

THOMAS BAYES

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.

Phil. Trans. 53, 370-418 (1763), doi: 10.1098/rstl.1763.0053, <http://rstl.royalsocietypublishing.org/content/53/370>

Probably everybody in science has heard the name of (Reverend) Thomas Bayes (born in 1701 or 1702). The Encyclopaedia Britannica calls Bayes a “nonconformist theologian (presbyterian) and mathematician”. Before his studies in “divinity and logic” he was most likely privately educated. Later he received the major part of his education in Mathematics at the University of Edinburgh. Bayes became minister of his presbyterian church, first in London and then in Tunbridge Wells. He died in Tunbridge Wells in 1761.

This review is about the work for which he is best known today, that is, his Essay towards Solving a Problem in the Doctrine of Chances, communicated by Price to the Royal Society. It contains among other results a special case of his celebrated inverse-probability formula. Before reviewing his Essay it may be helpful to give some more background.

It was not hard to find out more about the communicator to the Royal Society indicated in the title. It was Mr. Richard Price (1723-1791), a Welsh moral philosopher. Price was active as a writer in radical republican and liberal causes like, among others, the American revolution. Price also wrote on statistics and finance, works less known today, but to which, apparently, he owed his nomination as a Fellow of the Royal Society (F.R.S.). The style of his communication advertising Bayes' work shows much respect towards John Canton (a highly regarded physicist), but it shows true interest in Bayes' work, and also self-confidence of his own judgment of Bayes' work. Price invested his effort to attract Canton's attention to this work of Bayes. No doubt, much credit should be given to him for this.

The “A.M.” part in Canton A.M.F.R.S., by the way, probably just means “Artium Magister” (MA) although this way of aligning a degree and a distinction would nowadays be unusual. I owe this explanation to Frank P. Kelly, F.R.S. Kelly also informed me about *S.M. Stigler's* paper “Who discovered Bayes' Theorem?” [Am. Stat. 37, 290-296 (1983; Zbl 0537.62004)] on which we should comment in the present “classical” review.

First a few comments on the style of Bayes' essay. At the beginning, the reviewer found this essay hard to read. Learning on the way Bayes' terminology made it then easier; nevertheless, although the arguments are rather elementary, this takes still some effort. Here it is amazing to see again what a difference terminology can make. For instance we may wonder nowadays why an author, educated as Bayes was, would not create a word for the binomial coefficient $\binom{n}{k}$ instead of speaking of something like "that coefficient that will be attached to the term containing a^k if the expression $(a+b)^n$ is developed into its parts according ...", but this is the way he *always* wrote.

There are many calculations in this essay. Price was apparently worried that Canton may hesitate for that reason to publish the Essay, pointing out that he does not expect Canton to go through the details. He assured Canton that he himself checked all the calculations and found no mistake, taking all responsibility for possible errors. (How good to know for the present reviewer, who could confine himself to sampled checking, agreeing with Price.)

Now, what exactly is Bayes' famous PROBLEM? It is announced in Section I of the Essay and, reads:

Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its (specific event) happening in a single trial lies somewhere between any two degrees of probability that can be named.

The interchange of "chance" and "probability" and terms like "degrees of probability" and others should not confuse us. If we go on and focus on the result we see what is meant. Today we would say: "Given that n independent Bernoulli trials with unknown constant success probability p have brought k successes, what is the probability that the parameter p lies between two given bounds?" And, ironically, we are seduced to ask "Mr. Bayes, do you mean in a Bayesian setting with a given prior density for p ?" Indeed, this is what he meant in today's language, and his prior, without saying so, is the uniform prior on $[0; 1]$.

In his essay, Bayes answers this question via several intermediate results he derived, and for which he needed basic definitions. We sample a few of his definitions and notations:

Definition 1.5.: The probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed and the value of the thing expected upon it's happening. Indeed, if we get a success with probability p and a reward R upon a success and nothing on a failure it is true that $p = E(\text{reward})/R = Rp = R$. Terms like expectations or conditional expectations are sometimes imperceptibly interchanged. Again the reading shows he makes no error.

