

## **ESSAY ON BAYES**

**by W. Stanners**

**Abstract** It has been said, fairly plausibly, that "Bayesian inference is one of the most widely known eponyms in all of science". But unlike common scientific eponyms, it is by no means clear exactly what "Bayesian" means, and what it has to do with Bayes. "Bayesian", and the dozen or so words and phrases which are usually associated with it, seem to be more like unspecific words of the English language, deployed by an author as he wishes, rather than fixed technical terms. The obscurity of the language, relative to the precise meanings associated with, say, Newton's laws or Heisenberg's uncertainty principle, is matched by the obscurity of the history - the virtually unknown Bayes, the posthumous paper, the impenetrable and incoherent style, the muddled logic, the virtual silence on his work for 200 years, the sudden emergence in the last several decades, not of new knowledge, but of new Bayesian additions to the vocabulary. This note surveys the notions and the history. It concludes that the Bayesian vocabulary is vague and pretentious, and serves no useful purpose.

## ESSAY ON BAYES

by W. Stanners

### 1. The Bayesian vocabulary

Bayes, Bayesian, Bayes' rule, Bayes' Theorem, Bayesian inference, Bayesian inverse probability, conditional probability, prior "causes" with "subjective" probabilities, posterior "events" - references to Bayes are common. "Two centuries after Bayes' death, Bayesian inference ... [is] ... one of the most widely known eponyms in all of science" (Stigler, 1986) . Bayes' paper "must rank as one of the most famous memoirs in the history of science" (Barnard, 1958). How many generations of students have puzzled over the numerous contextual references to Bayes, perhaps tried in vain to dig a little deeper, and then contented themselves with mastering the jargon and the procedures?

For let us think of other "eponyms in all science" - perhaps Newton's laws of motion, the laws of Hooke and Boyle, Heisenberg's uncertainty principle.. There is rarely any mystery. Everyone who has any need to know them knows exactly what they are. There are unique dictionary definitions.

What is the Bayesian method? I quote from the entry under this heading in the Routledge Dictionary of Economics (1995). "A method of revising the probability of an event's occurring by taking into account experimental evidence. Bayes's theorem of 1763 originally stated that the probability of  $q$  conditional on  $H$  (prior information) and  $p$  (some further event) varies as the probability of  $q$  on  $H$  times the probability of  $p$ , given  $q$  and  $H$ ." No one needing to know will find enlightenment here!

The problem of knowing what Bayes is, as opposed to knowing how to use his name as another word in the English language, is not an easy one to get to grips with. As a striking example of this, one may cite Stigler himself. Having read there that "Bayesian inference" is one of the most widely known eponyms in all science, one looks it up in his index. There are six page-references. None have more than the usual contextual uses of the term. Nowhere is it explained what it is.

This extends everywhere. It is common for the word Bayesian to appear in a text for the first time with no explanation of its significance. One context will imply that subjectiveness is of the essence, another will describe Bayesian problems with no trace of subjectivity. One may read of the crucial importance of Bayes regarding "inverse probability", while not being informed as to what this may mean, and then find in various contexts that it can mean almost anything to do with probability. In some problems, there are causes and effects, in others merely events occurring together, in others still (including Bayes' own original problem) the entities are not events at all.

This seems to be the story of an OK word, or collection of words and phrases, rather than of a scientific or mathematical concept.

## 2. Bayes and "his" Theorem

In all of this, Thomas Bayes is blameless. When he died in 1761, a more or less unknown Presbyterian clergyman, his "Essay" was found among his papers. It made no reference to his "theorem", or to "inference", or to "causes", or to "subjectivity". No one paid any attention to it after it was published three years after his death, with a preface by Richard Price. Although his father, another fairly unremarkable dissenting clergyman, was memorialised in the Dictionary of National Biography in 1885, Thomas was not noticed in this way until 1993, 232 years after his death. The Encyclopaedia Britannica did not mention him until 1958, and had still little about him in 1970. When the Bayes industry began around 1950 or 1960, the only readily available copies of his essay were, apart from the original volumes, some photocopies circulated by the US Department of Agriculture.

Perhaps the firmest ground from which to set off is what is called "Bayes' Theorem", or Bayes' rule. The first acquaintance students make with this is a goggle-making formula, which I will refrain from giving now. Algebra usually simplifies, but when the simple but evolving steps of an arithmetical procedure are involved, the putting of the procedure into symbols has the effect, perhaps satisfying to textbook writers and lecturers, but not to students, of making the simple seem horrendously difficult. So it is with "Bayes' Theorem", which has the added complexifier of Bayes' name added to it. Many scientists and engineers will in their normal work have applied the trivial arithmetic of "Bayes' Theorem" many times without ever having been introduced to it.

It always a matter of choice what imagery is used to describe this sort of arithmetical process. The usual choice is of prior events, and posterior events, and the probability of one event occurring, given that another has already occurred. It will be seen later that there are applications where this language can be applied only in a tortured way. I prefer to use the idea of a "path", and of flow or traffic along the path, rather than a probability.

