The Mathematics Enthusiast

Manuscript 1597

Things Behind Bayes’ Theorem A review of Sharon Bertsch McGrayne’s The theory that would not die: How Bayes' theorem allowed to decipher the enigma code, to pursue the Russian submarines and to emerge triumphant from two centuries of controversy from a statistics education perspective

Ernesto Sanchez

Nathalia Morgado

Eleazar Silvestre

Things Behind Bayes' Theorem
A review of Sharon Bertsch McGrayne's *The theory that would not die: How Bayes' theorem allowed to decipher the enigma code, to pursue the Russian submarines and to emerge triumphant from two centuries of controversy* from a statistics education perspective

Ernesto Sanchez[1]
Center for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV), Mexico City

Nathalia Morgado[2]
Center for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV), Mexico City

Eleazar Silvestre[3]
The University of Sonora (UNISON), Sonora, México

If you are not thinking like a Bayesian, maybe you should.

~John Allen Paulos, 2011

A feeling emerges from reading McGrayne's book expressed as a paraphrase of a verse from Hamlet: *There are more things behind Bayes' theorem, Horatio, than dreamt of in your philosophy*. The book tells a story of the birth and survival of an idea around which social enterprises, personal odysseys, and human passions are masterfully woven by the author. The book has 17 chapters distributed in five parts and two annexes. These parts reveal a narrative structure worthy of Propp: The hero of the story (Bayes' theorem) is born and raised normally but rejected later. It is forced to act in secret but returns to the light and is submitted to several tests. The hero emerges triumphant from all of them. The above is just an outline of a complex but accessible chronicle that includes many brilliant and fascinating figures and extraordinary situations. We recommend the following three reviews that undoubtedly motivate you to immerse yourself in reading the book: Paulos (2011), Wainer & Savage (2012) and McGrayne (2015).

[1] esanchez@cinvestav.mx
[2] cindy.morgado@cinvestav.mx
[3] eleazar.silvestre@unison.mx

The Mathematics Enthusiast, **ISSN 1551-3440**, vol. 20, nos. 1, 2 & 3, pp. 121-131.

In this document, we present six key issues of the book we considered important and useful from a statistical education perspective; we formulate and support our claims with some data taken from different parts of the book since the material is abundant enough to delve into each of them.

Causes, Effects and Uncertainty

In the 1740s, an English Presbyterian named Thomas Bayes stated a version of the theorem that bears his name; after finishing it, he filed it without trying to publish it for the rest of his life. It would be Richard Price, a young friend of Bayes and a fan of mathematics, who would polish and publish it in 1763, two years after Bayes's death. McGrayne wonders: What motivated and inspired Bayes to state the theorem that would lead to 200 years of controversy? She confesses that we do not know with certainty, but she outlines the intellectual environment of England in which such figures as Newton, Hume, Berkeley, and de Moivre play an important role. She showed that in those days the problems related to cause, effect, and uncertainty were floating in the air, and perhaps this influenced Bayes to study the relationships between these concepts quantitatively. Later, in the 1770s, the French scientist and mathematician Pierre Simon, Marquis de Laplace, independently rediscovered the theorem of the probability of the causes of events. Over the years, the tenacious Laplace, among the many scientific projects that he developed, perfected the theorem in such a way that he ended up expounding it practically in the way we know it today. He applied it to solve problems of demography, jurisprudence, statistics, and astronomy. Earlier, in 1781, Laplace learned about the Bayes/Price paper via the Marquis de Condorcet and would recognize that: "The theory whose principles I explained some years after [...] he accomplished in an acute and very ingenious, though slightly awkward, manner" (Quoted in McGrayne, 2015). Bayes's paper revealed to Laplace the idea of assuming an initial assumption of *a priori* probabilities, and he immediately made it his own by incorporating it into his version. Thus, from its inception, there is a subjective probability associated with the theorem, which would make it the target of future criticism.

