Bayes' Theorem

An Expository Presentation *

By EDWARD C. MOLINA

BAYES' theorem made its appearance as the ninth proposition in an essay which occupies pages 370 to 418 of the Philosophical Transactions, Vol. 53, for 1763. An introductory letter written by Richard Price, "Theologian, Statistician, Actuary and Political Writer," [1] begins thus:

> "I now send you an essay which I have found amongst the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved."

A few lines farther on Price says:

> "In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times."

. . . .

> "Every judicious person will be sensible that the problem now mentioned is by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to a sure foundation for all our reasonings concerning past facts, and what is likely to be hereafter."

No one will dispute the importance ascribed to Bayes' problem by Price; in fact, a paper by Karl Pearson on an extension of Bayes' problem is entitled "The Fundamental Problem of Practical Statistics." Opinions differ, however, as to the validity and significance of the solution submitted in the essay for the problem in question. In view of this situation I shall limit myself today to an exposition of the fundamental characteristics of the problem Bayes' theorem deals with and shall give no consideration to its interesting applications.

The exposition may be outlined as follows: after specifying the class of problems to which Bayes' theorem pertains I shall:

* Read before the American Statistical Association during the meeting of the American Association for the Advancement of Science in Cleveland, Ohio, December, 1930.

[1] These titles are associated with the name of Price in the frontispiece portrait of him bound with the December, 1928, issue of *Biometrika.*

I. Discuss briefly two problems each of which will emphasize one of two kinds of *a priori* probabilities which should be constantly borne in mind when Bayes' theorem is under consideration,

II. Partially analyze a certain ball-drawing problem which will not only serve as an introduction to the algebra of Bayes' theorem but will later help to throw light on its significance,

III. Present Bayes' problem and the related theorem,

IV. Make some remarks on the value of the theorem and the controversies which it raised.

In carrying out this plan I shall find it convenient to ignore the historic order of events.

When probability is the subject under consideration one anticipates problems such as: A coin is about to be tossed 15 times; what is the probability that heads will turn up seven times? A sample of 100 screwdrivers is to be taken from a case containing 1000 screwdrivers of which 300 are known to be defective; what is the probability that the sample will contain 25 defectives?

These are direct, or *a priori*, probability problems. In each of them the nature of a game, or an experiment, is specified in advance and then a question is asked relating to one, or more, of the possible outcomes of the game or experiment. Problems of this type have occupied the attention of mathematicians since the days of Pascal and Fermat, the creators of the mathematical theory of probability.

An inverse class of problems of great practical significance, called *a posteriori* probability problems, came into prominence with the publication of Bayes' essay. In these we find specified the result or outcome of a game which has been played, whereas the question then asked is whether the game actually played was one or some other of several possible games. This type of problems is usually stated as follows:

> "An event has happened which must have arisen from some one of a given number of causes: required the probability of the existence of each of the causes."

I

Consider this example: during his sophomore year Tom Smith played on both the baseball and football varsity teams; we have been informed that he broke his ankle in one of the games; what are the *a posteriori* probabilities in favor of baseball and football, respectively, as the baneful cause of the accident?

Evidently the answer depends on the number of baseball and football games played during their respective seasons and also on the likelihood of a man breaking an ankle in one or the other of these two games. As a concrete case assume that:

1. At Smith's college an equal number of baseball and football games are played per season;
2. Statistical records indicate that if a student participates in a baseball game the probability is 2/100 that he will break an ankle and that, likewise, the probability is 7/100 for the same contingency in a football game.

In view of the first of these two assumptions our conclusions as to the cause of the accident may be based entirely on the information contained in the second assumption. The odds are two to seven, so that the *a posteriori* probabilities regarding the two admissible causes are:

For baseball, $2/(2 + 7) = 2/9$.

For football, $7/(2 + 7) = 7/9$.

Now consider this other example. A lone diner amused himself between courses by spinning a coin. We elicited from the waiter that in 15 spins, heads turned up seven times. Moreover, from our point of observation, the size of the coin indicated that it was either a silver quarter or a ten-dollar gold piece. What are the *a posteriori* probabilities in favor of the silver quarter and the gold piece, respectively?