Suppose a stream of water is separated into two streams in the ratio of 3 to 8. These numbers may be accurately determined, or may contain some degree of guesswork or estimation. Suppose that somehow the two streams are salinated in their separate channels or paths to levels denoted by 7 and 4, and that the stream is then re-united and mixed. Clearly the salt in the downstream flow will have come from the two channels in the ratio of 21 to 32. The proportion of salt downstream which comes from the first channel is clearly  $21/53$ , or about 40%. I venture to think that any reasonably numerate schoolchild could go through this, step by step, without great difficulty. If the problem was re-posed in terms of 9 or 57 channels, the same child would have little problem in seeing how to generalise or automate the calculation. Since the problem requires no more than ordinary common sense to understand, and nothing more than the arithmetical operations of multiplying, adding, and dividing to solve, any reasonably trained person of any civilised culture in the last 5000 years, could easily deal with it. The notion of *probability* is a little more modern, but once defined and understood, it is clear that in the above numerical example, 40% is also the probability that a molecule of salt comes from the first channel.

If the schoolchild was old enough to have begun algebra, he or she could readily render the generalised or automated procedure for 9 or 57 channels into one for  $n$  channels, with two pieces of data for each one. The same is true for any reasonably educated person since algebra was invented. If they happened on a particularly barbaric but unexceptionable symbolism, the result would be

$$P(B_i|A) = P(A|B_i)P(B_i)/\sum_i\{P(B_i)P(A|B_i)\}$$

and this head-breaker is what students see in their text books under the label of "Bayes' Theorem" or Bayes' Rule. The left hand side is 0.40. The right hand side is  $(3*7)/(3*7+8*4)$ , also a rather unwieldy formulation of something which is crystal clear as presented verbally above. Admittedly, the numbers on the right hand side are not probabilities, but the factors which would make them so cancel out.

Notice that my presentation contains absolutely none of the confusing jargon and mystification associated, and unfairly associated, with the name of Bayes, nothing about a theorem, rule, conditional probabilities, cause and effect, subjective judgment, Bayesian inference, inverse probability, although all of these are there. The so-called "Bayes' Theorem" is not a theorem. It required no discovery in 1763 (and Bayes himself made no such claim), or in 1774 when Laplace independently wrote down a similar "principle". It is simply ordinary arithmetic which anyone in the last 5000 years might have used, without tuition from Bayes or Laplace, given a suitable problem.

### 3. The bridge from de Moivre to Laplace

And *that*, the problem, is the novelty, not the "Bayesian" language which surrounds it. It is improbable that a Babylonian ever asked what fraction of Tigris silt came from which tributary. He would have understood the question and the method of answering it perfectly, but would have wondered why anyone would ask it. Indeed, many textbook examples and mock exam questions are highly improbable or unconvincing in the same sense. But there are modern concerns, as will be illustrated below, which do require this sort of arithmetic.

How did Bayes come to be honoured, quite wrongly, by this very posthumous incorporation into some areas of statistical language?

Bayes was a clever man. Stigler credits him with being the first, with Thomas Simpson (another English oddity, who started his career as a weaver), "to deal with the inference problem". Although it was Laplace who did most to "turn the approach of de Moivre on its head", these men led the way. I am not sure this is so, as far as Bayes is concerned (even aside from the fact that Laplace had, in his earlier years, never heard of Bayes), but first, what is the inference problem? Stigler, I have no doubt, knows exactly what he means by it, but presumably assumes his readers do too, for he does not specifically set it out.

He rather presents a progress from de Moivre to Laplace via Simpson and Bayes (among a few others) as if this was from lower to higher, from the easier to the more

difficult. It is rather as if one might present the evolution from Beethoven to Mendelssohn as an upward progress.

Inference, by which crudely is meant proceeding from the known observations to the hidden, guessed at, or deduced reality behind them, i.e., going in the reverse or "inverse" direction from nature, effect-to-cause instead of cause-to-effect, is rather pictured as an intellectual hurdle. Thus, Stigler says that, "for de Moivre, chance lay in the data, not in the underlying probabilities. The successful mathematical treatment of the inference problem was to require the abandonment of this restrictive view."

I believe that this is a wrong perception. Far from being the higher activity, inference is the very expression of human reason in its most primitive form. Daily we solve inference problems. If a door sticks, we swing it around, try to see where it rubs, *look for the cause*. I suppose no man ever looked at the rain, and said, this will cause rivers to flow. But he would in the course of time infer that the rain was antecedent to the river.

De Moivre would have been surprised to find that he was being portrayed as one failing to solve the inference problem. As is pointed out above, he could have solved Bayes' problem (the one Bayes did solve as opposed to the one he made a muddled claim to solving) with great ease, if he had been motivated to do so. His interest lay elsewhere. It was in a matter of enormous interest to his greatest predecessors and to the learned public - games of chance. Unlike work on the mathematics of inference, the first examples of which related to the observations of astronomers, de Moivre's work was not, and not intended to be, of any *use*. On the other hand, it was widely acclaimed, recognised and published.

It will later appear that Bayes' problem (what it was rather than what he said it was) was a problem of the de Moivre type, and *not* of inference, so it is worth saying in a little more detail what this type was. The gaming world was one of equal chances. The coin had two identical sides, apart from the distinguishing stamp. The die had six identical faces. The lottery had identical balls drawn blind from an urn. The playing cards were dealt with the identical faces up. The spinning wheel could equally stop in any orientation. The binomial series of Bernoulli and de Moivre gave the distribution of any combination of heads and tails in a given number of throws. In this world, the real physical characteristics were the given, inherent in the fair construction of the devices. Nothing further was to be found out about those. That is, no "inference" was involved. The enquiry was on the outcome of games of chance using those devices. Suppose an urn has 37 black balls and 63 red ones. If you pick one after another, noting the colour and putting it back, continually updating the ratio of black to red, how close to  $37/63$  could you expect to get as the number of draws increased? A totally useless question, but one of interest to fellow members of the Royal Society. I suppose a way of putting it is that de Moivre was interested in mathematics, whereas Simpson, insofar as he applied himself to the problem of reconciling different astronomical observations, was dealing with physics or science.

