

## THE NATURE OF PROBABILITY

*Sir Ronald Fisher*<sup>1</sup>

IT IS NO SECRET—it is a fact that I have stressed particularly in a recent book of mine on scientific inference<sup>2</sup>—that grave differences of opinion touching upon the nature of probability are at present current among mathematicians. I should emphasize that mathematicians are expert and exceedingly skilled people at the particular jobs that they have had experience of—in particular: exact, precise deductive reasoning. In that field of deductive logic, at least when carried out with mathematical symbols, they are of course experts. But it would be a mistake to think that mathematicians as such are particularly good at the inductive logical processes which are needed in improving our knowledge of the natural world, in reasoning from observational facts to the inferences which those facts warrant. Now when we are presented, as we are at the present time in the 20th century and perhaps especially in this country, with grave differences of opinion of this sort among entirely competent mathematicians, we may reasonably suspect that the difficulty does not lie in the mathematics—or at least only incidentally or accidentally in the mathematics—but has a much deeper root in the semantics or an understanding of the meanings of the terms which are used.

It's not the first time that grave differences of opinion among mathematicians have occurred on this very question of probability. Looking over the history of the subject, I

<sup>1</sup> This paper represents the substance of an address given in November 1957 at Michigan State University.

<sup>2</sup> *Statistical Methods and Scientific Inference* (Edinburgh: Oliver and Boyd, 1956).

think we can say that a crucial set of circumstances occurred at an early period, in the 17th and 18th centuries, at the time when the interest of mathematicians in the area of probability hung upon the high social prestige of the recreation of gambling, and mathematicians were constantly being approached by persons of the highest social standing, worthy of every respect and service, in order to solve the knotty problems that arose in this recreation; and this activity was manifestly the mainspring of the interest of the galaxy of distinguished mathematicians who, at that period, gave their attention to the subject.

May I just mention a few names illustrative of that period: Pascal, Fermat, Leibnitz, Montmort (all of whom functioned principally in France), De Moivre and Bayes (in England), and Bernoulli (who didn't live quite in France because he was a member of a distinguished family of the town of Basel). And I am inclined to say that all of those founders of the mathematical theory of probability understood the meaning of the word in one way, and they had the great advantage of coming to an understanding of the word which they used in their work, in that they were brought frequently into contact with its practical applications in the real world.

Now one of the difficulties in the teaching of mathematics in the present century is the difficulty of representing in mathematical departments those arts, crafts, skills, and technologies to which statistics is now being actively applied. It would seem an almost impossible task to staff a mathematical department, to get even a representation of the immense variety of practical affairs in which mathematics or statistics is applicable and is now being used. That is a problem for the organizers of education.

My own problem is a much narrower one. I want to make clear what I mean by probability; I want to make clear, so far as I can, why it is that quite a number of mathematicians fall into what I consider to be manifest fallacies in this field.

My business, you see, is one in semantics, the meaning of the word; and the meaning of the word only comes into existence by usage, and so I define the usage that I am concerned with as that of these 17th and 18th century mathematicians. If we wish to speak about something else from that which they call probability, then I think we should find a different word; but I doubt if there is anything else of so great importance that we should consider. We can trace, I think, some of the difficulties of such a word to the mathematical mind. Clearly, the purpose of the notion of probability is to express—and express accurately, with mathematical precision—a state of uncertainty; and states of uncertainty are not familiar in the processes of exact deductive reasoning.

Probability is, I suggest, the first example of a well specified state of logical uncertainty. Let me put down a short list of three requirements, as I think them to be, for a correct statement of probability, which I shall then hope to illustrate with particular examples. I shall use quite abstract terms in listing them.

- (a) There is a measurable reference set (a well-defined set, perhaps of propositions, perhaps of events).
- (b) The subject (that is, the subject of a statement of probability) belongs to the set.
- (c) No relevant sub-set can be recognized.

I expect that these words will acquire a meaning from the examples I have to give.