If the lone diner were a professor from one of our eastern universities we would not hesitate a moment in declaring that the coin spun was a quarter. But it happens that the gentleman was a member of the Cleveland Chamber of Commerce, dining at the Bankers' Club. We must, therefore, give the matter more careful consideration. The number of quarters and gold pieces usually carried by a banker and the probabilities of obtaining the observed result by spinning coins are relevant; let us assume, therefore, that:

1. The small change purse of a Cleveland financier contains, on the average, ten-dollar gold pieces and quarters in the ratio of eight to three.

Moreover, we may assume (in fact we know) that:

2. If either a quarter or a gold piece is spun 15 times, the probability that heads will turn up seven times is approximately 1/5.

The second of these two items of information makes the *a posteriori* probabilities depend entirely on the first item. Clearly the odds are eight to three and we conclude;

For a quarter, *a posteriori* probability $= 3/(3 + 8) = 3/11$.
For a goldpiece, *a posteriori* probability $= 8/(3 + 8) = 8/11$.

Now regarding the general *a posteriori* problem,

"An event has happened which must have arisen from some one of a given number of causes: required the probability of the existence of each of the causes,"

what do the two examples we have just considered suggest? In both problems we inquired into:

1. The frequency with which each of the possible causes is met with BEFORE THE OBSERVED EVENT HAPPENED. This frequency is called the *a priori existence* probability for the corresponding cause.
2. The probability that a cause, if brought into play, would reproduce the observed event. This probability will hereafter be referred to as the *a priori productive* probability for the cause in question.

In the case of the broken ankle, the *a priori existence* probabilities were equal and took no part in our conclusion; we based the *a posteriori* probabilities entirely on the *a priori productive* probabilities. We did just the opposite with reference to the coin spun by the Cleveland financier; on account of the equality of the *a priori productive* probabilities we deduced *a posteriori* probabilities in terms of the unequal *a priori existence* probabilities.

It is apparent that our two examples represent extreme cases. In general, the solution of an inverse or *a posteriori* problem, involving a number of causes, one of which must have brought about a certain observed event, depends on both sets of direct, or *a priori* probabilities. Those of the first set give the frequency with which the various causes were to be expected before the observed result occurred; those of the second set give the frequencies with which the observed result would follow from the various causes if each were brought into play.

II

Bearing in mind the two distinctly different sets of *a priori* probabilities required in arriving at *a posteriori* conclusions regarding the possible causes of an observed event, we must now give some thought to the algebra of the subject before taking up Bayes' problem and theorem. For this purpose consider the following bag problem:

A bag contained M balls of which an unknown number were white. From this bag N balls were drawn and of these T turned out to be white. What light does this outcome of the drawings throw on the unknown ratio of the number of white balls to the total number of balls, M, in the bag? Let x be this unknown ratio.

Two cases of this problem may be considered:

Case 1.—After a ball was drawn it was replaced and the bag was shaken thoroughly before the next drawing was made;
Case 2.—A drawn ball was not replaced before the next drawing.

These two cases become essentially identical when the total number of balls in the bag is very large compared with the number drawn. Case 1 will serve as an introduction to Bayes' problem; later we will find it highly desirable to consider Case 2.

We are confronted with $(M + 1)$ possible hypotheses or causes before the drawings took place:

$$\begin{array}{rl} 1 - & \text{the unknown value of } x \text{ is } x_0 = 0/M, \\ 2 - & \text{the unknown value of } x \text{ is } x_1 = 1/M, \\ 3 - & \text{the unknown value of } x \text{ is } x_2 = 2/M, \\ \cdots & \cdots \\ k + 1 - & \text{the unknown value of } x \text{ is } x_k = k/M, \\ \cdots & \cdots \\ M + 1 - & \text{the unknown value of } x \text{ is } x_M = M/M = 1. \end{array}$$

Let $w(x_k)$ be the *a priori* existence probability for the k'th hypothesis; by this is meant the probability in favor of the k'th hypothesis based on whatever information was available regarding the contents of the bag prior to the execution of the drawings.