"Inconsistent" means "disjoint", "contrary" means "complementary", etc., but one quickly gets used to this. Sometimes things become little puzzles, however, and not only for non-native English speakers, I guess. See for example Prop. 6 on page 382: *The probability that several independent events shall all happen is a ratio compounded of the probabilities of each.* Why ratio? Why not the product? What Bayes is bound to have meant is that this probability is the product of the respective probabilities which turns out to be a ratio (fraction), of course, since all his probabilities are understood as rational numbers.

As so often, there is reward in going on with the effort of reading. One understands the way Bayes was thinking. The essential point is that he seemingly preferred to think in terms of games and expected rewards, i.e., he had a clear preference to translate probabilities into expectations of gain or loss functions. Interestingly, as much as we profit (nowadays) from the simple identity of the probability of an event A and the expectation of its indicator I_A , the more complicated the arguments become if one does not use this intrinsic link. Bayes had to go through different interpretations of loss functions instead.

Why Bayes thought in terms of expectations of loss function is not clear, in particular as he was essentially concerned with the "uniform" case of a priori equally likely events. Why then not use the intuitive notion as in Laplace's definition of probability of an event A as being the number of cases which are "favourable" for A divided by the number of possible outcomes of experiments?

British education with its traditions may have played here a role. The British are known to like games, and to express probabilistic or statistical statements quite often in terms of games. Perhaps this was always the case. This may explain, by the way, why the word "odds" is not only an English-language creation but can also be heard in Britain at least as often as the word probability. Thinking in terms of games has certainly done no harm to probability theory in Britain and/or to its distinguished probabilists, thus the non-British should not wonder. Also, Bayes may have had independent reasons for his preference.

Thomas Bayes is particularly known for his “formula” to compute inverse conditional probabilities. As Stigler and others point out, Bayes' formula (theorem) is never clearly stated in his essay. He only proved a special form of it, and Proposition 5 is closest to an explicit statement. However, the way to the formula is apparent in several instances of the Essay.

Bayes' formula as seen today: The formula is a versatile tool, both in theory and applications. Clearly, it had an important impact on many problems and it is difficult to imagine it could possibly lose, one day, its interest. Hence it is immortal and one of the jewels in Mathematics one would like to have discovered oneself: very useful, intuitive (I think), and in modern language of Probability very easy to prove. Recall today's notation (Ω, F, P) for a probability space and $P(A/B)$ for the conditional probability of an event A given an event B . The latter is defined by $P(A/B) = P(A \cap B)/P(B)$, provided that $A, B \in F$ and $P(B) \neq 0$. Bayes' “inversion formula” says

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

To prove this equality it suffices to use the definition of a conditional probability and the commutativity of the set operation \cap , namely

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

$P(B)$ can be written for any partition $\{A_1; A_2; \dots\}$ of Ω as $P(B) = \sum_k P(B/A_k)P(A_k)$. So we can replace A above by an arbitrarily chosen but fixed A_k . Hence $P(A_k/B)$ is expressed in terms of the $P(B/A_j)$'s and absolute probabilities so that the name “inversion” is adequate.

Let us take an example (which is not in the essay and will only serve as a comparison). We have two urns: urn I contains one black and two white balls, and II contains two black and three white balls. One urn is chosen (in obscurity) according to $P(I) = 0.4$, $P(II) = 0.6$, and then one ball is sampled at random from that urn. If the ball is black (B), what is the probability that it stems from urn I? By Bayes' formula we get

$$\begin{aligned}
 P(I|B) &= \frac{P(B|I)P(I)}{P(B)} = \frac{P(B|I)P(I)}{P(B|I)P(I) + P(B|II)P(II)} \\
 &= \frac{(1/3)(2/5)}{(1/3)(2/5) + (2/5)(3/5)} = \frac{5}{14}.
 \end{aligned}$$