Let us consider any uncertain event. A child is going to be born. I don't know enough about the present state of medical science to know whether experts exist who are really capable of saying in advance of what sex the child will be. But let us imagine ourselves in the technology of the 19th century, when certainly no such statement could be made with any confidence. This is my first example of a matter in which we

are in the state of uncertainty; that is to say, we lack precise knowledge, but we do not lack all knowledge. On inquiry at the registrar, we may find that in his experience, or in the experience of much larger numbers recorded by registrars in different parts of the world, a fixed proportion of the births has been of boys and the remainder of girls. Let us suppose he tells us that in 51 per cent the births are those of boys (a little more than 51 per cent in most populations). To the registrar, the birth which is about to take place, though intensely important to ourselves, is just another birth. To him it belongs to this set of his experience of sex at birth, and he very properly informs us that the probability of a boy is 51 per cent, having made reference to this measurable reference set as the basis of his statement.

Secondly, we satisfy ourselves as to the existence of relevant sub-sets. I need not use the word "random" because all I need say can be said under "(c)," which is the most novel in its formulation if not in its idea, the most novel of the requirements I have listed. This is a formulation which I submit to your judgment as a competent formulation of what is needed if we are to speak without equivocation of a probability of something in the real world.

The registrar might raise such a question as this: Is it a white birth or a colored birth? In his experience, the sex ratio might be different. Very well, then, it's a white birth. We have recognized a sub-set of white births, and he must turn to his tables and find out what the proportion is in respect to white births, ignoring those which do not belong to the particular sub-set to which our event belongs. Or again, his experience might have shown that first births have a higher sex ratio than births in general. He will then inquire whether our birth is a first birth or not. If it is a first birth, it belongs to a relevant sub-set. It is now recognized and takes the place of the reference set with which we started.

Exactly the same considerations may be applied to any

other case of uncertainty. Let us take the case of deliberately arranged uncertainty, which occurs in games of chance.

I mentioned the importance of the recreation of gambling as calling attention of mathematics to this new concept of probability in the 17th century. The concept was unknown to the Greek mathematicians; it was also unknown to the Islamic mathematicians, perhaps because gambling was forbidden by the Prophet. But it was not only the taste for gambling, I think, which made the difference; it was the fact that by the 17th century the technology of the manufacture of the apparatus of games of chance had reached a point at which the calculations of mathematicians have some relevance. They were not playing with knuckle-bones; they were playing with very well made dice.

Consider the gambler who has laid a stake on the assertion that an ace will be thrown. It's worth a lot of money to him. He doesn't want to mistake your meaning if you say, as perhaps De Moivre might have said, the probability of an ace is one-sixth. In saying that, he is saying that this is just one throw out of all the possible throws that might be made, and he will regard these possible throws as a reference set, measurable, of which the fraction exactly one-sixth are aces. His reasons for doing that don't immediately concern us. It is a common sense reason, perhaps, that the die has been supplied by a reputable maker, that it has six faces, that the aim of the maker has been to make it approximately a perfect cube, and to make sure that the center of gravity is equally distant from each of those faces.

Contrast that, however, with a much more sophisticated and typically useless definition of probability, which is sometimes fed to mathematical students. It goes something like this:

$$Pr\left\{\left|\frac{a}{n} - \frac{1}{6}\right| > \epsilon\right\} \rightarrow 0$$

If  $a$  aces occur in  $n$  trials, then the difference in absolute value between the fraction  $\frac{a}{n}$  and  $\frac{1}{6}$  will have a probability of exceeding any positive number  $\epsilon$ , however small, a probability which will tend to zero as  $n$  tends to infinity.

You see, that is someway away from the real world already. The gambler deserves something better than that. He may ask you, "What do you mean, 'tends to infinity'?" "Well, you go on rolling, and you don't stop—you go on rolling; you go on rolling until the die is worn to a sphere; you go on rolling until the sun goes out; but still you haven't reached infinity and are still a long way off." And then, it's not only that; as a practical man he doesn't like that, of course. "But," he says, "I asked you what you meant by probability, and here you are, you've brought in the same notion of probability in your definition. How do I know what *that* probability means?" We have a perpetual regression defining probabilities in terms of probabilities in terms of probabilities; that is a purely logical objection to the definition. But the real objection, if I may say so, for the practical gambler who wants to know about his stake, is that it says nothing about the particular throw in which he is interested. It says something about what we should ultimately regard as the reference set, certainly; but it says nothing whatever about his particular throw. And of course it might occur to him that though this was true of throws in general, yet in particular groups of throws within that general set, in particular sub-sets, the fraction might be different, perfectly consistently with this general statement.