Let $B(T, N, x_k)$ be the *a priori productive* probability for the k'th hypothesis; by this is meant the probability of obtaining the observed result (T whites in N drawings) when the value of x is k/M.

Then, the *a posteriori* probability, or probability after the observed event, in favor of the k'th hypothesis is

$$P_k = \frac{w(x_k)B(T, N, x_k)}{\sum_{k=0}^{M} w(x_k)B(T, N, x_k)}. \tag{1}$$

For Case 1 of our bag problem we have

$$B(T, N, x_k) = \binom{N}{T} x_k{}^T(1 - x_k)^{N-T},$$

where $\binom{N}{T}$ represents the number of combinations of N things

[2] This is the Laplacian generalization of Bayes' formula, although in some textbooks it is referred to as "Bayes' Theorem." A relatively short demonstration of it is given by Poincaré in his *Calcul des Probabilités*. See also Fry, *Probability and Its Engineering Uses*, Art. 49.

taken T at a time. Substituting in (1) we obtain, after canceling from numerator and denominator the common factor $\binom{N}{T}$,

$$P_k = \frac{w(x_k)x_k{}^T(1 - x_k)^{N-T}}{\sum_{k=0}^{M} w(x_k)x_k{}^T(1 - x_k)^{N-T}}. \tag{2}$$

If in equation (2) we give k successively the values a, $a + 1$, $a + 2$, $\cdots b - 1$, b and add the results we have

$$P_a + P_{a+1} + \cdots + P_b$$

or

$$P(x_a, x_b) = \frac{\sum_{k=a}^{k=b} w(x_k)x_k{}^T(1 - x_k)^{N-T}}{\sum_{k=0}^{M} w(x_k)x_k{}^T(1 - x_k)^{N-T}} \tag{3}$$

for the *a posteriori* probability that the unknown ratio of white to total balls in the bag lies between a/M and b/M; both inclusive.

III

Bayes' Problem

Consider the table represented by the rectangle $ABCD$ in Fig. 1. On this table a line OS was drawn parallel to, but at an unknown distance from, the edges AD and BC. Then a ball was rolled on the table N times in succession from the edge AD toward the edge BC. As indicated in the figure, it was noted that T times the ball stopped rolling to the right of the line OS and $N - T$ times to the left of that line.

What light does this information shed on the unknown distance from AD to OS? In more technical terms, what is the *a posteriori* probability that the unknown position of the line OS lies between any two positions in which we may be interested?

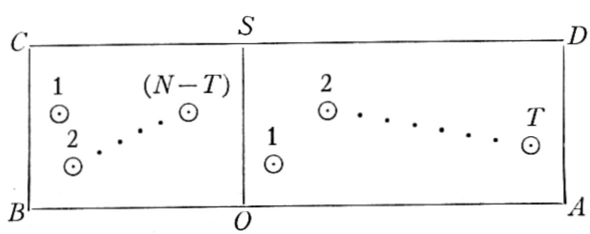

Fig. 1.

Each rolling of the ball was executed in such a manner that the probability of the ball coming to rest to the right of OS is given by the unknown ratio of the distance OA to the length BA of the table; likewise, the probability of the ball stopping to the left of OS is given by the ratio of the distance BO to the length BA.

$$\text{Set } x = OA/BA, \qquad 1 - x = BO/BA.$$

The only difference between this problem and the bag of balls problem is that now the possible values of x are not restricted to the finite set $0/M$, $1/M$, $2/M$, $\cdots$ $(M-1)/M$, M/M; in the table problem x may have had any value whatever between the limits 0 and 1. Therefore equation (3) will answer the question asked provided we substitute definite integrals in place of the finite summations. This substitution gives us, for the desired *a posteriori* probability that x had a value between x_1 and x_2, the formula

$$P(x_1, x_2) = \frac{\int_{x_1}^{x_2} w(x)x^T(1-x)^{N-T}dx}{\int_0^1 w(x)x^T(1-x)^{N-T}dx}. \tag{4}$$

