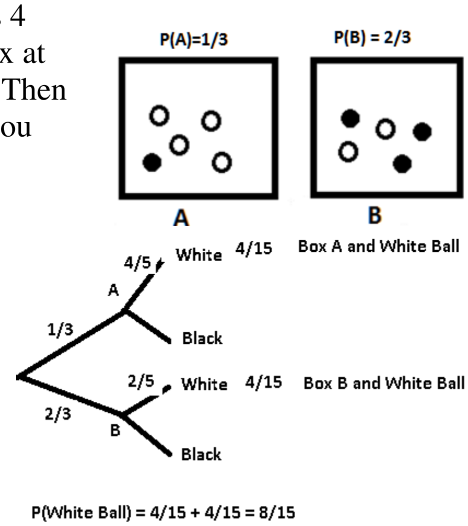
History of Probability (Part 4) - Inverse probability and the determination of causes of observed events. Thomas Bayes (c1702-1761).

By the early 1700s the work of Pascal, Fermat, and Huygens was well known, mainly couched in terms of odds and fair bets in gambling. Jacob Bernoulli and Abraham DeMoivre had made efforts to broaden the scope and interpretation of probability. Bernoulli had put probability measures on a scale between zero and one and DeMoivre had defined probability as a *fraction of chances.* But then there was a 50 year lull in further development in probability theory. This is surprising, but the "brains" of the day were busy applying the newly invented calculus to problems of physics – especially astronomy.

At the beginning of the 18^{th} century a probability exercise like the following one was not too difficult.

Suppose there are two boxes with white and black balls. Box A has 4 white and 1 black, box B has 2 white and 3 black. You pick one box at random, with probability 1/3 it will be Box A and 2/3 it will be B. Then you randomly pick one ball from the box. What is the probability you will end up with a white ball?



We can model this scenario by a tree diagram. We also have the advantage of efficient symbolism. Expressions like $P(A) = \frac{1}{3}$ had not been developed by 1700. The idea of the probability of a specific event was just becoming useful in addition to the older emphasis on odds of one outcome versus another.

Using the multiplication rule for independent events together with the addition rule, we find that the probability of picking a white ball is 8/15.

The question that received attention nearer the end of the 18^{th} century was the inverse. *Suppose you know that someone who played this game ended up with a white ball, what is the probability it came from Box A?*

This is called a problem of **inverse probability**. You know how the story ends but you don't know for sure how it began. Questions of inverse probability are of great general interest, not just in gambling. For example, you observe the outcome that a person escaped the flu this year. Why? What was the reason, the "cause?" Is it *because* she had a flu shot? Maybe. Can you assign a probability to the "maybe?" If you can then you can use this probability to estimate the chance that *in the future* someone who has had a shot will escape the flu. This kind of reasoning from accumulated experience has long been the basis of establishing risk in the insurance industry.

A probability for an event that depends on some other event or condition, is called a ***conditional probability***. The chance of getting the flu evidently is influenced by (is conditional on) whether or not you had the flu shot.

The significant advance in the mathematics of probability in the late 18^{th} century was to develop rules for calculations with conditional probabilities. In the opening example about the white and black balls we would like a formula for determining the probability that the person chose Box A *given that* we know she ended up with a white ball. This is usually symbolized now as **P(A | White ball)**. The vertical line is read as "given that."

If we simply ask what is the probability in this experiment that a person picks Box A, the answer is 1/3, which is the unconditional probability of A. P(A)= 1/3. (There is no vertical line in the symbol.) But if we ask this *after we know she ended up with a white ball* that increases our sense that she had picked Box A. The conditional probability P(A| white ball) is greater than the unconditional probability, P(A).

A significant breakthrough in calculating conditional probability was published in a paper in 1764 by the English mathematician, **Thomas Bayes (c1702-1761)**. Bayes derived a rule, which we now express as an equation, for determining a conditional probability. Today we call this rule **Bayes' theorem**,

In our modern notation it can be expressed this way: $\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)}}$**.** The symbol ∩ means "and."

In our example, A represents box A being chosen, and B represents white ball being chosen.