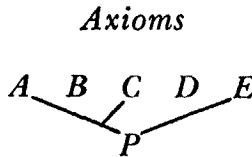
Consider a few possible sub-sets. Here's a recognizable sub-set: throws made on Friday. He can recognize that sub-set of possible future throws, and he knows his throw is one of them. But so far as we know, shall we say, according to the axioms on which the mathematicians were advising the gambler, throws made on Friday do not give a different frequency

of aces from throws made on other days. So it is recognizable, but not relevant. It doesn't alter the estimate. And then, perhaps you say, odd numbers: 1, 3, or 5. A very relevant sub-set, if it could be recognized. But the makers of dice and other apparatus of gambling have taken care—they have taken a great deal of trouble to make sure, in fact—that such a sub-set cannot be recognized before the dice are thrown. And, thirdly, let us suppose that our gambler has heard of Professor Rhine of Duke University, and that in the opinion of Professor Rhine, some of his students have the remarkable gift of precognition. The gambler perhaps makes an agreement with such a student to sit by his side while he is rolling the dice and give him a nudge when an ace is coming. Here you have, let us say, two possible cases. Perhaps the prophet is some good—and what that means is that the sub-set of throws in which he gives the signal to his patron has a proportion of aces which is greater than one-sixth—it is possible it might be a third if he is a pretty good prophet. And in that case I submit that the gambler has a recognizable and a relevant sub-set, and that to him, on his knowledge, on his information, on his data as we sometimes say, the probability is not one-sixth, but a third. On the other hand, if, after some experience he comes to the conclusion that his prophet is no good at all, he will not lose his knowledge of the probability—it will merely revert to its value of one-sixth. He will now be in the position of saying that there is a measurable set with a frequency of one-sixth, and there is no relevant and recognizable sub-set which I should prefer to it.

Now that, I hope, sounds easy, and I want to get a little closer to the psychological difficulties which cause difference in understanding as to the meanings of these words.

The first difficulty is that we are making a statement of uncertainty, and that statements of uncertainty are not familiar in the ordinary course of deductive mathematical argument. They introduce special logical requirements. You notice, my

third condition was that no sub-set should be recognizable. It is a postulate of ignorance. How are we to take account of postulates of ignorance, as we have to do in inductive reasoning? In the ordinary course of deductive reasoning, the reasoner is supplied with what I shall call, for the moment, "axioms"—the term doesn't matter very much—and if he can prove what he wants to prove by using axiom *A*, axiom *C*, and axiom *E* to give the proposition, he is perfectly entitled to do so because he is arguing with certainty, and the truth of axioms *A*, *C*, and *E* are not at all precluded or interfered with by his axioms *B* and *D* that have not entered into his argument.



But suppose he were making a statement of uncertainty. Then *B* and *D* do matter. In inductive reasoning the whole of the data, or the available axioms, or the available observations, has to be taken into account, and it is only because of that particularity of inductive reasoning that axioms of ignorance matter. There the postulate of ignorance asserts that certain things are not known and that the validity of the argument requires that they should not be known; and of course this is fundamental to any correct statement of uncertainty. If all sorts of other additional information could be sprung on you at any stage in the argument, you might discover there was no uncertainty at all, or, more easily, that the degree and nature of uncertainty which you have arrived at is totally different from what should have been arrived at if everything had been taken into account.

Now, at the end of the last century, a group of rather distinguished mathematicians, Hilbert, for example, and Peano,



set out on a project which was to show that the whole of mathematics could be deduced with strict irrefragable logic from certain chosen axioms. Peano had a shot at setting up such axioms that would suffice for the deduction of the whole of mathematics. That project was influential—it still is influential, I think, in spite of the setbacks that it has received. It was influential, for example, in producing Whitehead and Russell's *Principia Mathematica*. It was quite fundamental to Keynes' book on probability.

But difficulties have arisen. It was fairly easily demonstrated, and it came as a surprise to a good many people, that if a system of axioms allowed of the deduction of any contradiction (any fallacy, if you like)—if it allowed the proposition  $P$  and also the proposition *not-P* to be deduced by the ordinary rigorous processes from the same system of axioms—then that system of axioms contained latent *all* contradictions, in the simple sense that any proposition whatever could be deduced from them.

