5 Bayes

Texts

Stigler, Stephen(1986) *The History of Statistics : The Measurement of Uncertainty Before 1900*
Bayes, Thomas(1763) *An essay towards solving a problem in the doctrine of chances*
Philosophical Transactions of the Royal Society of London 53,370-418.
Reprinted in Biometrika,45,293-315.
www.Britannica.com

1 Bayes himself

Thomas Bayes(c1701,1761) was a Nonconformist minister in Tunbridge Wells. He was a Fellow of the Royal Society, though he published no mathematical works in his lifetime. The essay for which he is famous was read before the Royal Society in 1763 by Richard Price; Price contributed an introduction and an appendix to the published paper. Price refers to an introduction prepared by Bayes, and interpolates his own remarks in the paper, so that it is sometimes hard to tell if comments are due to Price or Bayes. The paper and ideas had little impact until they were rediscovered by Laplace in 1774-1781.

There is a fair amount of doubt whether this is really Bayes's picture.. the dress is supposed to be anachronistic. See Bellhouse and Stigler for discussion.

2 Definition of probability

Bayes's definition of probability:

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.

By chance I mean the same as probability.

The sloganized version is 'probability is the amount you bet over the amount you get'. Bayes, being a minister, would not have wanted to use gambling terms. Stigler states that Bayes's *' acceptance of subjectively determined probabilities seems complete and unambiguous*'. I do not see any subjective element in the definition: the value at which an expectation depending on the happening of the event *ought* to be computed. When he says ought he is saying this is the way the probability should be, not allowing for personal opinion.

Bayes is following the traditional route of defining probability in terms of some other related concept. Bernoulli uses *limiting frequencies*. Laplace uses *possibilities*. Bayes uses *expectations*. You could as well define expectations in terms of probabilities (although all good people know that the better starting point is expectation). Either way you do it, it is just a mathematical exercise in going from one concept to the other, with nothing expressed about the underlying nature of the concept, or how the concept will result in practical calculations. We could define frequentist expectation as the limiting observed average value of some variable under repeated , similar,(we do not want to say independent here just yet, it might hint at circularity) experiments. Then expectation is objectively defined, and so is the derived concept probability.

Mr Price in his introduction to Bayes essay(he did not let us see Bayes's own introduction):

Mr Bayes has thought fit to begin his work with a brief demonstration of the general laws of chance. His reason for doing this, as he says in his introduction, was not merely that his reader might not have the trouble of searching elsewhere for the principles on which he has argued, but because he did not know whither to refer him for a clear demonstration of them . He has also made an apology for the peculiar definition he has given of the word chance or probability. His design herein was to end all dispute about the meaning of the word, which in common language is used in different senses by persons of different opinions, and according *as it is applied to past or future facts. But whatever different senses it may have, all (he observes) will allow that an expectation depending on the truth of any past fact, or the happening of any future event, ought to be estimated so much the more valuable as the fact is more likely to be true, or the event more likely to happen. Instead therefore, of the proper sense of the word probability, he has given that which all will allow to be its proper measure in every case where the word is used.*

Of course people have been squabbling about probability ever since, without disagreeing with the connection between expectation and probability. To my mind it is a mere mathematical connection with no philosophical consequences, just as Bernoulli's theorem is.

3 Bayes's Problem

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

By chance I mean the same as probability, says Bayes, but you might ask, why use two words for the same concept then? The answer is that his statement of the problem will look mighty awkward if he is computing the probability that the probability of its happening in a single trial lies somewhere between any two degrees of probability. There are two kinds of probability here, chance is *epistemic* probability, the other probabilities are *frequency*(aleatory, Hacking calls them) probabilities.
The modern usage is to call frequency probabilities chances, arising in games of chance, and call probabilities epistemic.

You can have unknown chances, the coin comes up heads with unknown probability p to be estimated. You cannot have unknown epistemic probabilities since these probabilities are supposed to express your ignorance. If you contemplate a toss of the coin, and ask what its epistemic probability is of coming up heads; you might reason its long run frequency is the unknown p, which is (epistemically) uniformly distributed. The expected value of p is 1/2; so the probability of heads on the next toss is 1/2. As data accumulates, the probability of heads changes, but it is always specified.

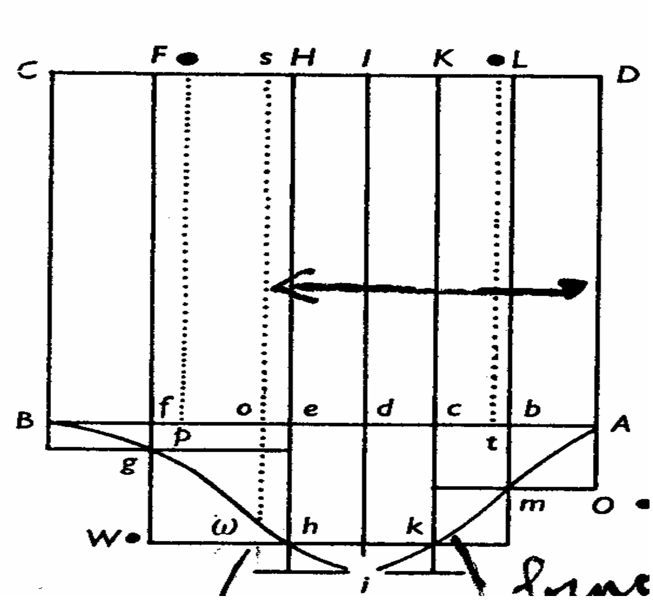
You can talk about unknown chances to be estimated, but you cannot talk about unknown epistemic probabilities to be estimated.

