rape and murder cases. In such cases, prosecuting attorneys may provide the court with a figure for P(E), the probability that the evidence could be found among the general population, whereas the figure of relevance in deciding guilt is P(G|E). As Bayes’ formula shows, the two values can be very different, with P(G|E) generally much lower than P(E). Unless there is other evidence that puts the defendant into the group of possible suspects, such use of P(E) is highly suspect. The reason is that, as with the cancer test example, it ignores the initial low prior probability that a person chosen at random is guilty of the crime in question.

Instructing the court in the proper use of Bayesian inference was the winning strategy used by American long-distance runner Mary Slaney’s lawyers when they succeeded in having her 1996 performance ban overturned. Slaney failed a routine test for performance enhancing steroids at the 1996 Olympic games, resulting in the United States athletic authorities banning her from future competitions. Her lawyers demonstrated that the test did not take proper account of the prior probability and thus made a tacit initial assumption of guilt.

In addition to its use, or misuse, in court cases, Bayesian inference methods lie behind a number of new products on the market. For example, the paperclip advisor that pops up on the screen of users of Microsoft Office — the system monitors the user’s actions and uses Bayesian inference to predict likely future actions and provide appropriate advice accordingly. For another example, chemists can take advantage of a software system that uses Bayesian methods to improve the resolution of nuclear magnetic resonance (NMR) spectrum data. Chemists use such data to work out the molecular structure of substances they wish to analyze. The system uses Bayes’ formula to combine the new data from the NMR device with existing NMR data, a procedure that can improve the resolution of the data by several orders of magnitude.

Other recent uses of Bayesian inference are in the evaluation of new drugs and medical treatments, the analysis of human DNA to identify particular genes, and in analyzing police arrest data to see if any officers have been targeting one particular ethnic group.

CHAPTER VI BOMB BAYES OPEN

Another particularly fine example of Bayes’ theorem in action is displayed in the following example. On January 17, 1966, while attempting a mid-air refuel at 30,000 feet off the coast if Palmares Spain, a Strategic Air Command B-52 Statofortress Bomber (See illustration 1), collided with an air borne KC-135 Stratotanker fuel tanker aircraft. The collision killed all four of the crewmembers on the KC-135 and three of the seven crewmembers on the B-52.

When the collision occurred, the aircraft was carrying four Hydrogen Bombs. As result of the collision, all four bombs departed the aircraft in the air. Three of the four bombs were recovered almost immediately along the shore of Spain (See illustration 2). However, search teams could not locate the fourth bomb. It was lost and had presumably fallen to the bottom of the Mediterranean ocean.

1966 was a time of tremendous strain in the relations between the then Soviet Union and the United States of America. Tensions were running high partially because the on-going conflict in Vietnam and failed Bay-of-Pigs invasion of Cuba. The cold war was at its peak and competition for superior nuclear weapons design between the United States and the Soviet Union was fierce. When the United States government notified the government of Spain of the incident, its government, fearing a radiation leak, demanded a clean up and assurances from the United States that they would recover the bomb. The United States military knew the Soviets were aware of the accident and in an effort to retrieve the bomb and glean secrets from its design and construction, were looking for the bomb too. United States president Lyndon Johnson refused to believe his military’s assertions that there was a good probability the bomb could not and would not ever be recovered, by ether side, because of its presumed depth and condition. The president demanded its recovery.

A team of was assembled to try to pinpoint the location for a search and to attempt to retrieve the weapon once it was (if ever) located. The group was attempting to use Bayes’ Theorem. A group of mathematicians were assembled to construct a map of the sea bottom outside Palomares, Spain. Once the map was completed, the U.S. Navy assembled a group of submarine and salvage experts to place probabilities that Sontag describes as”?Las Vegas-style bets?” [pg. 63] of each of the different scenarios (outcomes), that might describe how the bomb was lost and what happened to it once it departed the aircraft. Each scenario left the bomb in not just a different place, but in a wide variety of places. Then, each possible location (all inclusive), was considered using the formula that was based on the probabilities created in the initial phase of the equation, that is the phase that set the initial probabilities.