There is a story that emanates from the high table at Trinity that is instructive in this regard. G. H. Hardy, the pure mathematician—to whom I owe all that I know of pure mathematics—remarked on this remarkable fact, and someone took him up from across the table and said, "Do you mean, Hardy, if I said that two and two make five that you could prove any other proposition you like?" Hardy said, "Yes, I think so." "Well, then, prove that McTaggart is the Pope." "Well," said Hardy, "if two and two make five, then five is equal to four. If you subtract three, you will find that two is equal to one. McTaggart and the Pope are two; therefore, McTaggart and the Pope are one." I gather it came rather quickly.

That wasn't, however, the worst that befell the theory of the axiomatic basis for mathematics. It pinpointed the need for some means of demonstrating that a system of axioms *was* free from all contradictions, because if it wasn't it could lead

to anything. And then the blow fell, which was due, I believe, to Gödel, who put forward a very long, very elaborate, and extraordinarily ingenious proof to the effect that you could not, basing your reasoning upon a given system of axioms, disprove the possibility that that system could lead to a contradiction. Now that was a surprise to people, but I don't think it ought to have been. After all, suppose a Ph.D. student came, breathless with excitement, and said, "I have *proved* that this system of axioms is free from all contradictions." You'd say, "Did you prove it using only those axioms?" He might say, "Yes, I have written out a chain of propositions which demonstrate that these axioms are free from contradiction." Well, I suppose you'd look at him with mild surprise, and you might say, "I suppose you know that if this system of axioms *did* contain a contradiction, you could prove exactly those same propositions." And so you have the situation that certain propositions which purport to prove the truth, the truth of the theorem, could be equally well demonstrated by the ordinary rigorous processes of deductive reasoning if they were false. And I don't know how much we would give, then, for the chain of theorems which purported to prove that the system of axioms was free from contradictions. It would seem to be a little absurd to imagine that such a thing was possible.

Now, if I were to illustrate the mathematics, it would not appeal to a large proportion of the audience. But I want to give a few comparatively slight illustrations of how the controversies that I have alluded to affect our practical mathematical reasoning. Some of us think that if one had a sample which was known to be drawn from a normal population—a sample of  $N$  observations,  $X_1, \dots, X_N$ —that by taking the mean of that sample (that is, by adding up the individual observations and dividing by their number), and by taking the mean square deviation, using the sum of  $(X - \bar{X})^2$ , treating it appropriately, as Gauss suggested, and getting what is called the

sample variance of the mean,  $s^2 = S/N(N-1)$ —some of us believe that one can then make probability statements of the kind that the true mean ( $\mu$ ) of the population is less than a calculable limit with an exactly known probability. In fact, the statement can be made that the probability that the unknown mean of the population is less than a particular limit, is exactly  $P$ . Namely  $Pr(\mu < \bar{x} + ts) = P$  for all values of  $P$ , where  $t$  is known (and has been tabulated as a function of  $P$  and  $N$ ).

This is exactly the sort of specification of our uncertain knowledge of the constants of nature that scientists have for a hundred years thought they possessed about them. The conditions required are more stringent than has been generally realized, but these conditions can be met in a number of useful cases, and in these cases the quantity under discussion, although of course not known with exactitude, is accurately specified as a random variable about which exact probability statements can be made for all possible values of the probability.

This is a single example of a large number of such inductive inferences that are made by the same process of reasoning. They have been disputed, I think principally on this ground, that it is not clear to all mathematicians that a probability statement is based on data, and that it is no defect in such a probability statement that it would be different if the data were different.

Let me examine this simple example. We have a limit which we can calculate, and it is undoubtedly true that this limit exceeds  $\mu$  with given probability in the reference set defined by any value of  $\mu$ . If a population with a mean  $\mu$  were sampled repeatedly, we would certainly get this quantity exceeding  $\mu$  with a given probability. That, I believe, is not disputed. It is also true that if we take the statement in general we have proved it for all  $\mu$ , and therefore for the reference set for all samples from all populations. Each

sample has peculiar values  $(\mu, \bar{x}, s)$ , and for this enlarged reference set it is true that  $Pr(\mu < \bar{x} + st) = P$ , where  $t$  is "Students" deviate corresponding with the (one-sided) probability  $P$ .