Referring to the tree diagram above we therefore get

$P(\text{Box A }|\text{White Ball}) = \frac{P(\text{Both Box A and White Ball happen})}{P(\text{White Ball is chosen})} = \frac{\frac{4}{15}}{\frac{8}{15}} = \frac{1}{2}$, which is greater than $P(\text{Box A}) = \frac{1}{3}$.

Bayes' rule is central to any attempt to link the calculation of the probability of an event to prior information. Bayes' own version was restricted to binomial type questions of success and failure, but later versions are more general.

"Bayesian Analysis" now refers to methods where the statistician specifically assigns (guesses?) reasonable probabilities for prior events, essentially making the whole analysis a matter of conditional probabilities. There is vigorous research over the question of how to choose potential prior probabilities. The Bayesian label came into use only in the 1950s, particularly as formerly secret methods of intelligence used by British statisticians in WWII were published. In the United States early Bayesian analysis was put to unusual use. "… in the mid-1950s, Fred Mosteller and David Wallace began a major collaboration to investigate who wrote the disputed Federalist papers. The Federalist papers were written mainly by Hamilton and Madison to help convince Americans to ratify the constitution in the 1780s. The authorship of most of the papers was clearly identifiable, but historians had been unable to establish whether a small number of them had been written by Hamilton or Madison. The work began in the summer of 1959 and … they set to work on what was to become the first serious large-scale Bayesian empirical investigation involving heavy use of computers. (Fienberg)

Early development of Bayes' Theorem and Inverse Probability
Thomas Bayes gets credit for the first publication of this rule, but it is unclear how much direct impact his paper had at the time. Bayes had left money in his will to pay an acquaintance, **Richard Price,** to edit and publish his papers.** Price was a member of the Royal Society and an expert in insurance risk. He submitted Bayes' edited paper to the Journal of the Royal Society with the title ***An Essay towards solving a Problem in the Doctrine of Chances*****.** The phrase, "Doctrine of Chances," is an older expression for "theory of probability." This essay is the one for which Bayes gets the credit for his theorem.

Here is the opening of the article and the first calculation by the rule.

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763. I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved.

PROP. 5.

If there be two ſubſequent events, the probability of the 2d $\frac{b}{N}$ and the probability of both together $\frac{P}{N}$, and it being 1ſt diſcovered that the 2d event has happened, from hence I gueſs that the 1ſt event has alſo happened, the probability I am in the right is $\frac{P}{b}$.

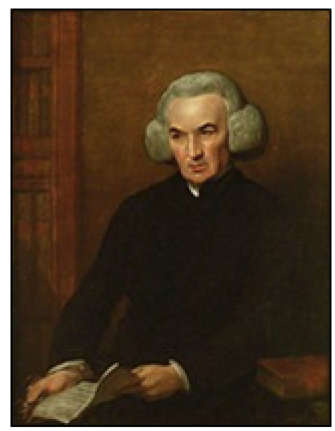
In modern symbols, Bayes' Proposition 5 is: Given $P(B) = \frac{b}{N}$, $P(A \cap B) = \frac{P}{N}$ then $\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)}} = \frac{P/N}{b/N} = \frac{P}{b}$.

The ideas in Bayes' paper did not reach a wide audience until the French mathematician **Pierre-Simon Laplace** took up similar problems a few years later and figured out the same rule himself. In 1774 Laplace's first paper on this topic was published, ***Mémoire sur la probabilité des causes par les événemens*** **(*Memoir on the Probability of Causes Given the Events*)**. It is not just that Laplace presented his version of Bayes' rule, but that he developed new mathematical ways to solve the equations that were needed to establish the more general version of the rule. Here is a bit from the opening paragraphs of his memoir. Notice that he calls DeMoivre a "geometer." The mathematicians from England had been following Newton's version of calculus, which was highly geometric. Laplace, in France, was more algebraic, following Leibniz.