The theory created by the 18th century theologian and amateur mathematician provided a way to apply quantitative reasoning to what we normally think of as a scientific method. That is, when several alternative theories, such as the case of the missing H-bomb, about an outcome exist, you can test them conducting experimental tests to see, whether or not those consequences actually occur. Put another way, if an idea predicts that something should happen and it does actually happen, it strengthens the belief in the validity of the idea. It acts as a spoiler too, if an actual outcome contradicts the idea, it may weaken the belief in the idea.

After a betting round to assign probabilities of the location of the bomb, the locations were then plotted again, sometimes great distances from where logic and acoustic science would have place them. The bomb, according to S.D. Bono, the Historian of the National Atomic Museum in Albuquerque NM, had been connected to two parachutes designed by the Sandia Corporation. “Sandia was the general contractor of the weapon system itself as well as the parachutes” (S.D. Bono, personal communication, February 14, 2001). The parachutes complicated the issue further because no one knew if they functioned or not. Whether they did or not, and what condition they were in, could have had dramatic effect on the bomb’s ultimate resting place. As part of applying Bayes’s theorem, the researchers asked the experts individually how they expected the event unfolded. Going over each part of the event with each participant. The team of mathematicians wrote out possible endings to the crash story and took bets (set probabilities), on which ending the believed to be most likely. After the betting was over, they used the odds created to assign probabilities to several locations identified by the betting. The site was again mapped, but this time, the most probable locations were marked according to relevancy.

The team used the Bayes theorem to map out where they believed the bomb was. According to their calculations, the most probable location of the bomb was a tremendous distance from where the other three bombs were located and a good distance from where most of the aircraft’s debris had hit the water. The team provided their data to the search team who immediately began the search.

The search was complicated by the fact that they had pinpointed a location that was at the bottom of a deep undersea ravine, making the search very difficult. After just two weeks of searching, president Johnson called the man responsible for the search to Washington to be briefed on the search. Upon learning the technique that had been used to determine the location of bomb, Johnson was furious. He could not believe the hope of finding the bomb before the Russian’s was tied to what appeared to outsiders, to such a plan of betting on where experts thought it would be. He called for another review of the data and circumstances by yet another group of mathematicians. The second group of mathematicians gathered, looked at the method used to determine the bombs location and could develop nothing better, and reported to president Johnson that there was no better way. Within days of the second group reporting to Johnson and after weeks of searching and revising the probabilities based on actual results, the bomb was located and eventually retrieved. The bomb was exactly where the team’s latest Bayesian calculations said it would be. The theory developed by an 18th century minister had found America’s lost Bomb.

CONCLUSION

As demonstrated by the examples shown in this paper, Bayesian thought has found its place in statistical thinking. It provides a mathematical approach to what is often called a hunch, and takes advantage of information an expert may have, but it unable to put into words. It allows one to refine ones theories about outcomes based on what has actually occurred, in practical use, it works well.

There are some that say its ineffective for the very reason it has gained much favor, but for many it serves a useful purpose in mathematical thinking. A amateur mathematician changed the was we see and calculate probabilities, and had it not been for his friend, Richard Price, who sent his essays to the Royal Society after his death, we would be without this very useful theory.

Bibliography

REFERENCES

Allenby, Greg M. (1980, August). Cross-Validation, the Bayes Theorem, and Small-Sample Bias Journal of Business and Economic Statistics. pp.171-179.

Adams, E., and Rosenkrantz, R.D. (1980, December). Applying the Jeffrey decision model to rational betting and information acquisition. Theory and Decision pp.1-20.

Aronson, J, (1989). The Bayesians and the raven paradox. Nous 23:221-240.

Arrow, K (1971). Essays in the theory of risk-bearing. Chicago: Markham Press.