That, however, does not settle the matter. There are two conditions which should be satisfied in addition. I would like to emphasize these because you will find examples in the literature where this sort of inference is drawn without any reference to the conditions, and usually drawn with reference to what is really irrelevant, namely, certain beliefs about tests of significance—"the theory of testing hypotheses," or perhaps the theory of decision functions. The two requirements that *are* necessary flow from the third condition which I laid down for a correct statement of probability, namely, that no relevant sub-set should be recognizable.

Now suppose there were knowledge a priori of the distribution of  $\mu$ . Then the method of Bayes would give a probability statement, probably a different one. This would supersede the fiducial value, for a very simple reason. If there were knowledge a priori, the fiducial method of reasoning would be clearly erroneous because it would have ignored some of the data. I need give no stronger reason than that. Therefore, the first condition is that there shall be no knowledge a priori. And the second condition is that in calculating the limit, the second term of the inequality concerned, we should have used exhaustive estimates. The two estimates that we are concerned with are the mean and variance (estimate of the mean, estimate of the variance), and those happen to be exhaustive in a mathematical sense when calculated from the normal distribution, but not from other distributions. If they are exhaustive, then it is known that given these two quantities,  $\bar{X}$  and  $s^2$ , the distribution of any other statistic whatsoever (that is to say, any function whatever of the observations) would, subject to the restriction of fixing the values of  $\bar{X}$  and  $s^2$ , have a distribution indeed and take many values, but its

distribution would be independent of the unknowns  $\mu$  and  $\sigma$ . And, therefore, no such value could provide information about  $\mu$ . But if the statistics used in this argument had not been exhaustive, then it would be possible to find other functions of the observations which even under the restrictions that  $\bar{X}$  and  $s$  are fixed, would have information to give about the unknown  $\mu$ . Such a value, calculated from the sample, would define a sub-set of cases which might well give a different probability from that which we have arrived at. So the rigorous application of that third specification of what is needed for a true statement of probability brings in the two requirements for a valid argument of this kind.

Now of course I haven't listed all or anything like all of the fallacies that have been introduced, largely springing from the same roots, but as I suppose is familiar, whether you think of error or whether you think of sin, one leads to another. Once a person has harbored an error in his undergraduate days, carefully implanted there by some distinguished but muddle-headed professor, he may go on for a long while without being enabled to work it out by his own powers of thought. At least it's scarcely conceivable that the mathematicians of the 19th century should have harbored the notion of inverse probability from about 1812, when Laplace published his *Théorie Analytique*, to what I suppose would be the best terminus, 1886, when, speaking of my own country, Crystal published his great *Algebra*, in which he took the unprecedented step of throwing out the whole business of probability altogether as being too hopelessly unsound to be included in a good book on algebra. That was good for the teaching of algebra, and I am inclined to think, though it is a matter of judgment, that it was also good for statistical studies in England. The same movement of thought was going on, to some extent, in other countries, but not quite so abruptly and dramatically as it did in England, and the result in England was that the study of probability, when it

re-emerged from its temporary eclipse, re-emerged well embedded in a much larger discipline which is commonly known as statistics at the present time.

Of course, there is quite a lot of continental influence in favor of regarding probability theory as a self-supporting branch of mathematics, and treating it in the traditionally abstract and, I think, fruitless way. Perhaps that's why statistical science has been comparatively backward in many European countries. Perhaps we were lucky in England in having the whole mass of fallacious rubbish put out of sight until we had time to think about probability in concrete terms and in relation, above all, to the purposes for which we wanted the idea in the natural sciences. I am quite sure it is only personal contact with the business of the improvement of natural knowledge in the natural sciences that is capable to keep straight the thought of mathematically-minded people who have to grope their way through the complex entanglements of error, with which at present they are very much surrounded. I think it's worse in this country than in most, though I may be wrong. Certainly there is grave confusion of thought. We are quite in danger of sending highly-trained and highly intelligent young men out into the world with tables of erroneous numbers under their arms, and with a dense fog in the place where their brains ought to be. In this century, of course, they will be working on guided missiles and advising the medical profession on the control of disease, and there is no limit to the extent to which they could impede every sort of national effort.