#### 4. Bayes' Essay

Where does Bayes (1764) come into this? He himself gives no clue. Stigler says that Bayes' essay "is extremely difficult to read - even when we know what to look for"! Karl Pearson is quoted as saying that "Bayes' actual method of approaching the problem is obscure". Bayes makes no attempt to explain what he is doing. What he seems to have done is to solve an extremely involved and ingenious, but totally specific, idealised, and determined gaming problem in the de Moivre style, and then written down in three lines a totally unspecific problem apparently in the real unknown world, and finally provided a muddled and, I would say, erroneous verbal bridge which he seems to imply relates his totally specific solved problem to his totally unspecific real world problem. Needless to say, his determined problem, although hideously involved, is of the "Bayesian" type mentioned above, and in dealing with it, Bayes uses the commonsense arithmetical procedures of addition and division I have set down above. (Note that "multiplication" is missing. This is because Bayes, true to the gaming tradition, and as Laplace would also do 13 years after his death, had equiprobable "prior" events. His version of "Bayes' Theorem" is  $P(B_i|A) = P(A|B_i)/\sum_j P(A|B_j)$  ).

Bayes passes over this line in his working without comment. Like me, he regarded it as just a line of arithmetic.<sup>note 1</sup> But it is this which has been pulled out, dressed up, and presented as "Bayes' Theorem".

Savage (1960), an early promoter of Bayes, essentially agrees with this when he says, "what Price says [in his preface to Bayes' essay] convinced me that Bayes was aware of 'Bayes' Theorem' in full generality". This acknowledges that Bayes did not himself indicate that he was aware of his own theorem. However, my position as stated above, is in agreement with the sense of Savage. Bayes was perfectly capable of doing the type of arithmetical operations now called "Bayes Theorem". Where I part company is in not finding this worthy of remark. It is just arithmetic.

Bayes' specific problem is presented by him in terms of rolling two balls on a "levelled table". However, this is only a device for producing random numbers, so I will follow Savage (1960) in replacing the table with a generator of random numbers between 0 and 1.

Imagine that a random number  $x$  is drawn ("random" gives the gaming characteristic of equiprobability). A further 100 (Bayes says  $n$ , but I will concretise somewhat) draws are made, and  $X$  of those are less than  $x$ . The expectation value of  $X$  is, of course  $100x$ , i.e., if  $x=0.65432\dots$ ,  $X$  would be expected to be around 65 out of 100. Bayes now turns the picture on its head. If the sequence 1-draw-followed-by-100-draws is repeated again and again, and the result noted only if  $X$  has one pre-fixed value, say 65, thus collecting those values of  $x$  associated with  $X=65$ , what would the distribution of  $x$  be? This picture is analogous to my water channels, even if with difficulty. The  $x$ 's are drawn and sent down one of 100 channels depending on their size, i.e., if  $x=0.35432\dots$  it is sent down the 36th channel. If the corresponding  $X$  value is 65, the  $x$  is somehow tagged. In "Bayesian" language, which Bayes does not use, the probability of  $0.35 < x < 0.36$ , given that  $X=65$ ,  $P(0.35 < x < 0.36 | X=65)$ , is the tagged count in the 36th channel divided by the sum of tagged counts in all channels.

That

sentence again expresses "Bayes' Theorem". Noting this for all channels gives the probability distribution of  $x$ , given that  $X=65$ .

It was mentioned above, in a parenthesis, that Bayes assumed equiprobability for  $x$ , on the basis of the construction of the table and the balls. Since modern Bayesians seem to feel that the name of Bayes gives respectability to subjective probabilities, this point is worth stressing. All values of  $x$  in this problem are *by definition* equally likely. There is nothing subjective here. The frequency distribution of *those*  $x$ 's which are associated with *one* specific later value of  $X$  is of course non-constant, peaking in fact at  $x=X$ .<sup>note 2</sup>

What is the import or importance of this problem, it might well be asked? And this seems to have been roughly the reaction for around 200 years, although knowledge of Bayes was by no means absent (Stigler says that Bayes work was known to specialists from 1780 onwards).

Bayes too seems to have wondered what the significance was. He decided to make his solution to this wholly specified problem the answer to a quite general one, namely:

*"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 two degrees of probability that can be named."*

He states this unspecific problem right at the beginning of his essay. With no further elucidation, no explanation of why this problem may be of interest<sup>note 3</sup>, these three lines are followed immediately by several pages devoted to what Savage calls a "whole short course on probability". Then follows his solution, as paraphrased above, of the specific problem involving the levelled table and the rolling balls.

A section headed "Scholium" then seems to attempt to show that the latter is equivalent to the stated unspecific problem, in involved but unemphatic terms. He never says at any point, "QED".

His reasoning seems to me absurd, and Stigler reports Karl Pearson and Ronald Fisher as having more or less the same, if less emphatically expressed, opinion. Fisher thought that this was the reason Bayes did not publish his essay, and that seems to me more than likely. Stigler, however, thinks that Pearson and Fisher are mistaken.