A word on intuition: The reviewer always liked Bayes' formula because (rare event) he happened to discover it independently in an unprepared exam. Indeed, intuition tells us to look for an equivalent model with a uniform choice of urns and a uniform choice of balls. Here is the solution to the above example in unsophisticated shorthand notation:

$$\left\{ \frac{2}{5} \rightarrow (1B, 2W) \middle| \frac{3}{5} \rightarrow (2B, 3W) \right\} \leftrightarrow_{\text{urns}} \{2(1,2)|3(2,3)\} \leftrightarrow_{\text{balls}} \{2(5,10)|3(6,9)\}$$

The answer is $10/28=5/14$ since in the last 5-urn model each ball is equally likely to be chosen, now having 10 black balls on the left and 18 black balls on the right.

What we have done is to create a *uniform urn-and-balls-model*, i.e., adapt the number of urns according to the probabilities $P(I)$ and $P(II)$ and then go over to equal numbers of balls in each urn (respecting the relative frequencies of colours). This works for an arbitrary number of urns and arbitrary contents and may be considered as an algorithmic version of Bayes' formula for rational numbers. This is a very intuitive approach, I think, and for a smaller number of urns and balls by the way not bad as an algorithm. The reviewer found nowhere a reference to this, although he thinks it unlikely that he was the first one seeing this.

This algorithm (or anything similar of this kind) may convince us that many people may have discovered Bayes' Theorem independently. We should keep this in mind for the arguments given below.

Bayes' formula and its origin: Bayes did not yet have the notion of a probability measure, or even "random variable", of course. The latter took roughly one and a half centuries more to be born. Bayes' essay is another instance where we can see that these were major steps in the history of probability, making many things so much easier. Bayes came up with his result through calculations. We are not allowed to be turned off by this; we should appreciate his insight.

We will possibly never know whether Bayes was really the first to see the formula; see Stigler's interesting article for an extensive analysis of this question. Stigler gives an amusing Bayesian argument to indicate that the odds are three to one against him. More precisely, Stigler gave the three for the English mathematician N. Saunderson (1682-1739) who was indeed, for multiple reasons, a very remarkable scientist. The original part of Stigler's argument is that he applied a Bayesian argument to "beat Bayes". We should add that all frequentists of the bad kind would probably enjoy seeing what can be done with priors! The three-to-one result of this analysis of Stigler's is not serious, and not meant to be serious, I guess. Also, with Stigler's "law of eponymy" (cf. [S.M. Stigler, Transactions of the New York Academy of Sciences, Ser. 2, 39, 147-158]), Stigler is allowed to stay faithful to his own dogma and to preach for his chapel, as the French would say. Nevertheless, Stigler's arguments based on a considerable amount of historical research are relevant and show that certain questions of the true origin of the formula stay open.

This reviewer, possibly biased by his own experience with Bayesian problems reported above, still sees things differently, however. He has little reason to doubt that Bayes' essay is his own and not a reproduction of ideas of others. And it is good to see that Stigler is careful to avoid any implication of this kind. Bayes had, as I see it, looked at the typical questions leading to his formula, though the formula itself, we say it again, seems to be nowhere clearly stated. He had good ideas about what he wanted to say on the way to it. If Price claimed that Bayes had solved a problem A. de Moivre could answer only partially, then this can be defended. We should not be confused or even become suspicious by the style of an author who, like Thomas Bayes, had to create his own terminology before being able to express his thoughts.

I also see a true motivation for Bayes' essay. Bayes' research was motivated (this was also true for Price) by religious-philosophical questions, as for instance questions concerning the different proofs of existence of God. For a minister, as Reverend Bayes was in his official position, we agree that such questions must have been a strong motivation to give an answer, and this quite independently of having broader interests.

Taking all these points together (and I believe that Stigler would not disagree) the reviewer can conclude that, unless the contrary is proven, we are all entitled to be faithful to the name Bayes' Theorem.

F. Thomas Bruss (Brussels)