The Central Idea

Although the theorem offers a procedure for calculating the probability that an event has been produced by a cause considering the different possible causes, it can be reformulated and interpreted in terms of initial beliefs that are updated based on recent information as represented in a simple one-liner:

Figure 1. Beliefs explained

Initial beliefs + Recent objective data = A new and improved belief

The epigraph that McGrayne chose for his book is a sarcastic phrase by John Maynard Keynes that briefly expresses the meaning of Bayes's theorem: "When the facts change, I change my mind. What are you doing, sir?" Despite its simplicity, the idea touches on a deeper problem of scientific thought: How do we use evidence to update our knowledge? Paulos (2011) suggested that the way Bayes's Theorem answers this question is: "If you want to assess the strength of your hypothesis given the evidence, you must also assess the strength of the evidence given your hypothesis." Following Paulos's description, the theorem states that the posterior probability of a hypothesis is equal to the product of (a) the prior probability of the hypothesis and (b) the conditional probability of the evidence given the hypothesis, divided by (c) the probability of the new evidence, which is:

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}$$

$P(H)$ and $P(E)$ are the prior probabilities and $P(H|E)$, $P(E|H)$ are the posterior probabilities.

The Heart of the Debate

After Laplace's death, European thinkers and scientists were at the forefront of the attacks on Laplace. John Maynard Keynes mentions that George Boole, John Venn, and Joseph Bertrand rejected Laplace's reasoning, arguing respectively that his hypotheses are arbitrary, not in keeping with the experience, and ridiculously so. However, intellectuals such as Augustus de Morgan, William Jevons, Rudolf Lotze, Emanuel Czuver, and Karl Pearson accepted Laplace's thinking. In the 1970s, Charles Sanders Peirce would promote the use of probabilities based on frequencies. Thus, there were two antagonistic tendencies distinguished, namely the Bayesians and the Frequentists, even though Laplace had proposed a combined and delicate balance in using *a priori* probabilities, and frequencies.

We should point out that the heart of the controversy was between the procedure of conjecturing prior probabilities with subjective criteria defended by Bayesians against the need to start from objective probabilities defended by Frequentists. Statisticians accepted the applications of Bayes' theorem when they knew a prior probability of a particular situation, or computable through a valid procedure; but this condition is very restrictive because many interesting situations do not fulfill. The debate begins when Bayesians fill this gap by making a conjecture about prior probabilities; this procedure is subjectivity run amok for frequentists. Two eminent statisticians struck the hardest blow against Bayes. One of them was Ronald Aylmer Fisher, for whom Bayes' theorem constituted "a mistake, perhaps the only mistake to which the mathematical world has so deeply committed itself", "My personal conviction [...] that the theory of inverse probability is founded upon an error and must be wholly rejected" (quoted in McGrayne, 2012, p.48). The other was Jerzy Neyman for whom, despite his rivalry with Fisher, agreed with outlawing the *a priori* subjective probabilities in using Bayes' theorem.

The Applications

While Frequentist statistical inference led by Fischer, Pearson, and Neyman flourished in the community of statisticians, Bayesian methods were secretly advancing far from the spotlight. McGrayne reveals that statisticians like Alan Turing were using Bayes' theorem to decipher Germany's naval Enigma code and, thus knowing the positions of the German U-boats, established trajectories for the supply ships that allowed England to resist the German offensive during World War II. Unfortunately, during the Cold War, the creative contributions of Turing to Bayesian methods were classified and kept secret. Also, a hydrogen bomb lost at sea near a Spanish town called Palomares due to an accident involving a US military plane; the search method created to find the missing hydrogen bomb used and developed Bayesian techniques that have subsequently been used to search large areas for missing ships or planes.