The argument of the Scholium (which is quite short) seems to be as follows.

In the fully specified case, says Bayes, "before  $[x]$  is discovered ... I can have no reason to think it [the number of 'successful' draws,  $X$ ] should rather happen one possible number of times than another".

For the unspecific case, involving "an event concerning the probability of which we absolutely know nothing antecedently", Bayes says that, "concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another".

Since in both cases, what he has "no reason to think" can be expressed in identical words, he concludes:

"In what follows therefore, I shall take for granted that the rule [the equiprobability of  $x$ ] given concerning [the specified case] is also the rule to be used in relation to any event concerning the probability of which nothing at all is known antecedently to any trials made or observed concerning it. And such an event I shall call an unknown event."

He disregards the fact that what he (quite correctly) has "no reason to think" flows in the first case from *his* specification of equiprobability, but in the second case from ignorance. His choice of words "no reason to think" cloaks the fact that in the first case, he has every reason to think, indeed to know, since he defined the levelled table and balls in that way, that all numbers are equally probable, whereas in the second, he not only has no reason to think this or that, he has in fact no reason to think anything at all. Bayes is implying that a man waiting for a number to come up from a machine known to him to be a random number generator is in exactly the same state of ignorance as a man waiting for a number to come up via a totally unknown process.

It is convenient to interpolate here additional evidence that both Price and Bayes saw "Bayes' Theorem" as just arithmetic. Price, in his introduction to the Essay, reported Bayes as thinking that the "rule" of equiprobability referred to above was the *only* obstacle in his way. "It appeared to him", he says, "that the rule must be to suppose the chance the same that it should lie between any two equidifferent degrees; which, if it were allowed, *all the rest might be easily calculated in the common method of proceeding in the doctrine of chances* [my italics]".

Perhaps the most telling demonstration of the wrongness of Bayes' "rule" (the postulate of equiprobability for unknown events) is the following.

Suppose Bayes was confronted with a real table, and real balls, and a real person rolling them. This would still be better than his claimed position of knowing absolutely nothing antecedently, since he would at least know that it was a visibly "level" table, and the balls as smooth and round as anyone could reasonably expect. But his calculations for the idealised table and balls would still not apply to the real situation, although this is exactly his implication. They would apply to *some* extent, to the extent that the real situation of table, balls, and person rolling, approached the ideal definition. The gap between reality and ideal, however, is unknown, and certainly unquantifiable. Bayes could certainly say, "The table looks level, the balls look round and smooth, the ball-roller looks capable and honest, so, for want of anything better, I'll *make a working assumption* that my calculation for the idealised situation applies". However, this is not what he says. He says that his calculation for the case where he knows everything antecedently *does apply without reservation* in any arbitrary situation of recurrent yes/no events "of which we know absolutely nothing antecedently". If this was true, we would have the paradox that his result applied if he was confronted with a situation of which he knew absolutely everything, or absolutely nothing, but not to a situation where he knows *nearly* everything!

In my view, this is simply a muddle.

A wider aspect of the muddle is the disjointedness of the essay, perhaps not surprising in a document which the author did not prepare for publication. There is no textual bridge between "the stated problem" and the "levelled table problem" (the "Scholium" comes later), no expansion or discussion of why each problem is being treated. We have no inkling<sup>note 4</sup> as to why Bayes lit on the type of problem which is now associated with his name, namely one which enquires, "given that this observed event may be associated with several alternative provenances, what is the probability that it is associated with any one given provenance?"

The point to note here is that Bayes' work did not *in fact* lead anywhere. The curiosity is that his problem was of a sort, as will be alluded to later, that crops up in current science, particularly the soft sciences. But, given the problem, the solution is simple arithmetic, and can be tackled without reference to Bayes. Bayes' fame is a vector pointing from the present to 1764, not the other way round.

Stigler makes me aware that the above is inevitably unfair to Bayes, not I think in its accuracy, but in its brevity. Laplace clearly did lead somewhere, yet his initial work was similar to that of Bayes. But he had the luck or the good sense to treat a physically undetermined, hence inferential problem from the start. If an urn contains only red and black balls of *unknown* number, and a ball is drawn at random, its colour noted, and replaced, and the ratio of red to black is  $X$  after  $n$  draws, what is the probability of the real ratio being  $x$ ? Bayes' problem was a fully *determined* problem. He knew everything there was to know about the levelled table and balls, just as Bernoulli and de Moivre knew everything about their coins and dice games. Like them, he calculated the play of chance on the outcomes. Laplace, on the other hand, was explicitly concerned with a problem of inference, namely, given the ephemeral and chancy value of  $X$ , what is the hidden but concrete truth,  $x$ ? The difference is not totally real. Laplace coupled a much better mathematical ability with a shrewd (or unwitting) silence on his assumption, like Bayes, of the equiprobability of  $x$ . The problem which agonised Bayes was ignored by Laplace. Where Bayes came to an explicit and wrong conclusion, Laplace skated over it. Thus we are at liberty to suppose that he regarded the implication of equiprobability as merely a reasonable working assumption, *faute de mieux*, and not as a proven statement of truth<sup>note 5</sup>. He was able, unlike Bayes, to achieve interesting numerical results. Replacing red and black balls with male and female children, he was able to infer that a real ratio of 1:1 for male and female births in Paris was highly unlikely, given the observed ratio. He was thus on the main historical experimental inference path leading 150 years later to Gosset and Fisher.