Equation (4) is useless until the form of the *a priori existence* function $w(x)$ is specified; this depends on the *way* in which the line OS was drawn. Bayes assumed that the line OS, of unknown distance from AD, was drawn through the point of rest corresponding to a preliminary roll of the ball. This amounts to postulating that all values of x, between 0 and 1, were *a priori* equally likely. In other words, with Bayes, the *a priori existence* function $w(x)$ was a constant which, therefore, did not have to be taken into consideration.[3] Thus, instead of equation (4), Bayes gave the equivalent of the following restricted formula:

$$P(x_1, x_2) = \frac{\int_{x_1}^{x_2} x^T(1-x)^{N-T}dx}{\int_0^1 x^T(1-x)^{N-T}dx}; \tag{5}$$

I say "the equivalent of" (5) because in Bayes' day definite integrals were expressed in terms of corresponding areas.

Equation (5) constitutes Proposition 9 of the essay, but is usually referred to as Bayes' theorem.

[3] The existence function $w(x)$ does not appear either explicitly or implicitly anywhere in Bayes' essay. This fact raises the question as to whether or not Bayes had any notion of the *general* problem of causes.

IV

Equation (5) is a very beautiful formula; but we must be cautious. More than one high authority has insinuated that its beauty is only skin deep. Speaking of Laplace's generalization and extension of the theorem, George Chrystal, the English mathematician and actuary, closed a severe attack on the whole theory of *a posteriori* probability [4] with the statement that "Practical people like the Actuaries, however much they may justly respect Laplace, should not air his weaknesses in their annual examinations. The indiscretions of great men should be quietly allowed to be forgotten."

Chrystal's advice as to the attitude one should assume toward "the indiscretions of great men" is excellent, but in the case under consideration, it was the plaintiff rather than the defendant who committed indiscretions; this is discussed in a paper by E. T. Whittaker [5] entitled "On Some Disputed Questions of Probability."

The discussions and disputes, which began shortly after the birth of the formula in 1763 and which have not as yet subsided, may be divided into two classes:

1. Discussions concerning problems in which it is known that the *a priori* existence function is not a constant.
2. Discussions concerning problems in which nothing whatever is known concerning the *a priori* existence function.

The discussions of Class 1 are out of order in so far as Bayes' theorem is concerned; recourse should be had to formula (4), Laplace's generalization of the Bayes' theorem, when it is known that $w(x)$ is not a constant. Failure to differentiate explicitly between equations (4) and (5) has created a great deal of confusion of thought concerning the probability of causes. The discussions of Class 2 have centered on what Boole called "the equal distribution of our knowledge, or rather of our ignorance," that is to say "the assigning to different states of things of which we know nothing, and upon the very ground that we know nothing, equal degrees of probability." Regarding the legitimacy of this procedure Bayes himself contributed a very important scholium which appeared in his essay on pages 392 and 393. The argument in this scholium, based on a corollary to Proposition 8 of the essay, may be summarized as follows:

Assuming that all values of x are *a priori* equally likely and that the N throws of a ball on the table have *not yet* been made, the probability

[4] "On Some Fundamental Principles in the Theory of Probability," *Transactions of the Actuarial Society of Edinburgh*, Vol. 11, No. 13.

[5] *Transactions of the Faculty of Actuaries in Scotland*, Vol. VIII, Session 1919–1920.

that T times the ball will rest to the right of OS and that the remaining $N - T$ times it will rest to the left of OS is (as shown in the corollary)

$$P = \int_0^1 \binom{N}{T} x^T(1 - x)^{N-T}dx = \frac{1}{N+1}, \tag{6}$$

a result in which T does not appear. In other words, any assigned outcome for the throws is no more, or no less, likely than any other outcome, if *a priori* all values of x are equally likely. But, wrote Bayes in the scholium, when we say that we have no knowledge whatever *a priori* regarding the ratio x, do we not really mean that we are in the dark as to what will be the outcome when we proceed to make N throws? If so, then equation (6) justifies the assumption that *a priori* all values of x are equally likely.