4 Bayes table

Bayes assumes an independent sequence of Bernoulli trials:

Prop. 7
If the probability of an event be a, and that of its failure be b in each single trial, the probability of its happening p times, and failing q times in p + q trials is Ea^pb^q *if E be the coefficient of the term in which occurs* a^pb^q *when the binomial* $(a + b)^{p+q}$ *is expanded.*

Then he introduces his famous table:



POSTULATE. I. I suppose the square table or plane ABCD to be so made and levelled, that if either of the balls o or W be thrown upon it, there shall be the same probability that it rests upon anyone equal part of the plane as another, and that it must necessarily rest somewhere upon it.
2. I suppose that the ball W shall be first thrown, and through the point where it rests a line os shall be drawn parallel to AD, and meeting on and AB in s and o; and that afterwards the ball o shall be thrown p + q or n times, and that its resting between AD and os after a single throw be called the happening of the event M in a single trial. These things supposed :

LEM. I. The probability that the point o will fall between any two points in the line AB is the ratio of the distance between the two points to the whole line AB.

LEM. 2. The ball W having been thrown, and the line os drawn, the probability of the event M in a single trial is the ratio of Ao to AB.

Prop. 8

If upon BA you erect the figure BghikmA whose property is this, that (the base BA being divided into any two parts, as Ab, and Bb and at the point of division b a perpendicular being erected and terminated by the figure in m; and y, x, r representing respectively the ratio of bm, Ab, and Bb to AB, and E being the coefficient of the term $a^p b^q$ *when the binomial* $(a + b)^{p+q}$ *is expanded* $y = Ex^p r^q$*. I say that before the ball W is thrown, the probability the point o should fall between f and b, any two points named in the line AB, and with allthat the event M should happen p times and fail q in p + q trials, is the ratio of fghikmb, the part of the figure BghikmA intercepted between the perpendiculars fg , bm raised upon the line AB, to CA the square upon AB.*

Prop. 9

If before anything is discovered concerning the place of the point 0, it should appear that the event M had happened p times and failed q in p + q trials, and from hence I guess that the point 0 lies between any two points in the line AB, as I and b, and consequently that the probability of the event M in a single trial was somewhere between the ratio of Ab to AB and that of Alto AB: the probability I am in the right is the ratio of that part of the figure AiB described as before which is intercepted between perpendiculars erected upon AB at the points I and b, to the whole figure AiB.

5 Bayes Scholium for unknown probabilities

Scholium

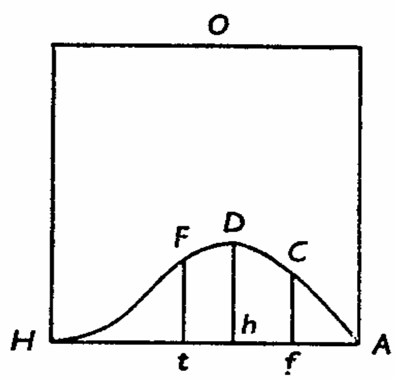
From the preceding proposition it is plain, that in the case of such an event as I there call M , from the number of times it happens and fails in a certain number of trials, without knowing anything more concerning it, one may give a guess whereabouts it's probability is, and, by the usual methods computing the magnitudes of the areas there mentioned, see the chance that the guess is right. And that the same rule is the proper one to be used in the case of an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it, seems to appear from the following consideration; viz. that I concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen anyone possible number of times than another. For, on this account, I may justly reason concerning it as if its probability had been at first unfixed, and then determined in such a manner as to give me no reason to think that, in a certain number of trials, it should rather happen anyone possible number of times than another. But this is exactly the case of the event M. For before the ball W is thrown, which determines it's probability in a single trial (by cor. prop. 8), the probability it has to happen p times and fail q in p +q or n trials is the ratio of AiB to OA, which ratio is the same when p +q or n is given, whatever number p is; as will appear by computing the magnitude ofAiB by the method of fluxions. And consequently before the place of the point o is discovered or the number times the event M has happened in n trials, I can have no reason to think it should ratl happen one possible number of times than another .

In what follows therefore I shall take for granted that the rule given concerning the event M M in prop. 9 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.

COR. Hence, by supposing the ordinates in the figure AiB to be contracted in the ratio of E to one, which makes no alteration in the proportion of the parts of the figure intercepted between them, and applying what is said of the event M to an unknown event, we have the following proposition, which gives the rules for finding the probability of an event from the number of times it actually happens and fails.

Prop. 10

If a figure be described upon any base AH (Vid. Fig.) having for it's equation $y = x^p r^q$ *where y, x, r are respectively the ratios of an ordinate of the figure insisting on the basE right angles, of the segment of the base intercepted between theordinate and A the beginning of the base, and of the other segment of the base lying between the ordinate and the point H, to the base as their common consequent. I say then that if an unknown event has happened p times and failed q in p+q trials,and in the base AH taking any two points as f and t you erect D the ordinates fC, tF at right angles with it, the chance that the probability of the event lies somewhere between the ratio of Af to AH and that of At to AH, is the ratio of tFCf, that part of the before-described figure which is intercepted between the two ordinates, to ACFH the whole figure insisting on the base AH.*


O
F
D
C
H
h
A
t
f