I sum up Bayes' achievement as follows.

1. He did a calculation similar in principle to those done on well specified equiprobable gaming systems such as coins, dice, cards, lottery draws, by predecessors such as Bernoulli and de Moivre. The detail of the calculation involved the same, essentially trivial, arithmetical procedures that are today associated with "Bayesian" calculations, as illustrated above, but that is about all that Bayes' work has in common with that term.

2. Bayes' problem (the one he calculated, not the one he verbally claimed to have solved) did not deal with the following aspects of the notions now associated with his name.

2a. The original Bayes' problem was not framed in terms of a prior event or cause whose probability of happening (or having happened) is calculated, *given* the happening of a posterior event or effect. His prior entity was not *an event*, but "the probability that the probability of [the prior event] happening in a single trial lies somewhere between two degrees of probability that can be named".

2b. What is now called "Bayes' Theorem" is a misnomer. It is not a theorem, and Bayes' made no claim to it. It simply refers to the use of basic and/or operations in the arithmetic of probability. Bayes' innovation was not in this so-called theorem, but perhaps in conceiving the ingenious problem which called for those operations.

2c. Bayes contributed nothing to the "problem of inference". The problem he calculated was a fully specified one, in the sense that gaming problems involving coins or dice are fully specified. The problem he *announced* at the outset of his essay was framed as an inferential one, that is, one in which only the outcomes are known and the properties of the underlying system are to be inferred from them, but there is no rational connection between his calculation and the initially stated problem, only a muddled and erroneous verbal assertion of a connection. Bayes did not publish his essay. It is quite possible, or likely, that his own opinion of the soundness of this assertion was not different from the one expressed here.

2d. The aspect of the weighing of competing *causes* of observed happenings, often associated with "Bayesian" procedures was not present in the case Bayes treated. Nor was the idea of "updating" old estimates by the amalgamation of new information.

2e. Perhaps the most "mystical" notion associated with modern "Bayesian" statistics is the role of subjective data - the quantification of hunches, convictions, beliefs, or feelings. There is absolutely nothing to support this in Bayes' essay. He asserted that "the rule" for dealing with equiprobable events, "is also the rule to be used in relation to any event concerning the probability of which nothing at all is known antecedently to any trials made or observed concerning it", but he appeared to put this (in any case quite wrong) view forward as a correct and factually justified one, not as a personal guess or belief, or as an approximately correct working assumption.

## 5. An example of a modern "Bayesian" presentation

A typical modern presentation of the use of the "Bayesian method" is taken from an article on "Genetics and heredity" in the Encyclopaedia Britannica, 15th edition (1989).

"For a variety of reasons, the parental genotypes frequently are not clear and must be approximated from the available family data. The Bayes theorem, a statistical method first devised by the English clergyman-scientist Thomas Bayes in 1763, is useful in those cases, most often when autosomal dominant or sex-linked recessive traits are

involved. The method assesses the relative probability of two alternative possibilities (e.g., whether a consultand is or is not a carrier). The likelihood derived from appropriate Mendelian law (prior probability) is combined with any additional information that can be obtained from the consultand's family history (conditional probability). A joint probability is then determined for each alternative by multiplying the prior probability by all conditional probabilities. By dividing the joint probability of each alternative by the sum of both joint probabilities, the posterior probability is arrived at. The latter is the likelihood that the individual whose genotype is uncertain either carries the mutant gene or does not. An example of this method for a sex-linked recessively inherited disease (muscular dystrophy) is given below.

"The consultand [the person wishing to have genetic counselling] wishes to know her risk of having an affected child. It is known that the consultand's grandmother is a carrier, since she had two affected sons. What is uncertain is whether the consultand's mother is also a carrier. The Bayesian method for calculating the risk is as follows:

"Likelihood that consultand's mother a carrier	is	is not
Prior probability	1/2	1/2
Conditional probability	1/4 (for each of her sons there was a 1/2 chance of being unaffected	1
Joint probability	$1/2 \times 1/4 = 1/8$	$1 \times 1/2 = 1/2$
Posterior probability	$1/8(1/8 + 1/2)$ =1/5	$1/2(1/8 + 1/2)$ =4/5

"If the mother of the consultand is a carrier (risk=1/5) there is then a 1/2 chance that the consultand is a carrier, so her total empiric risk is  $1/5 \times 1/2 = 1/10$ . If she has a child, there is a 1/2 chance that it will be male and a 1/2 chance that the male will be affected. Hence the total empiric risk of the consultand having an affected child is  $1/10 \times 1/2 \times 1/2 = 1/40$ ."

It is natural, of course that users of the Bayes vocabulary will not necessarily have any close knowledge of Bayes, but it is clear that the author is unaware that Bayes had been dead for two years in 1763, and did not "devise" his "theorem"<sup>note 6</sup>. The use of the vocabulary - theorem, prior probability, additional information, conditional probability, joint probability, posterior probability - burdens rather than clarifies the text, and the effort to skew the presentation in order to accommodate these terms makes the train of thought less, not more, comprehensible.

What are the bare bones of this calculation?

1. The grandmother of the "consultand" certainly carried the gene.
2. The mother of the consultand was born with a 1/2 chance of carrying the gene, but by having two unaffected sons (the consultand's brothers), she reduced this chance.