Bayesian methods were applied to situations to assess probabilities of singular events without data from previous occurrences. The frequentist methods do not offer any strategies to solve problems in such situations. McGrayne highlights that besides the search for the lost hydrogen bomb in the sea, the problem of estimating the probability of an event considered impossible would be like those of an accident at a nuclear plant and two planes collided in midair. In both cases, Bayesian analyzes found that the probabilities, while small, were not negligible. When those events occurred, the losses in both cases corroborated the Bayesian estimates. Mosteller and Wallace (1984) used a Bayesian approach to resolve the authorship of Federalist articles. McGrayne also describes how John Tukey led an effort to predict election results better and faster than anyone else through the secret use of Bayes' theorem. The theorem was also used in medicine to improve medical diagnoses through accumulated data processing. The practical statistician Jerome Cornfield analyzed previously accumulated data on lung cancer using the Bayes' theorem and published an important paper arguing the existence of a causal relationship between smoking and the development of this disease.

The Technology

In most cases, Bayes' theorem is used reiteratively to update probabilities each time new data emerges; with each iteration, the number of operations needed to resolve a problem multiplies and performing these computations becomes a real challenge. Laplace developed analytics methods to deal with the computations he needed to perform with his data while Alan Turing designed and constructed a specialized machine to process his own. In the years following World War II, technology that would allow Bayesians to materialize their ideas in important applications was still lacking so they were restricted to study problems that implied a scarce number of manageable integrals. But everything changed, starting in the 1990s; technological development allowed access to potent desktop computers, new computational methods, and software to execute all sorts of computational tasks. These resources made it possible for Bayesians to resolve important problems by avoiding the obstacles related to the computational burden. Therefore, Bayesian theory is widely used in the digital world and in many applications; furthermore, attacks and disqualifications have been left aside and the theory has been accepted as a legitimate theory in statistics. McGrayne (2015) mentions:

> Thanks to Bayes, we can filter spam, assess medical and other risks, search the Internet for the web pages we want, and learn what we might like to buy, based on what we've looked at in the past. The military uses it to sharpen the images produced when drones fly overhead, and doctors use it to clarify our MRI and Pet Scan images. It is used on Wall Street and in astronomy and physics, the machine translation of foreign languages, genetics, and bioinformatics. The list goes on and on. (p. 159)

The Theoretical Advances

In the 1920s, Emile Borel, Frank P. Ramsey, and Bruno de Finetti, from France, England, and Italy, respectively, were three mathematicians who were working individually, and almost simultaneously, formulated the idea of conceiving probability as a measure of a person's degree of subjective belief, which could be calculated as a function of the bet the person is willing to take in the defense of such belief. Based on this conception of probability, Ramsey postulated the problem of finding the formulas to make decisions under uncertainty. As well, de Finetti provided subjective probability a solid mathematical ground in addition to applying his theory to the domain of financial economy. Despite the theoretical advances made by these authors, the statistics community made them a target for attacks and deliberated indifference. At the same time, their attention and recognition were drawn to the anti-Bayesian triplet: Ronald Fisher, Egon Pearson, and Jerzy Neyman. During the next decades, statistics would be dominated by the research agenda proposed by these anti-Bayesian statisticians. Despite such an adverse environment, geologist Harold Jeffreys would keep Bayes' theorem alive by applying it to the processing of data when determining the most likely epicenter of an earthquake. He also developed a new form of Bayes' theorem that allowed the amplification of its applications, developing a new set of formal rules for the selection of the *a priori* probabilities. The year the First World War began, Jeffreys published his Theory of Probability which was the first systematic exposition of the ways Bayes' theorem is applied in scientific problems. Principles of the Bayesian approach would be rebirthed after the war by means of Jack Good's pen, Leonard J. Savage and Dennis V. Lindley. Good, who worked with Turing during the war, wrote a book titled *Probability and the Weighting of Evidence* that would be published in 1950 where Good mentions that he retakes Alan Turing's ideas about his fight against the Enigma code in his book. Nowadays there are many expositions of Bayesian statistics.