> The theory of chances is one of the most curious and most difficult parts of analysis, due to the delicacy of the problems it produces and the difficulty of submitting them to calculation. It appears to have been treated with the most success by M. Moivre, in an excellent work entitled Theory of Chances. We owe to this able geometer the first research that was done on the integration of differential equations by means of finite differences.
>
> . . .
>
> [In this memoir] I propose to determine the probability of the causes of events, a question which has not been given due consideration before, but which deserves even more to be studied, for it is principally from this point of view that the science of chances can be useful in civil life.

To quote historian Stephen Stigler, "The influence of this memoir was immense. It was from here that 'Bayesian' ideas first spread through the mathematical world…"

****Richard Price**



In the history of mathematics, Price is mainly known as the man who edited the posthumous work of Thomas Bayes. In political history, you will find him to have been an influential supporter of the American Revolution.

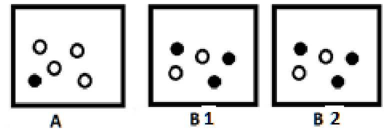
Here are a couple of quotations from the Wikipedia article on Price:

"Richard Price (23 February 1723 - 19 April 1791) was a Welsh moral philosopher, nonconformist preacher and mathematician. He was also a political pamphleteer, active in radical, republican, and liberal causes such as the American Revolution. He was well-connected and fostered communication between a large number of people, including Founding Fathers of the United States."

"At home, or at his church itself, Price was visited by Franklin, Thomas Jefferson, and Thomas Paine; other American politicians such as John Adams, who later became the second president of the United States, and his wife Abigail..."

In 1781 Yale University awarded two honorary doctorates, one to George Washington and one to Richard Price.

Exercises



1. In the opening problem, the probability of picking Box B was double that for Box A. We can represent the situation with this picture by making two copies of Box B. Now there are 15 equally likely balls in the sample space.

Based on this representation,

a. By looking at the collection of balls justify why P(White ball) = 8/15

b. Justify why P(Box A | White Ball) = ½

c. Put the information about the balls into this table.

How can you see here that P(Box A | White Ball) = ½?

	White	Black	Total
Box A			5
Box B1			5
Box B2			5
Total			15

2. This medical example is based on one from the book, *Willful ignorance* by Herbert Weisberg. An example like it is almost always given in an introduction to conditional probability, especially in medical applications.

a. A doctor is trying to determine if her patient suffers from a particular type of infection. Her diagnostic procedure will indicate the infection (correctly) with probability 90% if the infection is truly present. We represent this as: P(test positive | infection) = .90. The procedure also has a false positive rate of 5%. In symbols, that is P(test positive | no infection) = .05.

The big question is what is the probability the patient has the infection if she tests positive? In symbols, that is written as P(infection | test positive). To use Bayes' rule you also need to know the prevalence of the infection, P(infection), the overall portion of the doctor's patients who have it. Let us take the case where the prevalence is 2%. The infection is uncommon among her patients.

This problem is most easily solved by expressing the information in terms of a table of *frequencies* rather than probabilities. Arbitrarily make the total number of patients large, say 1000. Then fill in the cells using the probability percentages.

	Test is positive	Test is negative	
Infection			2% of 1000 =20
No infection			
Total			1000

The ratio of the two shaded cells will give P(infection | Test positive). Confirm that the probability the patient has the disease if she tests positive is only about 27%.

b. How would the answer have changed if the disease were fairly common, say, with prevalence = 80%?
c. Comment on how the answer depends on the prevalence.

3. (Just for fun.) Write a word-for-word modern English translation of Bayes' Proposition 5.

Sources

Victor Katz. *A History of Mathematics*, Addison Wesley Longman, 1998;
Herbert Weisberg. *Willful Ignorance*, Wiley, 2014
Thomas Bayes. An Essay towards solving a Problem in the Doctrine of Chances, *Journal of the Royal Society*, 1764, http://rstl.royalsocietypublishing.org/content/53/370.full.pdf
Stephen Stigler. Laplace's 1774 Memoir on Inverse Probability, *Statistical Science* 1986 v1 (3)
Stephen Fienberg. When did Bayesian Inference become "Bayesian?", *Bayesian Analysis,* 2006 Number 1
Images: Richard Price: www.bbc.co.uk/arts/yourpaintings/paintings/dr-richard-price-17231791-dd-frs-120186