By

how much? Imagine a whole population of women with her history (i.e., with a gene-carrying mother). Half would be free of the gene, half would have it.

3. If they all had one son, all of the sons of the first half, and half of the sons of the second half, would be unaffected. The remaining quarter would have affected sons, and, being now known gene-carriers, would drop out of the statistics.
4. If they all had a second son, again all the sons of the first half, and half the sons of the remaining quarter, would be unaffected. A further one eighth would join the pool of known gene-carriers.
5. The consultand's mother is one of the "surviving" half plus one eighth, but since she now belongs to a population with only one out of five eighths with the gene (the other three eighths having had at least one affected child), her risk of having the gene is now reduced from  $1/2$  to  $1/5$ . If this story were to be carried on, it would become clear that if the consultand had  $n$  healthy brothers, her mother's chance of carrying the gene would be reduced to  $1/(2^{n+1})$
6. As in the Encyclopaedia Britannica article, the calculation concludes with the consultand having a  $1/5 \times 1/2 \times 1/2 \times 1/2$  chance of having the gene, and having a male rather than a female child, and the son having MD - a risk for her of  $1/40$ .

The above presentation, while not being one that "a reasonably numerate schoolchild could do" (for conceptual, not computational, reasons), certainly uses only simple arithmetic, and appeals only to the basic rules of probability. Being devoid of jargon, it is in my view easier to follow.

However, my main point is not that, but the fact that it could be done if neither the historical Bayes nor the modern Bayes industry had ever existed. Even if the tabular calculation of the Encyclopaedia Britannica article is preferred, its comprehensibility would hardly be impaired if the entire introductory Bayesian paragraph were omitted. Its inclusion puts the reader to the task of refreshing his memory on what the "Bayes' Theorem" is, and then working out that the "event"  $A$  in the algebraic expression is, in this example, "having two healthy brothers". The mental gymnastics involved are not trivial, and they are quite unnecessary. With only a few words more explanation, the bare tabulation of the calculation would be sufficient.

## 6. Other "Bayesian" examples

Freund (1974), in addition to treating "Bayes' Theorem", has entries in his index under "Bayesian analysis", "Bayesian estimate", and "Bayesian inference". The term "Bayesian analysis" appears to apply when, for alternative business decisions (e.g., go on, stop), the numerical payoffs for alternative outcomes (e.g., new drug does, does not succeed) are weighted with known or assumed or guessed or assessed probabilities, thus giving "expected" or probability-weighted payoffs for each of the two decisions.

The terms "Bayesian estimate" or "Bayesian inference" apply when a real estimate of a mean and standard deviation is amalgamated with guessed values, using the normal rules. They are also applied to a related procedure, illustrated with a calculation showing how the "feelings" of three company executives of varying (guessed and quantified) reliabilities are tested against a real result to give an updated reliabilities, and hence a new weighted result. Thus, instead of throwing away their pre-conceived notions once they have real data, the executives merely "update" their notions to some intermediate value<sup>note 7</sup>.

In game theory (e.g., see Fudenberg and Tirole, 1991), if a Nash equilibrium cannot be obtained because of "incomplete information" in the payoff matrix, the game is turned into one of "complete" but "imperfect" information by constructing several payoff matrices, each having a certain subjective probability assigned to it. A Nash equilibrium can then be obtained, its provenance being indicated by calling it a "*Bayesian* Nash equilibrium"<sup>note 8</sup>.

The only traceable connection of these uses of the word "Bayesian" with the historical Bayes is the quite erroneous one of guessed probabilities, and in some cases the trivial use of "Bayes' Theorem" arithmetic<sup>note 9</sup>.

## 7. What does the modern word "Bayesian" mean?

If "Bayesian inference ... [is] ... one of the most widely known eponyms in all of science", it is through unnecessary references of the sort illustrated above.

Do such references have any succinct common theme? Given that the modern "Bayesian" has little to do with the historical Bayes, is it possible to discern any feature which might serve as a dictionary definition of the word? The vocabulary is used in such diverse cases that succinctness may be beyond reach, but there appear to be three ingredients which are present in varying degrees, including zero, i.e., none is of the essence. These are:

### **Arithmetical**

Any arithmetical manipulation corresponding to the "Bayes' Theorem":

$$P(B_i|A) = P(A|B_i)P(B_i)/\sum_i\{P(B_i)P(A|B_i)\}$$

is liable to evoke the whole list of Bayesian words and phrases. The  $P(B_i)$  may have a wholly subjective, partly subjective, random, equal, or purely factual character. The  $B_i$  may be "events" as in the modern canonical Bayesian story, or they may not, as in the two cases actually treated by Bayes. There may be "inference" or, as in Bayes' own worked out case, not. The common feature, in those cases, is not the various allegedly Bayesian characteristics but the narrowing of the sample space to the event A. In the above example from genetics, event A is the existence of two healthy sons, excluding all other combinations of the words "healthy", "unhealthy", "sons", and "daughters", and all numbers other than two. The idea of "provenance" is certainly essential, as is the idea of "several" provenances. All applications of the arithmetic of "Bayes' Theorem" involve process, i.e., ordering in time, and some more or less

hidden

(probabilistic) branching of the process. (My abstractions are almost as opaque as the "Bayesian" ones - my main point is that it is better and much easier simply to *do* the simple arithmetic. The abstractions serve no purpose.)