Reading McGrayne's book is fascinating because it illustrates the dramatic process from birth to consolidation of a statistical idea, its rejection and neglect from the statistics community and its marginal survival when used to resolve problems loaded with unmanageable uncertainty, to finally rebirth with an unexpected strength during the contemporaneous technological world. The book is so stimulating that for statistics educators who read it would evoke a reflection like Paulos': if the statistics curriculum for the school levels does not contain a Bayesian approach, maybe it should.

Addendum

From a statistics education research perspective, there is a restless feeling that emerges when wondering what has been researched about Bayes' theorem within the education domain. It is hard to fully know the wide extent of educational research done in the tertiary levels related to Bayes' theorem. In disciplinary fields that belong to medicine, psychology, informatics, artificial intelligence and others, the applications of the theorem are especially important and, therefore, it is likely that such communities had already dove into research in Bayesian reasoning.

But more popular in the educational domain is the research done in the field of Psychology, given that many of the results of the research program of Heuristics and Biases under Uncertainty (Kahneman, Tversky & Slovic, 1982) had a major impact in statistics education research (Shaughnessy, 1982). One question that guides such a research agenda is: Are people good intuitive statisticians? (Kahneman, 2011). A main finding of this program unveiled aspects of Bayesian problems in which most people's reasoning fails to provide an appropriate normative answer. A paradigmatic situation is represented by the Taxi problem (Bar-Hillel, 1980; Kahneman & Tversky, 1972) and revisited by Shaughnessy (1992). The problem goes as follows:

> Two cab companies operate in a city, the Blue and the Green (according to the color of cab they run). Eighty-five percent of the cabs in the city are Blue, and the remaining 15% are Green. A cab was involved in a hit-and-run accident at night. A witness later identified the cab as a green cab. The court tested the witness' ability to distinguish between blue and green cabs under nighttime visibility conditions. It found that the witness was able to identify each color correctly about 80% of the time but confused it with the other color about 20% of the time. What do you think are the chances that the errant cab was indeed Green, as the witness claimed? (Kahneman & Tversky, 1972; Bar-Hillel, 1980)

This is an interesting problem whose solution is reached with the help of the Bayes theorem; the solution is $p = 0.41$ which is not intuitive at all. A formal solution requires translating the information to mathematical symbology for one to apply the theorem, a task with its own level of complexity. There are instructional proposals to solve the problem in which, with the help of other representations (trees or rectangles), the symbolic load is lightened (Shaughnessy, 1992). But what is relevant from psychologists' point of view is that individuals – including experts –, when trying to provide an answer, reveal a strong tendency to ignore base probabilities (*a priori* probabilities) and support judgments relying only on the witness probabilities (posterior probabilities). Such omission is not due to distraction or neglect, but to a failure to consider why base probabilities should be involved. This reasoning bias is called the base-rate fallacy (Bar-Hillel, 1980). Bar-Hillel mentions that the fallacy's presence in reasoning about certain contexts can lead to inconvenient results with serious practical consequences (e.g., when assessing confidence in diagnostic tests or evidence in criminal trials). Indeed, to ignore base-rates can produce an overestimation of probabilities of target events in a problem given. Besides the characterization of the base-rate fallacy, educational research has also characterized the next students' difficulties related to the wider concept of conditional probability:

- *Translation of natural language to symbolic language*. As shown in the taxi problem, knowledge of the symbolic expression of Bayes' theorem is not enough to solve Bayesian problems found at the school level. These are usually formulated in natural language, and one must identify all components embedded in Bayes' theorem. Perceiving and distinguishing between simple, compound, and conditional probabilities in the text of a problem is a necessary requirement for applying Bayes' theorem in problem solving, but making such distinctions also constitutes serious learning difficulties for students. (Díaz, Ortiz & Serrano, 2007)

- *Fallacy of the transposed conditional*. Students frequently believe the fact that conditional probabilities and its transposed version were the same, i.e., they assume that $P(A|B) = P(B|A)$. Difficulties for distinguishing each are highlighted when such probabilities must be identified in information expressed verbally in a problems' text. (Falk, 1986; Ojeda, 1994; Yáñez, 2003; Díaz & de la Fuente, 2006)