Baker, F. and Evans, R. (2000). The Probability of Mr Bayes. Melbourne, University of Melbourne.

Barhard, G.A. (1958). Thomas Bayes – A biographical note. Biometrika 45:293-295.

Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London 53:370-418. Reprinted in facsimile in W.E. Deming, ed., Facsimiles of Two papers by Bayes (Washington, D.C.: U.S. Dept. of Agriculture, (1940).

Billingsley, P. (1979). Probability and Measure. New York: John Wiley.

Bonjour, L. (1985). The structure of empirical knowledge. Cambridge: Harvard University Press.

Dale, A.I. (1982, May). Bayes or LaPlace? An Examination of the Origin and Early Applications of Bayes’ Theorem. Archive for History of Exact Sciences pp.23-47.

Devlin, K.J. (1999). Turning Information into Knowledge. New York: W.H. Freeman and company.

Dorling J., and Miller D. (1981, April). Bayesian Personalism, Falsificationism and the Problem of Induction. Proceedings of the Aristotelian Society supp. Pp. 109-41.

Edidin, A. (1983, August). Bootstrapping without Bootstraps. In Earman

Edwards, A.W.F. (1978, April). Commentary of the Arguments of Thomas Bayes. Scandinavian Journal of Statistics pp.116-118

Fildes, R. (1983, February). An Evaluation of Bayesian Forecasting. Journal of Forecasting pp.137-151.

Fodor, J. (1984, May). Observation Reconsidered. Philosophy of Science pp.:23-41.

Horwich, P. (1982). Probability and Evidence. Cambridge: Cambridge University Press.

Jaynes, E.T. (1978). Bayesian Methods: General Background An introductory Tutorial. St. John’s College and Cavendish Laboratory, Cambridge England.

Kelly, W. and Chainani, G. (1983, Summer). Probability Considerations in Decision Theory. Cost Engineering pp.15-22.

Kyburg, H. (1978, July). Subjective Probability: Criticisms, Reflections and Problems. Journal of Philosophical Logic. Pg. 176

Kyburg, H. and Smokler, H., (eds.). (1980). Studies in Subjective Probability. New York: John Wiley.

Lehman, R.S. (1955). On Confirmation and Rational Betting. Journal of Symbolic Logic 20:251-262.

Logue, J. (1995). Projective Probability. Oxford: Oxford University Press

Maydew, R.C. (1966). America’s Lost H-Bomb! Palomares, Spain 1966. Kansas, Sunflower University Press.

Mellor, D.H. (1971). The Matter of Chance. Cambridge:Cambridge University Press.

Pearson, K. (1907, August). On the Influence of Past Experience on Future Expectation. The Philosophical Magazine pp.365-378.

Popper, K. (1961). The Logic of Scientific Discovery. New York: Science Editions.

Redhead, M.L.G. (1980, November). A Bayesian Reconstruction of Methodology of Scientific Research Programs. Studies in the History and Philosophy of Science pp. 674-347.

Seidenfeld, T. (1979, November). Why I am not an Objective Bayesian: Some Reflections Prompted by Rosenkranz. Theory and Decision pp.413-440.

Smith, A.F.M. (1986). Why isn’t Everyone a Bayesian? Comment American Statistician 40(number 1):10.

Smith, C. (1997). Theory and the Art of Communications Design. Seattle, State of the University Press.

Sontag, S and Drew, C. (1998). Blind Man’s Bluff. New York; HarperCollins

Spielman, S. (1977). Physical Probability and Bayesian Statistics. Synthese 36:235-269.

APPENDIX

Equation 1

Equation 2

Equation 3

Equation 4

ILLUSTRATIONS

Illustration 1

A B52 Stratofortress owned by the National Atomic Museum, similar to the one that crash off the coast of Spain in 1966.

Illustration 2

One of the three Atomic bombs found on the coast of Palmares Spain, in 1966.