### **Probability weighting**

A scientist or engineer usually deals in near-certainty. Even if probability is in the picture, it is usually fairly accurately computable probability. Life, and the social and life sciences, are too complex for that. Risk is not marginal but the norm. To act after weighing and balancing risks is the common, almost daily, experience. Occasions when this process can usefully be helped by arithmetic are bound to occur, hence the weighting of numerically calculated alternative payoffs or outcomes with calculated or guessed relative probabilities. For some reason this attracts a "Bayesian" label, but there is no discernable connection with either "Bayes' Theorem" or with Thomas Bayes. This is even simpler arithmetic than that involved in "Bayes' Theorem", and has none of the conceptual complexity which is sometimes present in the situations in which the "Bayes' Theorem" is applied.

### **Subjectiveness**

Subjectiveness is apparently an attractive strand of modern Bayesianism. It is evidently not the *essence* of Bayesianism. In the application of "Bayes' Theorem" to genetic counselling cited above, the various probabilities, and in particular the "prior" ones, are strictly objective or factual. Probability weighting too might well involve accurately known numbers. But very many examples of Bayesianism do incorporate probabilities derived from opinions, feelings, hunches, or simply (as in Bayes) equiprobabilities standing in for complete ignorance. In the life and social sciences, this is likely to be the normal case, since the basis for accurate calculation is simply not there. In the case of the "Bayesian estimate" given by Freund (see above), involving the quite standard arithmetic of amalgamating two estimates of mean and standard deviation, there is no feature which might qualify the procedure as "Bayesian" *apart from* the fact that one set of mean and standard deviation is a "hunch". But, as I hope has been made clear, there is absolutely no basis on which to associate Bayes with hunches, guesses, feelings, etc., at least, not any more than Laplace or any other theorist of probability.

So evocations of Bayes (B) seem to span a spectrum which might be written

$$B = ((BT \text{ or } PW) \text{ and/or } S)$$

where BT and PW are the arithmetic of "Bayes' Theorem" and Probability Weighting respectively, and S is the Subjectivity of numerical inputs, which latter may be the sole ingredient or a sauce added to other ingredients. Not only is there no essential link to the historical Thomas Bayes at any part of this spectrum, there is no coherent entity which it might be of practical use to give any one name to.

## **8. Why "Bayesianism"?**

As conjectured above, it seems likely that the modern resurrection of Bayes has accompanied the expansion of statistical methods into medical, biological and social science research, including economics and business studies. But why specifically Bayes - given that "Bayesianism" is not *necessary*, and indeed is a time-wasting

impediment? After all, we are not talking about quantum mechanics here, or even least

squares, or Gosset's t-test. We are talking about the most elementary arithmetical rules for combining probabilities.

Did the invocation of the name of Bayes start as an elaborate way of avoiding writing "let us assume", or as an obscure and portentous-sounding cloak for shaky statistics? Is it just famous for being famous? Is it in text books not because it needs to be but because it cannot be left out? I simply do not know. It is deeply puzzling.

I hope I have made it clear that I am not necessarily criticising the *work* which goes with the vocabulary. I have no objection to guessed probabilities and shaky statistics, if better probabilities or statistics are not available, and if the guesswork and shakiness is made explicit. But I think there is more than a hint that the vocabulary is sometimes used to convey the unspoken message that the mysterious Bayes somehow blessed and made OK the use of guesses and subjective estimates, so that caveats and excuses are not required, or to convey the impression that something clever is afoot, when it is not.

## 9. Conclusion

What I contend is that this Bayesian "eponym", unlike those associated with Newton, Hooke, Boyle, Heisenberg, Einstein, Bohr, etc., is obscure, pretentious, and serves no useful purpose. If somehow it could be eliminated, it would save everybody's time.

## Acknowledgement

Nearly everything factual in this note concerning Thomas Bayes and his place in the history of statistics is derived from the excellent work of Stigler (1986).

## Notes

1. Bayes' presentation is in terms of a sketched figure, and the "arithmetic" is done in terms of areas. Bayes' "proposition 9" is most relevant. Here, he is doing no more than, in Price's words, "calculat[ing] in the common method of proceeding in the doctrine of chances".
2. It is typical of the confusion which surrounds the Bayes myth that this simple and seemingly tautological remark is elevated to the status of the "Bayes-Laplace theorem" in one reputable reference. In the Encyclopaedia Britannica (1989), in the entry on "the history of mathematics", one reads: "In 1763 Thomas Bayes proved that, if  $m:n$  is the relative frequency of an event on  $n$  independent occasions, then  $m:n$  is also the most probable value of the event's probability, provided that any value of this probability is initially (a priori) as probable as any other value. The same theorem was proved independently by Laplace in 1774. ... The Bayes-Laplace theorem is the inversion of Bernoulli's theorem ... ". Apart from the fact that the "theorem" referred to is quite different from the slightly less tautological one normally referred to as "Bayes' (or Bayes-Laplace) Theorem", this short passage is very free with inaccurate history. For example, Stigler is categorical that Laplace in 1774 *stated* his version of "Bayes' Theorem" (the one normally known by this name), but did not prove it then. Proof is in any case trivial.
3. Price, in his introduction to Bayes' Essay, is more coherent. After outlining the work of Bernoulli and de Moivre on the fully determined gaming problem, he says: "But I know of no person who has shown how to deduce the solution to the converse problem to this; namely, 'the number of times an

unknown event has happened and failed being given, to find the chance that the probability of its

happening should lie somewhere between any two named degrees of probability." Price is here quoting Bayes more or less verbatim, but, so far as I am aware, Bayes himself does not indicate explicitly that he is addressing this converse problem (unless merely stating it can be called addressing it), and certainly, the problem he addressed in detail was not this converse problem. Laplace 10 years later did explicitly address the converse problem.