- *Causal and chronological conceptions of conditional probability*. Frequently, events that intervene in a conditional probability are organized through causal schemes because the situations' context suggests cause-effect relationships that are not significant to the problem. Time is involved in many contexts and can easily call for a chronologist conception of probability when is assumed that affects the outcome of the problem. (Tversky & Kahneman, 1980; Falk, 1986; Gras & Totohasina, 1995)

One final thought is that McGrayne's research on the development of Bayes' theorem may suggest new directions in educational research. Statistics education research should explore forms of introducing Bayes, guided by the central idea of answering the question: How to use evidence to update our knowledge? Attention to the development of a Bayesian component in teaching must solve the problem of how and when to introduce the subjective approach to probability. Although subjective probability is mentioned in several educational studies, it is not included in the statistics curriculum of pre-university levels and, despite the recommendations of Hawkins and Kapadia (1984), the research does not seem to offer any alternative yet for teaching the subjective approach to probability. It would also be necessary to find new problems for teaching related to the current applications of Bayes' theorem which could be supported with technological resources. It is necessary that we understand how the problem of learning and teaching Bayesian reasoning is transformed with the availability of technological resources. Progress in these aspects can constitute concrete referents to understand the theory of Bayesian statistics.

References

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211–233. https://doi.org/10.1016/0001-6918(80)90046-3

Díaz, C., Ortiz, J. J., & Serrano, L. (2007). Dificultades de los estudiantes de psicología en el cálculo de probabilidades inversas mediante el teorema de Bayes. *Revista de La Universidad de Granada Publicaciones*, *37*, 141–156.

Díaz, C. & de la Fuente, I. (2006). Enseñanza del teorema de Bayes con apoyo tecnológico. *Investigación en el aula de matemáticas*. *Estadística y Azar*. Sociedad de Educación Matemática Thales.

Falk, C. F. (1986). Promoting continuing education programs. *New Directions for Adult and Continuing Education*, *31*, 49–71. https://doi.org/10.1002/ace. 36719863106

Gras, R., & Totohasina, A. (1995). Conceptions d'élèves sur la notion de probabilité conditionnelle révélées par une méthode d'analyse des données: Implication - similarité - Corrélation (Students' Conceptions on Conditional Probability Revealed by a Data Analysis Method: Implication - Sim.) *Educational Studies in Mathematics*, *28*(4), 337–363.

Hawkins, A.. & Kapadia, R. (1984). Children's conceptions of probability - A psychological and pedagogical review. *Educational Studies in Mathematics* 15, 349-377.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. https://doi.org/10.1037/h0034747

McGrayne, S. (2015). The theory that never died. How an eighteenth century mathematical idea transformed the twenty-first century. *Mètode Science Studies Journal*, *5*, 159–165.

Mosteller, F., & Wallace, D. L. (1989). Deciding authorship. In J.M. Tanur et al (Eds.), *Statistics: A Guide to the Unknown* (pp. 115-125). Wadsworth Brooks.

Ojeda, A. M. (1995). Dificultades del alumnado respecto a la probabilidad condicional. *UNO*, *5*, 37–55.

Paulos, J. A. (2011). *The Mathematics of Changing Your Mind*. The New York Times.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 465–494). Macmillan Publishing Co, Inc.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgements under uncertainty. In. M. Fischbein (Ed.), *Progress in Social Psychology,* NJ: Erlbaum.

Wainer, H., & Savage, S. (2012). The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines and Emerged

Triumphant from Two Centuries of Controversy by McGrayne, Sharon Bertsch. *Journal of Educational Measurement*, *49*(2), 214–219. https://doi.org/10.1111/j.1745-3984.2012.00171.x

Yáñez, G. (2003). *Estudios sobre el papel de la simulación computacional en la comprensión de las secuencias aleatorias, la probabilidad y la probabilidad condicional*. Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México.