To clinch his argument it must be shown that the converse of equation (6) is true. That is, it must be shown that, if any outcome of throws *not yet* made is as likely as any other, then any value of x is *a priori* as likely as any other. This converse theorem was submitted to Dr. F. H. Murray who obtained an elegant proof based on a theorem of Stieltjes.[6]

In view of Bayes' corollary and his scholium, an analysis of our bag problem with reference to the "equal distribution of our knowledge, or ignorance" is in order.

Consider again Case 1 where each drawn ball is replaced in the bag before the next drawing is made.

Assuming each of the $(M + 1)$ permissible hypotheses to be *a priori* equally likely, the probability that N drawings, *not yet* made, will result in T white and $N - T$ black balls is

$$P = \sum_{k=0}^{M} \frac{1}{M+1} \binom{N}{T} \left(\frac{k}{M}\right)^T \left(1 - \frac{k}{M}\right)^{N-T}. \tag{7}$$

Equation (7) is not, in general, independent of T,[7] so that any one assigned outcome of N drawings is not as likely as any other outcome. This result is disturbing; at first sight it seems to discredit Bayes' scholium. We must, therefore, look into the matter more closely.

Bayes' problem corresponds to drawings from a bag containing an infinite number of balls. Therefore, even if drawn balls are replaced,

[6] *Bulletin of the American Mathematical Society*, February 1930.

[7] Consider, for example, the case of $M = 2$. Equation (7) reduces to

$$P = \frac{1}{3}\left(\frac{1}{2}\right)^N \binom{N}{T},$$

a result which is not independent of T.

the chance of a particular ball being drawn more than once is zero. But when N drawings with replacements are made from a bag containing a *finite* number, M, of balls, we are by no means certain of drawing N different balls; a particular white ball may be drawn several times over and, likewise, a particular black ball may appear more than once. It is not surprising, therefore, that Case 1 of the bag problem does not confirm Bayes' corollary.

Consider now Case 2, where the drawn balls are not returned to the bag. If k of the total balls are white and the rest black, the probability that a sample of N balls from the bag will contain T white and $N - T$ black is

$$\binom{k}{T}\binom{M-k}{N-T}\bigg/\binom{M}{N}.$$

Hence, if the permissible values 0, 1, 2, 3, $\cdots$ M for k are all equally likely *a priori*, we obtain instead of (7),

$$P = \sum_{k=0}^{M}\left(\frac{1}{M+1}\right)\binom{k}{T}\binom{M-k}{N-T}\bigg/\binom{M}{N} = \frac{1}{N+1}, \tag{8}$$

a result independent of any assigned value for T and identical with the result in the corollary to Proposition 8 of the essay.

Summary

Bayes' theorem is the answer to a special case of the general problem of causes. The special case postulates that the *a priori* existence probabilities for the various admissible causes of an observed event are equal.

In the essay Bayes recommends that his theorem be adopted whenever we find ourselves confronted with total ignorance as to which one of several possible causes produced an observed event. To justify this recommendation Bayes takes the attitude that: a state of total ignorance regarding the causes of an observed event is equivalent to the same state of total ignorance as to what the result will be if the trial or experiment has not yet been made. This interpretation is a generalization of the fact that in his billiard table problem, the assumption of equal likelihood for all possible positions of the line OS, gives equal probabilities for the various possible outcomes of a set of N ball rollings not yet made.

Laplace, Poincaré and Edgeworth [8] have shown that the *a priori existence* function $w(x)$, which appears in the Laplacian generalization

[8] Laplace: "Oeuvres," Vol. 9, p. 470. Poincaré: "Calcul des Probabilités," 2d edition, p. 255. Bowley: "F. Y. Edgeworth's Contributions to Mathematical Statistics," pp. 11 and 12.

of Bayes' theorem, is of negligible importance when the numbers N and T are large. Therefore, when this condition holds, one need not hesitate to use Bayes' restricted formula for the solution of a problem of causes.

The transmission, by Price, of Bayes' posthumous essay to the Royal Society marked an epoch in the history of the literature on probability theory. As mentioned at the beginning of this paper, Karl Pearson has called the extension of Bayes' problem the "Fundamental Problem of Practical Statistics."