4. See footnote 2.

5. Seal (1978) quotes Laplace, four years later (1778), as writing "one must suppose" all (unknown) possibilities to be equally likely. What else could one rationally "suppose" if one really knew absolutely nothing? This is rather different from Bayes' comparatively lengthy argument, concluding "therefore I shall take for granted". However that may be, note that Laplace, and Bayes more insistently, exclude the use of a "Bayesian" hunch in these circumstances.

6. This is probably a case of a wrong message being passed on, getting firmer as it goes. At random I pick up Maddala's "Introduction to Econometrics", 1992. He says, "... Bayes' theorem ... appeared in a text published in 1763 by Reverend Thomas Bayes, a part-time mathematician".

7. Curiously, the procedure embodied in this example is the basis of "Bayesian statistics", now the most vigorous arm of the academic Bayes industry. A check on an extensive library catalogue shows that books with "Bayes" or "Bayesian" in the title, virtually all dealing with "Bayesian statistics", number (in this catalogue) respectively 0, 2, 10, 24, and 30 for pre-1960, the 1960s, 1970s, 1980s, and 1990s. At or near the beginning of this movement was Savage (1954). Oddly, in the light of hindsight, Savage himself does not mention Bayes the man, and makes only a passing reference to "Bayes' Theorem", which he described as "easy to prove ... trivial to derive". He preferred to name his approach "personalist statistics". The use of the arithmetic of "Bayes' Theorem" is all-pervasive in this approach, since it insists that probability is in the mind of the observer (cf. the "feelings" of Freund's company executives) and not *only* in the data thrown up by the process being observed. These "prior" feelings are updated by the observed data to "posterior" values, by means of the "Theorem". So perhaps it was inevitable that Savage's followers would lazily and happily fall to calling themselves "Bayesians", and then to labelling those who deal with the data and nothing but the data, such as Ronald Fisher and every "working" scientist and engineer, as "non-Bayesians". The appeal is extended by claiming, quite plausibly but irrelevantly, that all human learning, including all scientific development, follows the hunch/updating model. I have checked that the 9th (1997) edition of Freund leaves its 1974 treatment of Bayes virtually unchanged, which reinforces the impression that "Bayesian inference" or "Bayesian statistics" is more of a sustained polemic (idealism versus realism) than a branch of practical data evaluation. F. P. Ramsey (1926), in what the DNB calls "the classic paper that laid the foundations for modern subjective interpretations of probability", makes the perceptive remark that "it [is] likely that the two schools are really discussing different things".

8. In connection with game theory, however, it may be noted that Dimand & Dimand (1997), a 3-volume compilation of 122 papers relating to the foundations of game theory "from its beginnings until 1960", while liberally strewn with references to Cournot, Bertrand, Edgeworth, Nash, Morgenstern, etc., etc., finds no occasion to refer to Bayes in an index of around 500 names. The nearest it comes, I think, is a single sentence in Vol. I, p480: "The point of view of subjective probability is set forth by Savage [The foundations of statistics, 1954 ]".

9. After completing this paper, I found that Binmore (1990) expressed a similar opinion in the context of game theory. He says, "the word 'Bayesian' itself represents an unwelcome distraction". And he adds (Binmore 1994): "I wish it were possible to persuade people to call Bayesian decision theory something else". Later he calls Bayes' rule a "trivial algebraic manipulation", although still repeating the received formula that it "was discovered by the Reverend Thomas Bayes some time before 1763". Binmore favours Savage as an "eponym"

## Bibliography

- Barnard, G. A., 1958, Thomas Bayes - a biographical note, re-printed in Pearson & Kendall, and in Press
- Bayes, Thomas, 1743, An essay towards solving a problem in the doctrine of chances, Philosophical Transactions of the Royal Society of London, re-printed in Pearson & Kendall and in Press
- Binmore, Ken, 1990, Essays on the foundations of game theory
- Binmore, Ken, 1994, Game theory and the social contract
- Dale, Andrew I., 1991, A history of inverse probability from Thomas Bayes to Karl Pearson
- Dimand, Mary Ann & Dimand, Robert W. (eds), 1997, The foundations of game theory
- Encyclopaedia Britannica 15th ed., 1989, "Genetics and heredity"
- Freund, John E., 1952 to 1997 (9 editions), Modern Elementary Statistics
- Fudenberg, Drew & Tirole, Jean, 1991, Game theory
- Pearson, Egon S. & Kendall, M. G. (eds), 1970, Studies in the history of statistics and probability
- Press S. J., 1989, Bayesian statistics
- Ramsey, F. P., 1926, Truth and Probability, re-printed in Polson & Tiao, 1995, "Bayesian Inference"
- Savage, L. J., 1960, Reading note on Bayes' theorem, re-printed in Dale
- Savage, L. J., 1954 & 1972 (2 editions), The foundations of statistics
- Seal, H. L., 1978, Bayes, Thomas, re-printed in Press
- Stigler, Stephen M., 1986, The history of statistics before 1900